107 年委託研究報告

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告

計 畫 委 託 機 關 ： 國 家 通 訊 傳 播 委 員 會

中華民國一〇七年十二月

107 年委託研究報告

PG10707-0082

數位匯流影音平臺服務品質量測方法之委託研究採購案

受委託單位

財團法人電信技術中心

計畫主持人

施嘉興

研究人員

呂維曦、楊智鈞、吳長益、周傳凱

簡文祥、薛茂呈、廖珮如

中華民國一〇七年十二月

目次

表次

圖次

提要

關鍵詞：數位匯流、QoS、QoE、服務品質、MOS

現今的網際網路應用已屬於數位匯流的主流時代，閱聽人透過開放式的網際網路，就擁有多樣的行動載具得以接收各種線上視訊內容、語音通訊等服務。而數位匯流線上影音服務快速發展已直接衝擊傳統廣播電視，甚至改變視聽者的收視習慣，近來出現的「剪線運動」（cord-cutting movement）[1]凸顯出使用者減少對傳統電視的依賴。而轉向透過網路、行動寬頻吃到飽服務下的線上影音服務，而目前台灣平均每人每月行動數據傳輸量已超過 14GB，更是高居世界排名第一[2]。

但因開放式網際網路不具有 QoS 保證的特性，加上不同的終端裝置具備多樣性的螢幕尺寸、各用戶多樣化的使用情境等因素，如何掌握用戶的體驗品質、進而優化架構以改善品質，成為線上影音服務營運與管理的重要課題，本研究針對數位匯流影音平臺服務品質量測方法進行研析，目前本研究將包含下列四大工作項目：

一、 國內影音服務產業現況：介紹目前的數位匯流影音服務的各種型態，與國內通傳事業之數位影音匯流服務架構解析，了解國內數位匯流影音平臺服務產業現況，並探討匯流影音產業面臨轉型挑戰與侵權困境。

二、 國際監理政策研析：透過了解數位匯流影音服務品質監理的政策意涵，再研析美國、加拿大、英國、法國、新加坡五個主要國家的數位匯流影音服務品質監理政策，探討網路中立性對數位匯流影音服務品質監理之影響，並說明服務品質與網路中立性的監理目前之挑戰，最後針對目前國內的線上影音平臺服務法規進行研析並提出法規方面之建議。

---

[1]台灣有線電視「剪線潮」警訊？用戶數上半年少了 6 萬 8422 戶，東森新聞雲（2018/8/19），https://www.ettoday.net/news/20180819/1238703.htm#ixzz5PdtidQGW (檢閱日期：2018/08/30)
[2]黃晶琳（2017/12/17）。台灣 4G 傳輸量 稱冠全球，經濟日報，https://money.udn.com/money/story/8888/2879232

三、 國際間影音服務品質量測方法研析：從技術面向研析國際間針對線上影音服務品質（QoS 與 QoE）之量測方法，從影音服務的量化品質面向，透過網路性能指標與使用者體驗感受的主客觀量測，來蒐集了解影音服務相關網路效能參數、影音服務體驗指標，量測方法可測量性與合理性等驗證。

四、 布建量測與結果：本研究案選擇在用戶端執行客觀 QoE 量測並以開發用戶端應用程式作為量測工具，以利直接蒐集與用戶實際感受最接近的度量指標結果或是事件紀錄，同時也可以一併記錄用戶端量測到的網路 QoS 參數並且進行分析與 QoE 之間的關聯性。並透過量測數據的驗證其量測工具之可靠度與可信度，從中發現到影片解析度與網路品質直接影響 vMOS 分數，越高解析度的影片串流需要更高品質的網路環境條件，當網路品質不佳時反而會因頻繁地發生卡頓造成嚴重地體驗品質下降。

五、 活動辦理記錄與成效：本案辦理總共辦理一場座談會、兩場說明會、三場教育訓練，透過辦理這些活動促進產、官、學界的交流。

透過上述五大工作項目，本研究案研析了國內影音服務產業現況、國際監理政策、並探討各個客觀 QoE 模型視訊品質量測方法，介紹了涵蓋研究論文、產業白皮書與標準文獻中的十種客觀 QoE 模型。並基於本中心既有之研究所提出具體量測方法為基礎進行本研究量測工具之建置，透過實際進行佈建超過 200 個用戶數量測所蒐集超過 20000 筆之數據，來驗證本研究量測方法與工具具有可靠度與可信度，有了可靠的評估指標與量測方法，即可提供串流影音服務營運與管理（或監理）公開服務品質資訊與訂定品質規範的參考，協助釐清消費者爭端與提升數位匯流影音平台之服務品質，以維護民眾之消費權益。

Abstract

Keywords: digital convergence, QoS, QoE, quality of service, vMOS

The Digital convergence application is growing rapidly for the internet application. Through the Open Internet, readers have a variety of mobile vehicles to receive a variety of online video content and voice communications. Online audio and video service has directly impacted traditional radio and television even changing the habits of the customers. And "cord-cutting movement" was a recent phenomenon which represents users reduce their dependence on traditional TV. Users can enjoy online audio and video services through the Internet services. In Taiwan, the average mobile data throughputs has exceeded 14GB for each person in a month, which is almost highest in the world.

However, the video streaming data transmission on internet is always in non-QoS environment. It’s become an important issue to improve quality and identify quality of user experience（QoE）for online audio and video service. This research we will focus on Digital convergence audio and video platform service quality measurement method. In this research will include the following work items: 1. Taiwan online audio and video service industry status. 2. The international political issue of supervision for digital convergence. 3. Analysis of International Quality Measurement Methods for online audio and video service. 4. Measurement tool and deploy method for measurement. 5. Symposium of this study.

Through the above work items, this research case studies the video quality measurement methods of each objective QoE model. We introduced ten objective QoE models including research papers, industry white papers, and literature. Based on our existing research foundation. We propose specific video mean opinion score（vMoS）methods and measurement tools. And we will construct measurement which verify the

measurement method of this study. We deploy over 200 measurement equipment and collect over 20000 data of vMoS and QoS measurement. And we verify reliability and validity from measurement result.

In this study, we analysis the political issue of Digital convergence, online video streaming, and provide our suggestion in study. And we develop a measurement APP tool for video streaming QoE measurement. We also verify reliability and validity from measurement result. So we can using this measurement tool to clarify QoE problem, and help online video service provider and ISP identify quality of video streaming problem, improve user experience..

第一章 國內影音服務產業現況

第一節 風起雲湧的數位匯流影音服務

在數位匯流時代，閱聽人擁有多樣的行動載具得以藉由影音串流技術（Video Streaming）接收線上影音內容（Online Video Content）。藉由網際網路提供數位匯流影音內容的方式，約略可分為兩大類：一是 IPTV（Internet Protocol Television）係由電信業者以其內建封閉式網路（walled garden）將電視節目內容傳送至訂戶收視，國內以中華電信推出之 MOD 服務為其代表[3]，其優點是信號穩定，影音品質由業者管控，缺點是目前中華電信受電信法規範而無節目編排自主權，且介面易用性常受使用者詬病；二是 OTT（Over The Top），由網路服務業者提供收視平臺，直接透過開放式公眾網路提供影音內容給使用者所使用之電子終端設備收視[4]，優點是使用者得自行選擇所欲收視之節目，但缺點則是平臺業者在網路傳輸上若無妥適安排，其常受網路壅塞情況而影響收視品質。OTT 與 IPTV 最大差異在於前者不受到特定專屬網路影響且內容多樣，而 OTT 收視方式也可透過聯網電視系統或機上盒等具備寬頻連線功能之設備，讓觀眾以電視終端得以觀看 OTT 內容，例如 Apple TV、Chromecast、XBOX 等，可擴充聯網能力，提供多樣化的網路應用服務[5]。

數位匯流線上影音服務快速發展已直接衝擊傳統廣播電視，甚至改變視聽者的收視習慣，近來出現的「剪線運動」（cord-cutting movement）[6]凸顯出使用者減少對傳統電視的依賴。對於擁有多螢裝置的現代人而言，隨身攜帶的手機、平板等，均能無縫隙地填補生活空白，隨時供使用者上網觀看，加上近來行動寬頻吃到飽費率，平均每人每月行動數據傳輸量已超過 14GB，高居世界排名第一[7]。

---

[3]許琦雪（2015/4/30）。OTT 競爭下有線電視產業的危機與轉機—跳脫傳統電視框架。NCC News，104 年 4 月號。
[4]葉志良（2015）。我國線上影音內容管制的再塑造：從 OTT 的發展談起。資訊社會研究，29 期，頁 47-92。
[5]DIGITIMES（2011/10/31）。Smart TV 應用環境與技術發展。
https://www.digitimes.com.tw/iot/article.asp?cat=130&id=0000256722_kg46jrky3vlzbd1o4yhe5
[6]台灣有線電視「剪線潮」警訊？用戶數上半年少了 6 萬 8422 戶，東森新聞雲（2018/8/19），
https://www.ettoday.net/news/20180819/1238703.htm#ixzz5PdtidQGW (檢閱日期：2018/08/30)
[7]黃晶琳（2017/12/17）。台灣 4G 傳輸量 稱冠全球，經濟日報，https://money.udn.com/money/story/8888/2879232

OTT 商業模式主要仰賴網路廣告收入，亦有來自用戶之訂閱費用（採定期或依觀看內容而計），雖部分內容仍需用戶自費收視，但整體而言，費用要較傳統多頻道視訊付費平臺要少，且提供 OTT 服務無須自建基礎網路與負擔設備成本，讓更多不同屬性業者趨之若鶩，投入此市場當中[8]。

隨著寬頻網路與多螢幕設備的普及，目前 OTT 服務已進入高峰，許多消費者也能接受付費收視。現提供 OTT 服務經營者，約有以下幾類：

一、 既有廣播電視業者：包括有線電視與無線電視，使用者只要支付月費或年費即可透過電視或網路收看內容，而國內的代表業者，如公視、民視、三立、中視、華視…等。

二、 內容業者：擁有大量內容的業者可將其影片讓使用者透過網路觀看，亦得將內容授權給其他平臺業者，如酷瞧、麥卡貝…等。

三、 內容整合業者（content aggregator）：這類業者整合各方來源的影音內容讓使用者透過此平臺即可收看多類型節目，國內較知名平臺有 LiTV、Catchplay、Line TV 等。

四、 設備業者/終端裝置製造商：逐漸成為主流的聯網電視，諸如 OVO，使用者購買裝置後即可收看免費或付費內容。

五、 電信業者：目前業者大多因應 OTT 興起，自行創立平臺以吸引使用者持續利用其平臺，例如中華電信的中華影視（目前改名為 Hami Video）、台灣大哥大的 myVideo、遠傳電信的 FriDay 影音。

六、 其他業者：如搜尋引擎、入口網站、媒體出版業或電商零售業等，紛紛投入 OTT 行列，例如國內 UDN。

---

[8]曾俐穎、陳人傑（2015）。眼球經濟新藍海：影音 OTT 平臺產業發展模式之研究。2015 中華傳播學會年會暨第 12 屆傳播與媒體生態學術研討會，高雄義守大學。

表 1 國內影音服務平臺代表業者屬性分類

屬性	國內代表業者
廣播電視業者	公視、民視、三立、中視、華視…等
內容業者	酷瞧、麥卡貝…等
內容整合業者	LiTV、Catchplay、Line TV…等
設備業者/終端裝置製造商	OVO、中華電信 MOD…等
電信業者	Hami Video、friDay 影音、myVideo…等

(資料來源:本研究整理)

OTT 發展初期會受到頻寬不足與不同網路間銜接的限制，造成傳輸品質降低。但近期更須關注的是內容授權議題，因為更高速且穩定的網路傳輸技術發展快速，導致許多影音內容在未取得合法授權之前就已在線上影音平臺之間流竄，迫使許多內容供應業者必須正視與線上影音平臺的合作關係。此外，影響更大的莫過於國際大型線上影音業者，在 2015 到 2016 年之間已有 LINE TV、Netflix、Dailymotion 以及愛奇藝大舉入侵台灣市場，讓國內中小型線上影音業者的經營情形更為險峻[9]。

[9]葉志良、何明軒（2016）。OTT 產業政策白皮書。元智大學大數據與數位匯流創新中心政策法規研究團隊。

誠如前述，有線電視、IPTV 與 OTT 是截然不同的影音內容傳輸型態，有線電視與 IPTV 是在具有一定頻寬與 QoS(Quality of Service)保證下傳送影音內容給訂戶收視，而 OTT 則係基於開放式網際網路作為傳輸，提供服務的傳輸品質較難維持，目前線上影音業者大多透過內容傳遞網路（Content Delivery Network, CDN）將內容傳送到消費者，以提供較為穩定的服務品質。其中 CDN 節點會在多重地點、不同網路上設置，節點之間會互相傳遞內容。在利用 CDN 服務下，可提供內容彙集及快速存取服務，對用戶的下載進行最佳化，以提高用戶的體驗品質。正因為開放式網際網路不具有 QoS 保證特性，再加上不同終端設備具有多樣性的螢幕尺寸、各用戶多樣化的使用情境等因素，如何掌握用戶的體驗品質，進而優化架構以改善品質，誠為數位匯流影音營運與管理的重要課題。

以下本節針對目前國內有線廣播電視系統、電信事業固網以及行動寬頻等通傳事業透過行動寬頻或自建 WiFi 網路（含 MSO 及固網業者提供家用 WiFi 網路），分別對於數位匯流影音相關服務之架構及產業現況提出分析、說明。

第二節 國內通傳事業提供數位匯流影音服務之架構

我國視聽媒體產業的發展，隨科技進步與政策開放，從早期獨占壟斷的無線電視，到中期的有線電視、行動電視、直播衛星、錄影帶業者等擴大影音內容，目前則因網際網路技術進步而邁入新的個人化電視時代，電視節目不在只是透過電視收看，亦可透過網際網路隨時隨地在各種手持式裝置上收看。通訊與傳播技術匯流後，新型態的視聽媒體傳輸平臺也相繼出現，IPTV、OTT 成為視聽眾接收訊息的來源，而傳統廣電事業與電信業者也在尋求轉型當中。目前有線廣播電視系統經營業者、電信事業固網業者與行動寬頻業者之經營業務，已從過去壁壘分明到今日逐漸融合，如下圖所示。

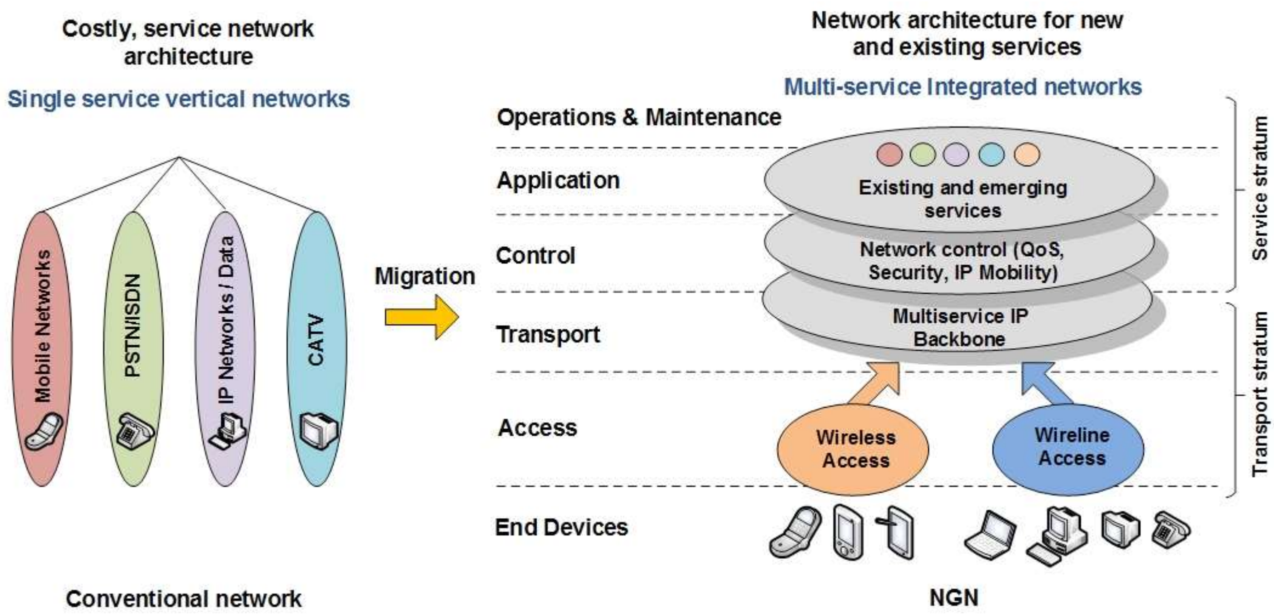


圖 1 網路產業水平與垂直架構[10]

[10]Rodríguez & Muñoz (2017). Review of Quality of Service (QoS) mechanisms over IP Multimedia Subsystem (IMS)

線上影音服務意指透過開放式網際網路（指未有 QoS 的網路環境），直接對用戶提供各種視訊內容、語音通訊等服務，其中影音內容通常以串流或下載方式傳送至用戶端的電視、電腦、智慧型手機或平板電腦等各種終端設備。與利用電信網路或有線電視系統的專用網路提供的 IPTV 相較，數位匯流線上影音服務係基於開放式網際網路作為傳輸，使得提供服務的傳輸品質較難維持，因此目前線上影音服務業者多透過 CDN 將內容傳送至消費者，以提供較為穩定的服務品質，提高用戶的體驗品質。

一、 有線廣播電視服務及其架構

有線廣播電視過去以類比訊號傳送與接收視聽內容，但隨著訊號數位化與壓縮科技的發展，原一條類比電視頻道頻寬只能乘載一個節目，但經數位化壓縮訊號後，原頻寬大小可同時播送 3 至 4 個標準畫質（SDTV）的節目，因節省許多頻寬，所以有更多的頻道空間得以釋出，使收視戶有更多節目選擇，或提供更多加值服務。數位化有線電視服務提供的模式，須將原先纜線進行數位化升級建設，下圖為數位有線電視系統與寬頻上網系統的運作架構，將原先纜線升級為光纖與同軸纜線混和網路（Hybrid Fiber Coaxial），建置數位化頭端系統，並能提供各項加值應用服務，業者亦需投入中介軟體（Middleware）與條件接取系統（Conditional Access System）建置，收視戶始得透過數位機上盒接收數位節目內容訊號以及進行互動服務，包括利用電子節目表單（Electronic Programming Guide, EPG）選取，可訂閱、搜尋、錄製所喜愛的節目內容，並具備隨選視訊功能；或藉由纜線雙向傳輸數據資訊、語音服務，整合電信與廣播技術之服務，提供電視電商股市下單、線上購物等功能，或甚至中介軟體可提供家庭連網，使數位電視加上網路傳輸成為可連網之數位電視。

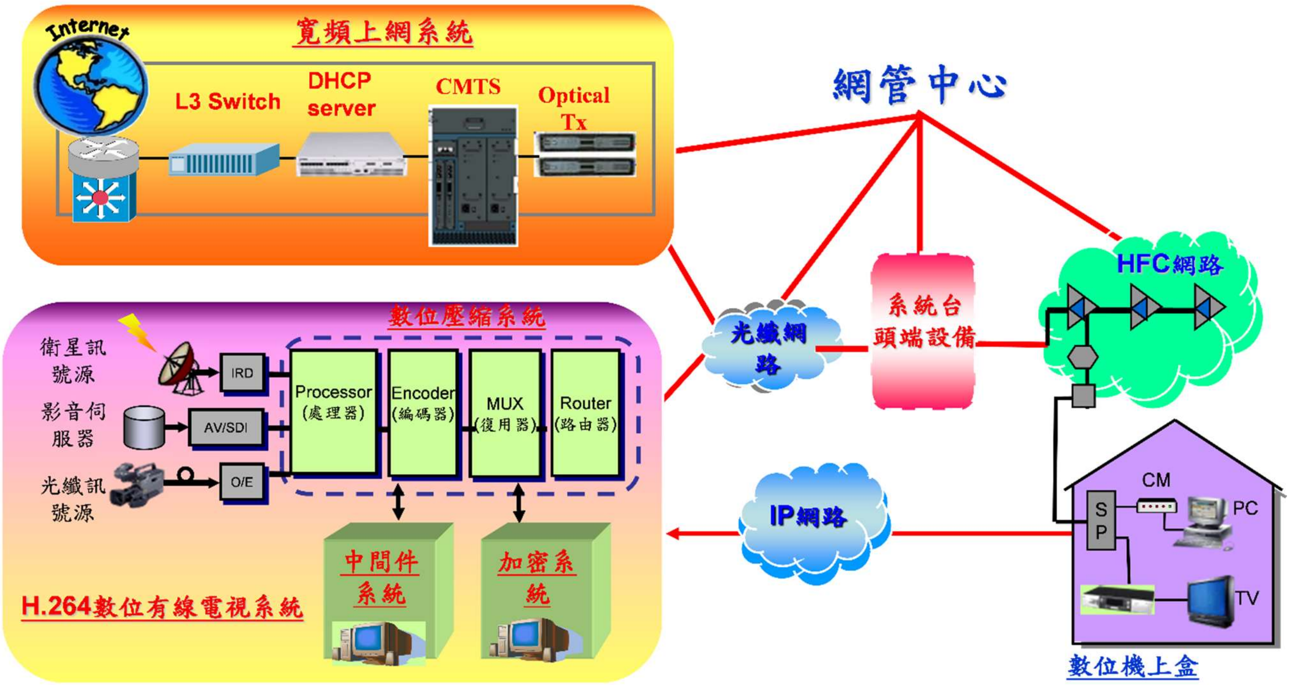


圖 2 數位有線電視系統與寬頻上網系統之運作架構[11]

二、 IPTV 服務及其架構

IPTV 是將原本電視節目內容播送的傳輸平臺轉換到以網際網路協定作為封包傳輸的系統，在該系統中，電視與視訊訊號使用網際網路協定上的寬頻網路連結分配給用戶，並起以「電視機＋機上盒」為主要終端設備，於數位機上盒轉換訊號後使用戶得於電視上收看視訊節目。由於用戶的終端設備，包括 IP 是特定的，且數位機上盒亦是由電信業者提供，故網路屬性為封閉式（walled garden）的網路環境。IPTV 因利用電信網路傳輸訊號，實務上多為結合電信業者與內容提供者所建構之封閉式傳輸平臺而提供服務，因此可整合電視、電話和寬頻上網等三合一服務（Triple Play），並可提供互動式個人化服務，電信業者藉此攻佔電視市場。

[11]凱擘大寬頻(2014)，凱擘數位頭端機房 https://www.slideserve.com/erling/5724429

國內的 IPTV 經營者主要有中華電信所推行的多媒體內容傳輸平臺（Multimedia on Demand, MOD）以及威達雲端電訊的 Vee TV，惟後者僅於台中大里地區推行，規模不足。如圖 3 所示 MOD 係透過中華電信寬頻網路，由連接至家中的電話線路及寬頻網路（ADSL 或光纖網路），藉此連接到多媒體服務系統，服務內容經由數位機上盒連接到電視。在 MOD 網路電視服務架構下，終端用戶須事先向中華電信申請寬頻接取服務以及額外的網路電視服務，由中華電信提供一台寬頻網路數據機及一台 IP 數位機上盒連接家中電視，讓終端用戶能在家中觀看影音內容。除提供一般電視頻道內容外，尚有 VOD、KTV、生活資訊、兒童專區等。

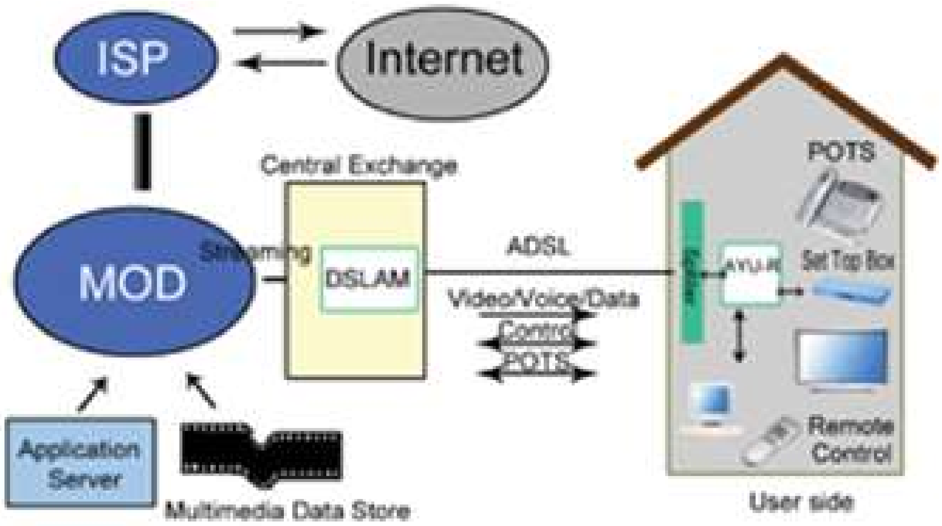


圖 3 中華電信 MOD 系統之技術架構[12]

[12]劉繼謚（2004）。解讀互動電視。https://www.ctimes.com.tw/DispArt/tw/網際管理系統/0412011123O4.shtml

三、 OTT TV 服務及其架構

OTT TV 是一種架構於網際網路之視聽媒體傳輸平臺，可完全獨立於傳統廣播電視。依據國際電信聯合會（ITU）對於 OTT TV 之定義，係指：「在網際網路之網站上觀看到一般廣播電視節目，亦即是 IPTV 定義以外之電視服務。」由於 OTT TV 是藉由開放式網際網路傳輸節目內容，因此線上影音業者實質上並無國內外之區別，只要有裝置能連上網際網路，無論是手機、平板、機上盒或遊戲機，使用者皆能觀看到影音內容；然而，OTT TV 常受到網路頻寬不足以及不同網路間銜接的限制，易造成傳輸封包延遲或遺失，節目品質難免受到影響。不過，隨著寬頻網路技術持續進步，網路接取服務已能提供更高速且穩定的網路傳輸，OTT 這類視聽傳輸平臺已逐漸侵蝕傳統廣播電視的市場大餅。

下圖以電信業者為例，業者提供用戶行動寬頻上網服務之際，同時也提供用戶便利的 OTT 影音服務，包括隨選視訊（VOD）以及及時頻道節目（Live Channels）。

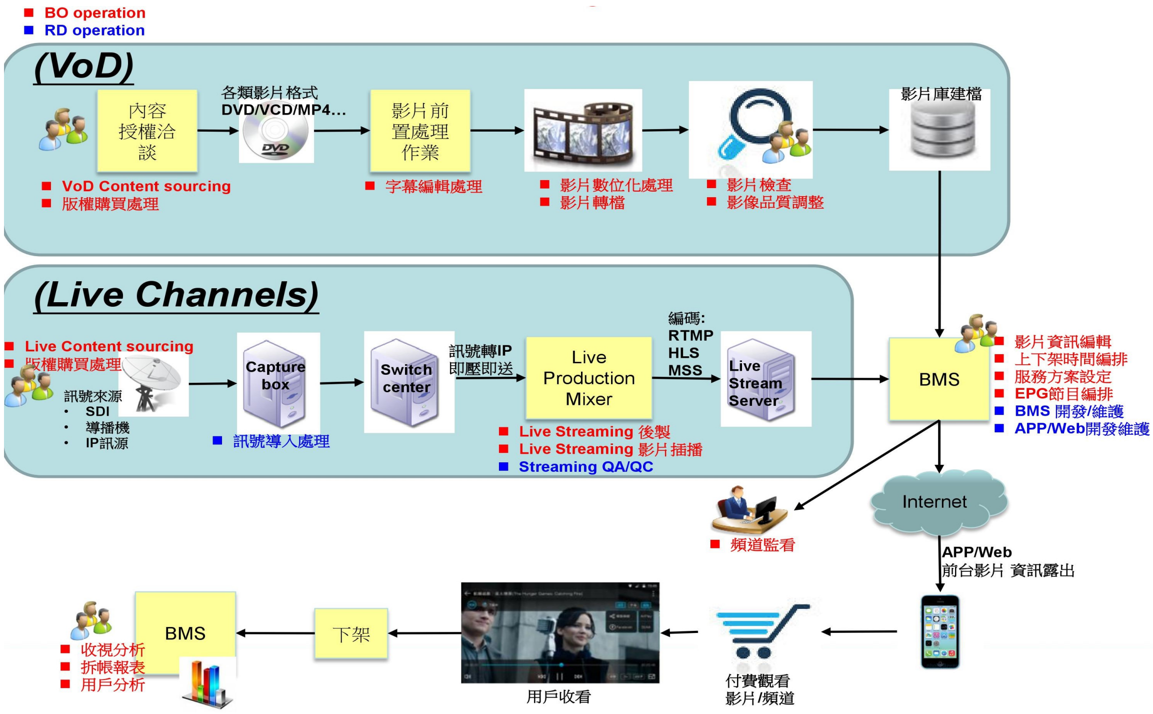


圖 4 行動通信業者提供 OTT 服務之技術架構[13]

[13]遠傳電信提供予本研究參考

第三節 國內數位匯流影音平臺服務產業現況

根據 NCC 於 2017 年匯流發展調查，臺灣 79.6%的民眾擁有智慧型手機已超過於擁有一般電視的 63.3%。電信產業、廣播電視與網際網路的數位匯流，加速了線上影音視訊服務（Video Streaming）的發展，特別是開放式網路、無 QoS 保證的線上影音服務。OTT 服務的收視增長改變了付費電視的優勢，根據 ABI Research 預估，2018 年 OTT 將達 4 億用戶，2022 年全球營收將達 514 億美元。

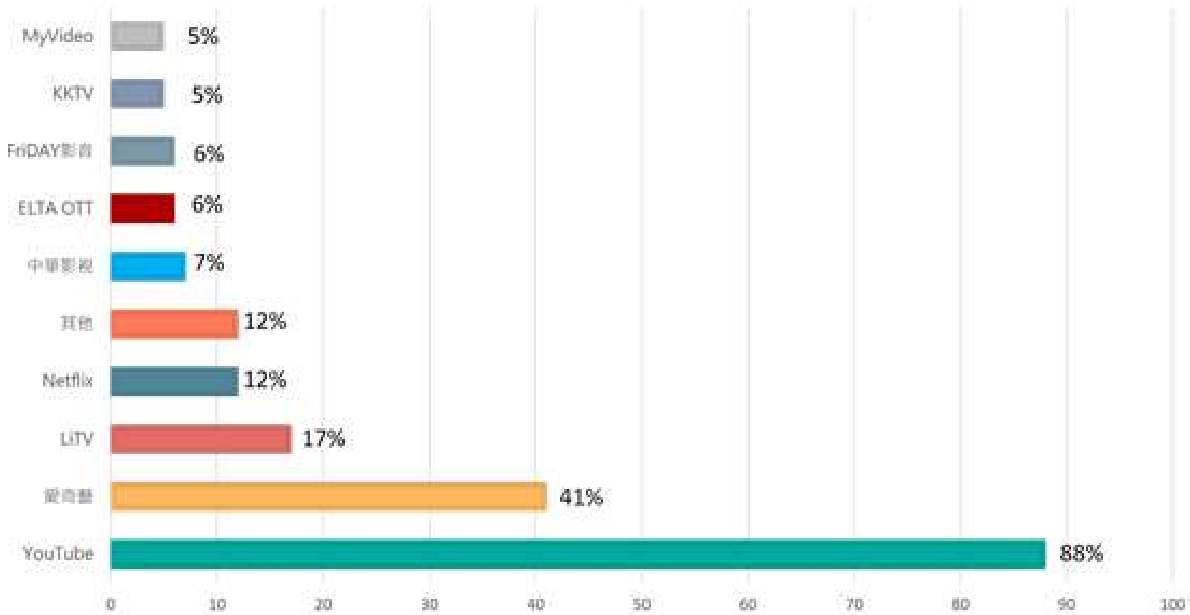


圖 5 台灣使用者平常觀看的 OTT 影音服務[14]

如上圖所示隨台灣 4G 網路日益成熟，廣電、電信與網路三路業者積極布局搶佔線上影音市場大餅，近年來除持續在台灣開台上線的新興線上影音影音平臺業者外，廣電業者因受到傳統電視廣告量下滑，積極進軍線上影音市場，而電信業者與有線電視結盟進行合縱連橫之團體戰，推出線上影音型態的隨選視訊；中華電信則是強打 MOD 服務，並強化主流頻道於 MOD 上架。以下就國內主要數位匯流影音平臺服務，逐一介紹目前的產業現況。

[14]OVO (2016/11/22)。台灣 OTT 電視使用行為調查。https://www.ovotv.com/blog/zh/2016/11/22/ottresearch/

一、 中華電信

中華電信目前除了經營原本的IPTV服務（MOD）外，亦積極發展OTT服務，於2014年推出線上影音付費服務「中華影視」，主打「電視節目隨身攜帶」，可多螢播放，亦有熱門體育賽事直播。2015年2月26日宣布成立OTT辦公室，整合MOD、行動Hami等線上影音業務。2017年7月中華電信將OTT服務「中華影視」改名為「Hami Video」[15]。2018年3月28日舉行「數位匯流事業處」揭牌儀式，宣布MOD成功突破170萬用戶，年底將可達成200萬用戶數目標。

Hami Video付費方式主要為「月租型」，並分為「電視」、「影劇」、「運動」三種方案，費用為69元至149元之間，可同時三個設備觀看。2018年3月，月租型付費會員數約為50萬。除了持續創造影視服務的變革與突破，數位匯流處將結合中華電信在物聯網、AR/VR/MR以及AI等技術能力，中華電信將引進各種發展前景的業務[16]。

二、 台灣大哥大

2012年11月台灣大哥大與國內外60家內容供應商，推出線上影音付費服務「myVideo」，其特色為「最大片庫」、「手機就是遙控器」、「與DVD同步」、「一個費用、任選影片」、「動態頻寬調整」、「無間縫視聽」、「不分網內外多平臺支援」。

myVideo提供多種付費方法，可使用信用卡或是看片金支付影片費用，看片金可在全省台灣大哥大myfone門市以現金購買儲值序號；而台灣大哥大用戶可直接使用台灣大哥大手機門號做為myVideo帳號，透過電信帳單進行付款，成為會員後能使用單一帳號登入多種不同設備，包括手機、平板、

---

[15]張家華（2017）。電信業者進入OTT市場之競爭策略研究 —以台灣大哥大myVideo為例。國立中正大學電訊傳播研究所碩士論文。

[16]中華電信（2018/3/28）。中華電信數位匯流事業處成立揭牌 強攻數位匯流龍頭，https://www.cht.com.tw/zh-tw/home/cht/messages/2018/msg-180328-162206

電腦及筆電等，不限任何電信公司用戶，皆可透過串流方式進行線上付費收看影音內容。

myVideo 付費方式分為「月租型」、「計次型」兩種，「月租型」為用戶依需求選擇月租天數，費用為 250 元至 799 元之間，在期限內可不限次數觀看標示有「月租免費看」的影片，或可進入月租館專區，任意挑選影片；「計次型」方式分「租借」或「購買單部影片」，租借後可於 7 天內開啟播放，開啟後在 48 小時內可無限次觀看，如選擇購買，可在「我的影片」裡下載影片到個人裝置中觀看。2018 年 5 月 myVideo 下載超過 350 萬，並累積付費會員 200 萬以上[17]。

台灣大和凱擘正在攜手開發 OTT 服務，雖然台灣大已有手機網路串流影音服務 myVideo，內容主要以電影為主，凱擘的優勢則是可以拿到比較多的電視內容，未來雙方合作互補優勢，以打造影、視雙棲的 OTT 服務，讓消費者看更多的影視內容[18] [19]。未來發展會繼續朝電視內建 App，以及與機上盒業者、智慧電視業者合作的方向前進[20]。

[17]台灣大哥大新聞中心（2018/5/7）。慶 APP 下載數破 350 萬 myVideo 把整個城市變成電影院。https://corp.taiwanmobile.com/press-release/news/press_20180508_767429.html
[18]江明晏（2015/8/5）。台灣大攜手凱擘 OTT 服務明年問世，Yahoo 奇摩新聞，https://tw.news.yahoo.com/台灣大攜手凱擘-ott 服務明年問世-083338994--finance.html
[19]何英煒（2016/7/27）。台灣大：Q3 獲利可望優於 Q2，中時電子報，http://www.chinatimes.com/newspapers/20160727000208-260206
[20]江明晏（2012/11/29）。OTT 潮流 台灣大攻行動影音，大紀元電子報。http://www.epochtimes.com/b5/12/11/29/n3741240.htm

三、 遠傳電信

遠傳電信優化先前的影音平臺「遠傳影城」，於 2015 年推出「friDay 影音」，以策展、直播、新聞為主，打造多元影音平臺。friDay 影音擁有超過 3,500 部的國內外電影，以及台灣、中國、韓國戲劇，並有七台全球新聞同步直播、大型賽事及頒獎典禮直播。其中在電影部分，「friDay 影音」結合自家影音內容不定期推出「策展」，歸納影片類型而推出主題影展，例如：「週末熱播」、「電影院看不到的好片」或「經典影展不看掉漆」等主題策展，方便用戶依喜好搜尋並觀賞[21]。而直播節目包括金馬影展、棒球賽事等免費頻道，其目的是客戶服務，是希望能滿足消費者在一個平臺上，對不同類型內容需求的便利性。實際上當有大型典禮或賽事直播時，能夠帶動 friDay 影音 app 下載數並增加新會員數，同時可搭配直播主題做策展，本身也是一種內容行銷[22]。

friDay 影音付款方式為「續租型」、「天數儲值」，費用為 199 元至 2388 元之間。以及部分免費觀看，只需加入會員即可觀看。2018 年 3 月 friDay 影音下載次數為 80 萬次，月租型用戶數約為 30 萬。

四、 亞太電信

亞太電信的 OTT 服務分為主打頻道的 Gt 行動電視及主打電影的 Gt 行動影城，以及 2016 年底集團推出 BANDOTT 影音服務。目前 Gt 行動電視已經累積近 30 萬付費用戶，且開始獲利，而行動影城目前用戶數約 3 萬戶[23]。有別於國內其他電信商的 OTT 服務，Gt 行動電視能在手機或平板等移動設備上使用，觀賞影片並非採取隨選隨看模式，而是由各大頻道商排定每日的節目流程，用戶可透過節目表了解喜好內容的播放時段以進行觀賞，如隨身攜帶家中電視一般。

---

[21]黃晶琳（2016/1/30）。遠傳 friDay 影音 多螢幕吸睛，聯合新聞網，http://udn.com/news/story/7240/1476029
[22]何佩珊（2016/11/22）。不靠打賞、不捧網紅，為什麼他們也要做直播，數位時代，http://www.bnext.com.tw/article/41958/the-reason-why-they-build-live-platform
[23]張家華（2017）。電信業者進入 OTT 市場之競爭策略研究 —以台灣大哥大 myVideo 為例。國立中正大學電訊傳播研究所碩士論文。

目前 Gt 行動電視可提供給網內與網外的用戶使用，亞太電信用戶只需月付 99 元即可享受服務，網外用戶則需月付 149 元。「Gt 行動影城」則推出包含台港、日韓、歐美等各類型影片，只要透過任一種行動裝置，即能享受隨選隨看的影音服務。用戶可自由訂購 Gt 影城內「影音暢看區」服務或下載「搜電影」程式，依照喜好選擇無限觀看或單次計費觀賞，Gt 行動影城無限觀看只要月付 139 元，讓用戶不出門租 DVD，也能在家使用隨選隨看的影音服務。

亞太電信推動「影音匯流」進入家庭趨勢，和鴻海旗下台灣富連網合作，於 2016 年 12 月 12 日正式引進 Android TV 機上盒，推出 BANDOTT（便當包飯）影音服務，打造網路影音平臺服務。BANDOTT 與愛奇藝、CATCHPLAY、Netflix、myVideo 等四大影音平臺策略合作，目標觀眾為國內追劇族群，而 myVideo 其實是台灣大哥大的線上影音平臺，因此台灣大與亞太電信除在 4G 網路合作外，在 OTT 領域也有合作關係[24]。

五、 新創線上影音平臺業者

根據通傳會統計，2017 年第四季全台有線電視訂閱戶相較於第三季小幅衰退 0.36%，是有線電視法公告施行以來首度呈現衰退趨勢。然而，線上影音則呈現爆發性成長，根據資誠聯合會計師事務所《2018 全球與台灣娛樂暨媒體業展望報告》指出，預估未來五年台灣線上影音影音市場將以 15.5%年複合成長率至 4.41 億美元。

除了傳統電信業者陸續開發線上影音服務，國內有越來越多的新創平臺陸續上線提供不同類型的影音內容，這些平臺有的是提供內容的整合業者，有的是自製節目的內容供應者，但因為其經營來源眾多，還是跟既有線上影音經營類型有些不同。

[24]林淑惠(2016/12/19)。新聞分析－攻 OTT 5 大電信掀眾打群架，中時電子報，http://www.chinatimes.com/newspapers/20161219000041-260202

2014 年的 LIVEhouse.in 跟一般提供電影、戲劇的平臺不盡相同，它提供較多不同類型的直播，囊括了電競、運動與新聞等的直播內容。2014 年成立的「酷瞧」，團隊幾乎都有電視製作經驗，平臺方針是以素人為主、全自製節目。平臺特別針對年輕一代族群，影視內容包含直播與各種娛樂性質的短片，也有一些原創短劇吸引年輕閱聽人目光。2015 年 TGC（替您錄）公司的 LiTV 屬於較正統的內容整合業者，靠著優惠方案與時下最受歡迎的連續劇，開始營運之後成長極快，LiTV 錢大衛董事長期望將美國 TiVo 系統商業模式帶入台灣，利用友善的節目單與使用者介面吸引閱聽人[25]。

其實國外 OTT 進入台灣並不一定會壓縮本土產業，或許是讓台灣戲劇成長的契機。以 LINE TV 為例，除了在台播出與韓星合作的韓劇之外，也積極推行在地化，與多家自產優質戲劇的台灣電視台合作，一方面獲得電視台金援能讓 LINE TV 持續拍攝自製劇並獲得更多注目；另一方面，也能讓電視台華劇透過平臺的播出與社群討論，找到年輕或特定族群的觀影喜好，進而能開發出更多商機。另外，因為台灣電視台長期迎合單一大眾族群，導致許多網路原生劇，即使兼具品質、新創議題及明星，仍難受到關注。但現在國內像是 CHOCO TV、KKTV 等平臺，成功的整合影音內容之後，目前已計畫自製優質節目，期望轉為部分內容供應商，以新穎題材鎖定適當的分眾族群，未來在線上影音平臺提供更多不同類型的影片內容。

[25]張恩齊（2017）。OTT 平臺重度使用者經驗之研究。世新大學資訊傳播研究所碩士論文。

第四節 我國數位匯流影音產業面臨轉型挑戰與侵權困境

來自美國的「Netflix」和中國出產的「愛奇藝」都在 2016 年進入台灣市場，他們有品牌知名度、串流技術和先行者的經驗，初登陸時就讓相關產業天搖地動，2016 年也因此被稱做「台灣 OTT 元年」。當時台灣本地線上影音尚在起步階段，但面臨外來 OTT 在價值創造和競爭之間對本國產業造成不公平競爭問題，包括賦稅不均、寬頻建設投資、影音服務對傳統媒體衝擊等，其中又以網路侵權為 OTT 產業最大的挑戰與困境[26]。

根據《OTT 網路電視數據報告》（OVO，2017）指出，台灣民眾最常使用的線上影音服務為 YouTube，其次為正版付費的愛奇藝、LiTV 與 friDay 影音（圖 6）；而根據另一項 i-Buzz 網路聲量調查，境外線上影音平臺的 Netflix 與愛奇藝兩者相加，甚至已占了近八成的聲量（Brain, 2018）（圖 7）。

圖 6 台灣民眾最常使用的線上影音服務[27]

[26]葉志良、何明軒（2016）。OTT 產業政策白皮書。元智大學大數據與數位匯流創新中心政策法規研究團隊。
[27]OVO（2017）。OTT 網路電視數據報告。展雋創意股份有限公司，
https://www.ovotv.com/blog/zh/2017/09/15/ovodata2017q3/

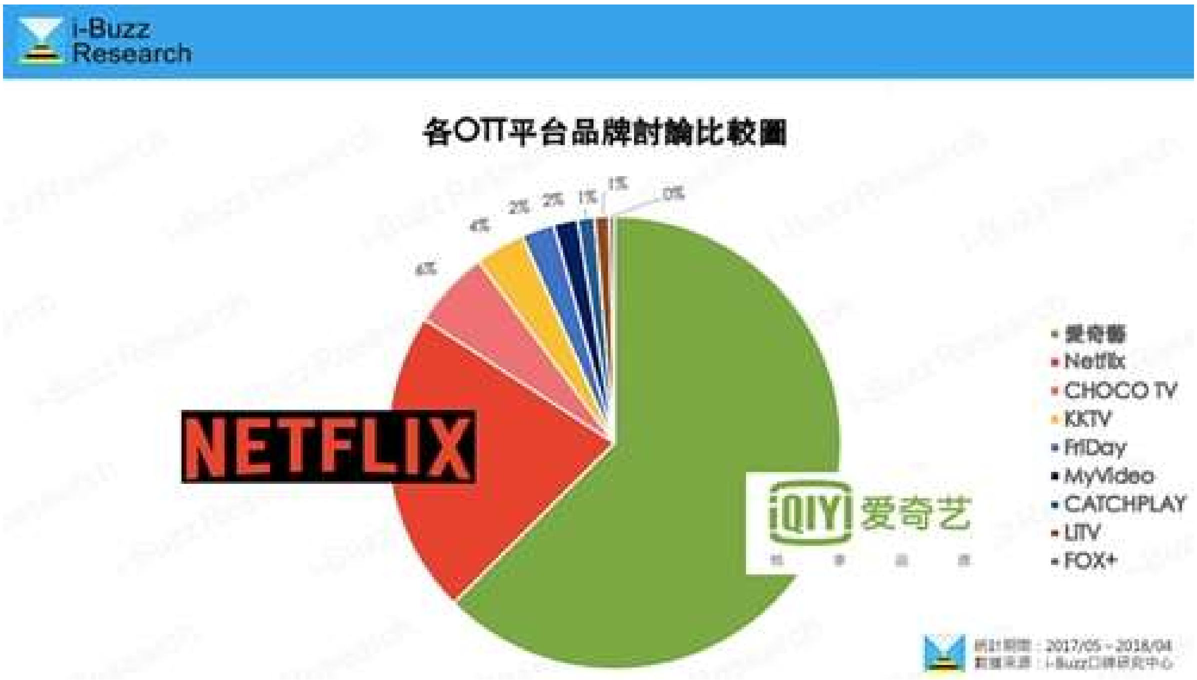


圖 7 各 OTT 平臺品牌討論比較[28]

台灣已有超過半數民眾（54.9%）使用過線上影音服務，但台灣的線上影音產業是否可像美國、韓國或中國一樣能提供足夠的數位內容供民眾觀賞？是否有獨立研發技術能力，抑或需要與外商策略合作以建立更好的商業模式[29]？在內容製作與研發技術上，台灣尚有可發展之空間，然而在產業發展策略上，實際上已嚴重落後於上述各國，不僅國外平臺業者大舉進入瓜分使用者眼球，部分境外業者甚至跨境經營，毋庸負擔國內賦稅，商業產值因此迅速流出。

[28]i-buzz (2018), “國外線上影視平臺大舉進攻 愛奇藝及 Netflix 齊撼動台灣市場”, http://www.i-buzz.com.tw/industry/article_page/?id=MTU5
[29]雲端暨聯網電視論壇（2018/7/3）。多螢媒體與匯流政策的對話—政策建言白皮書。

再者，台灣線上影音業者除採購節目外，亦努力提升平臺差異化，積極跨入產業鏈上游的影視製作，但對於所有線上影音平臺投資業者來說，侵權已成為線上影音產業最大挑戰與困境。網路侵權不僅打擊自製影片收益，間接因投資利潤無法回饋到投資者，無形中使得線上影音整體影視投資環境萎縮。是故，如何透過線上影音整體上、中游產業間共同抵制影音侵權網站廣告投放、切斷非法影音頻道收入，擴大聯合境外網站與政府機關聯手打擊侵權業者，應是健全台灣線上影音產業當務之急。台灣線上影音業者已於 2017 年 11 月 2 日組成「台灣線上影視產業協會」，其主要宗旨即在於與政府共同對抗侵權問題，營造台灣合法影視生態環境，提升民眾智慧財產權觀念。目前已蒐集十大影音侵權網站名單，提供政府相關部會參考。

台灣數位匯流影音平臺服務產業面臨強勁境外對手，如何提升本身平臺差異化，並積極製作優質內容，是所有業者共同目標；而面對另一個更強大的對手—影音侵權網站（或串流影音侵權機上盒），業者在面臨缺乏智財保護的環境之下，在產業整體發展上顯然較他國有相當的差距，因此對於侵權問題的處置上需要政府在修法與執法行動上的積極協助，始能徹底解決此一問題。

第二章 國際監理政策研析

第一節 數位匯流影音服務品質監理的政策意涵

目前有線廣播電視系統經營業者、電信事業固網業者與行動寬頻業者之經營業務，已從過去壁壘分明到今日逐漸融合，如下圖 8 所示。雖現有服務品質（Quality of Service, QoS）監理機制分別有「有線廣播電視營運計畫評鑑須知」、「固定通信業務服務品質規範」與「行動寬頻業務服務品質規範」等各種服務品質監理機制，但上述監理機制中 QoS 指標主要規範提供傳輸網路業者，尚無法直接反應消費者對於影音品質的真實感受。

考量市場競爭及數位匯流發展，跨業經營各種新興影音服務已成趨勢，對於影音服務技術現況及實際體驗品質（Quality of Experience, QoE）量測方式實有必要進行研究以確保消費者權益及作為處理相關影音服務消費爭議之依據。

英文	中文	英文	中文	英文	中文
Conventional network	傳統式網路	NGN	新世代網路	control	控制層
Wireless Access	無線存取	End Device	終端裝置	Application	應用層
Wireline Access	有線存取	Transport	傳輸層		

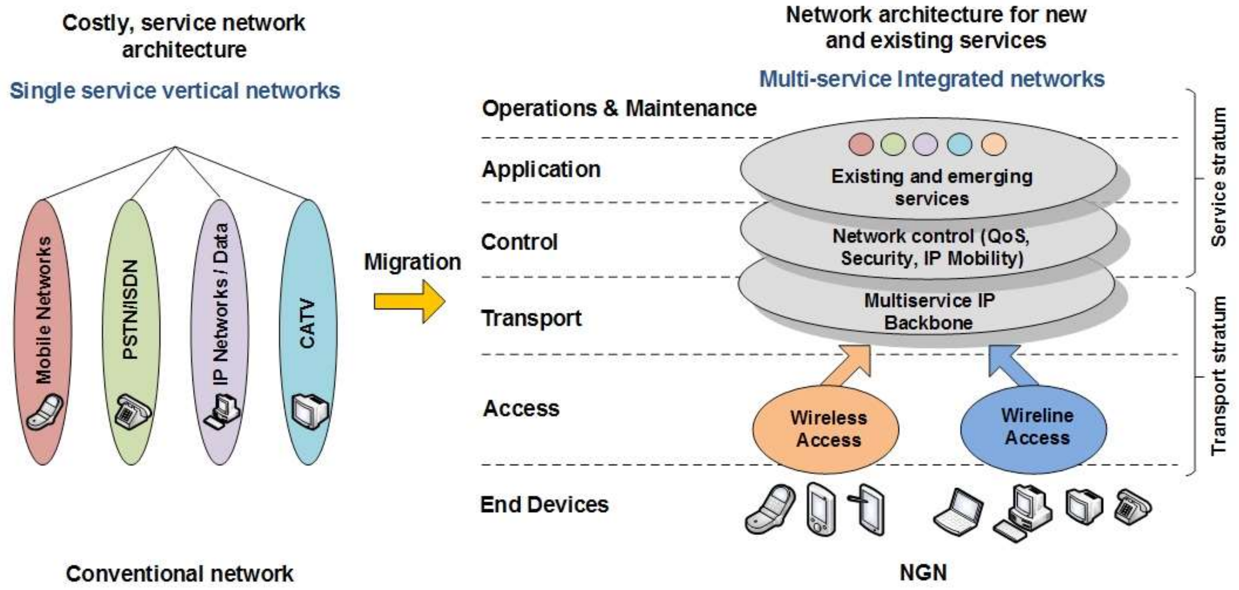


圖 8 網路產業水平與垂直架構[30]

[30]Rodríguez & Muñoz (2017). Review of Quality of Service (QoS) mechanisms over IP Multimedia Subsystem (IMS)

在數位匯流時代，閱聽人擁有多樣的行動載具得以接收線上影音內容（Video Streaming）。藉由網際網路提供數位匯流影音內容的方式，約略可分為兩大類：一是網路協定電視（Internet Protocol Television, IPTV）係由電信業者以其內建封閉式網路（walled garden）將電視節目內容傳送至訂戶收視，國內以中華電信推出之 MOD 服務為其代表[31]，其優點是信號穩定，影音品質由業者管控，缺點是目前中華電信受電信法規範而無節目編排自主權，且介面易用性常受使用者詬病；二是 OTT（Over The Top），由網路服務業者提供收視平臺，直接透過開放式公眾網路提供影音內容給使用者所使用之電子終端設備收視[32]，優點是使用者得自行選擇所欲收視之節目，但缺點則是平臺業者在網路傳輸上若無妥適安排，其常受網路壅塞情況而影響收視品質。OTT 與 IPTV 最大差異在於前者不受到特定專屬網路影響且內容多樣，而 OTT 收視方式也可透過聯網電視系統或機上盒等具備寬頻連線功能之設備，讓觀眾以電視終端得以觀看 OTT 內容，例如 Apple TV、Chromecast、XBOX 等，可擴充聯網能力，提供多樣化的網路應用服務[33]。

[31]許琦雪（2015/4/30）。OTT 競爭下有線電視產業的危機與轉機—跳脫傳統電視框架。NCC News，104 年 4 月號。
[32]葉志良（2015）。我國線上影音內容管制的再塑造：從 OTT 的發展談起。資訊社會研究，29 期，頁 47-92。
[33]DIGITIMES（2011/10/31）。Smart TV 應用環境與技術發展。
https://www.digitimes.com.tw/iot/article.asp?cat=130&id=0000256722_kg46jrky3vlzbd1o4yhe5

誠如前述，有線電視、IPTV 與 OTT 是截然不同的影音內容傳輸型態，有線電視與 IPTV 是在具有一定頻寬與 QoS 保證下傳送影音內容給訂戶收視，而 OTT 則係基於開放式網際網路作為傳輸，提供服務的傳輸品質較難維持，目前線上影音業者大多透過內容傳遞網路（Content Delivery Network, CDN）將內容傳送到消費者，以提供較為穩定的服務品質。其中 CDN 節點會在多重地點、不同網路上設置，節點之間會互相傳遞內容。在利用 CDN 服務下，可提供內容彙集及快速存取服務，對用戶的下載進行最佳化，以提高用戶的體驗品質。正因為開放式網際網路不具有 QoS 保證特性，再加上不同終端設備具有多樣性的螢幕尺寸、各用戶多樣化的使用情境等因素，如何掌握用戶的體驗品質，進而優化架構以改善品質，誠為數位匯流影音營運與管理的重要課題。

目前線上影音服務無論是國內外皆無量測標準，主要原因是線上影音服務類型眾多，各家廠商所實施或注重的服務也不盡相同，若需進行量測，是必須由廠商發布其應用程式介面(Application Programming Interface, API)或串流源供測試所需，但現階段市場生態如要執行相關量測勢必遭遇相當大的困難，雖然線上影音服務的量測準則依然處在模糊地帶，不過可以針對市售產品列舉較常使用之 KPI 做為參考，如：資源上下載速率、資源使用延遲時間、系統穩定度、影音幀速率、影音幀分辨率、影音壓縮率、網路頻寬速率、網路延遲時間、封包遺失率、網域名稱系統(Domain Name System, DNS)解析時間。

線上影音服務係基於開放的網際網路進行傳輸，使得提供服務的傳輸品質較難維持，**因此目前線上影音業者多透過內容傳遞網路（CDN）將內容傳送至消費者以提供較為穩定的服務品質**。在利用 CDN 服務下，可提供內容彙集及快速存取服務，對用戶的下載進行最佳化，以提高用戶的體驗品質（如圖 9）；再加上不同終端的使用情境，如何掌握用戶的使用情境，進而優化傳輸架構以改善服務品質，成為線上影音營運與管理的重要課題。

英文	中文	英文	中文
Fixed line access network	固定接入網	ISP	網際網路服務供應商
content distribution networks	內容傳遞網路	TLS	傳輸層安全性協定

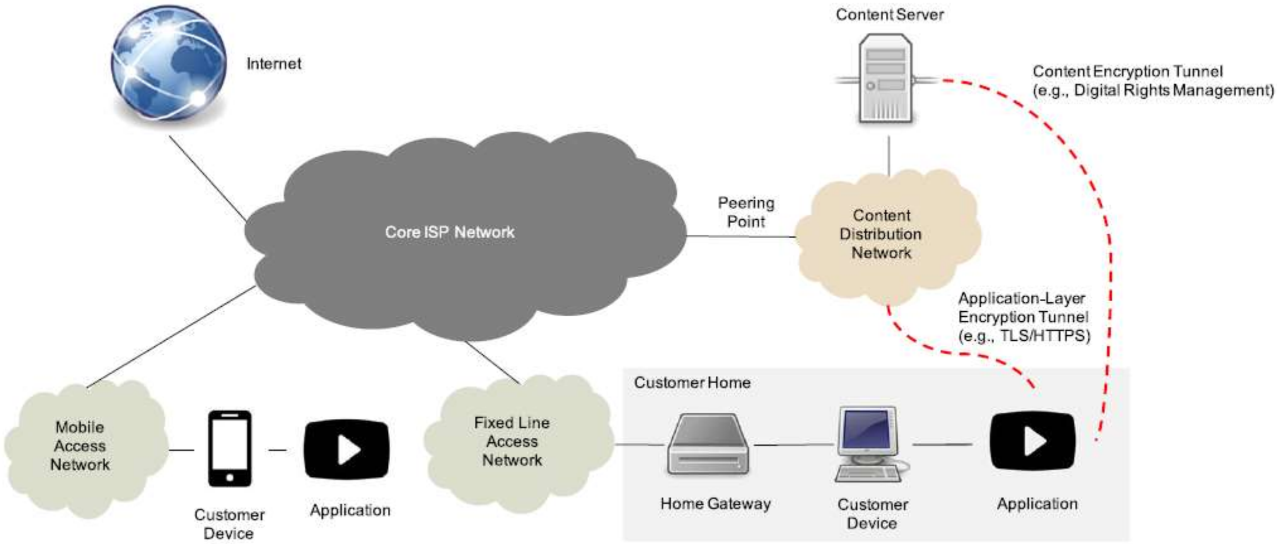


圖 9 線上影音平臺服務透過固網或行網傳遞至消費者的情形[34]

各國在視聽媒體產業的發展，從早期獨占壟斷的無線電視，到近期的有線電視、行動電視、直播衛星、錄影帶業者等擴大影音內容，再到當前透過網際網路傳輸個人化數位匯流影音內容的時代，各國監理機關基於網際網路的開放性、國家管轄權限制，以及行政管制成本等考量下，對於線上影音內容的監理也從高度規管逐漸轉變為低度管制，甚或不予管制的態度。

[34]Robitza, W., Ahmad, A., Kara, P., Atzori, L., Martini, M., Raake, A. & Sun, L. (2017). Challenges of Future Multimedia QoE Monitoring for Internet Service Providers. Multimedia Tools and Applications, 76(21), 22243-22266.

第二節 主要國家數位匯流影音服務品質監理政策研析

對於數位匯流影音服務品質，本研究除分析國際電信聯盟 ITU-T SG12 對於 QoS/QoE 相關技術進行標準化、主要 KPI 以及政策監理上的意涵外，特別針對以下主要國家，包括：美國、加拿大、英國、法國與新加坡，就數位匯流影音服務品質監理政策，分別說明之。本節對於網路中立性政策有關於數位匯流影音網路服務品質監理的影響，分析了當前美國在政策上的反覆以及歐盟在政策上所堅持的重點，做了簡要的說明。

一、 QoS 與 QoE 相關技術標準化與其政策監理意涵

對於數位匯流影音服務品質規範，目前各國幾乎不僅以純技術性的 QoS 僅衡量其傳輸面之品質，也從影響消費者對於服務的體驗品質 QoE 著手。各種媒體類型如何提升其服務之 QoS 與 QoE，學術界與業業界皆有相當多的討論（Qadir et al, 2015; Li et al. 2018）。

在傳統電話時代，國際電信聯盟電信標準化局（International Telecommunication Union, Telecommunication Standardization Sector, ITU-T）對於通話品質評估方式以及品質規劃原則的研究與標準化工作，係為確保端到端語音通信服務的品質。今日若僅基於預先設定的品質規劃，原則上是很難達到既定的 QoS 目標，這是由於電信網路所承載或提供的服務與應用類型越來越多，而且其數據流量與流向具有高度動態化特徵。從 ITU-T 第 12 研究組（Study Group 12, SG 12）近來持續對 QoS 與 QoE 相關技術的標準化進行研究，不僅包括各類服務與應用的 QoE 評估方式，也對 QoE 管理技術（包括可對每個應用的 QoE 狀況進行實時監測）進行研究（Janevski & Jankovic, 2017），並採取必要措施來解決相關問題（Baah-Acheamfuor, 2014）。按 OSI 層級化模型（如圖 10），QoE 在理解上要較 QoS 上位。

英文	中文	英文	中文
QoE Domain	體驗品質層級	Application	應用層
QoS Domain	服務品質層級	Transport/ Network	傳輸/網路層

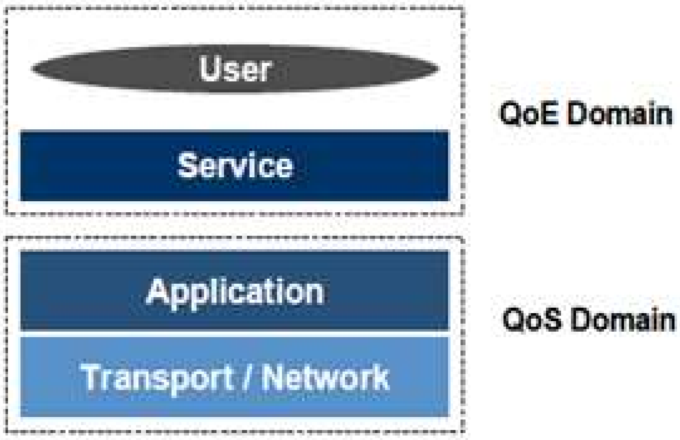


圖 10 QoE/QoS 層級化模型[35]

目前 ITU-T SG12 對於 QoS/QoE 相關技術進行標準化，其目的係為達到以下三項意義：保證用戶使用服務的便利性（protection of users' convenience）、品質衡量尺度的唯一性（uniqueness of quality scales）以及進一步提高端點對端點服務品質(achieving better end-to-end quality)(Takahashi, 2015, p.1-2)。目前 ITU-T SG12 與全球其他技術論壇以及標準化組織合作，對網頁瀏覽、線上視頻傳輸/分發、網路遊戲、視訊會議及其他應用的 QoS 與 QoE 進行研究，提供一整套的相關技術。下圖 11 是 ITU-T 對於 QoE 所採取概念檢驗層次，下層包括 QoS 與人類感知各項元素（human perception components）。

[35]Belias, V. (2013). A Study on QoE for Multimedia Systems. Bachelor Thesis on Informatics. Alexander Technological Educational Institute of Thessalniki.

英文	中文	英文	中文
factors	因素	emotion	情感
Individuality experience	個人經驗	Human perception component	人類感知成分
billing price	計費價格		

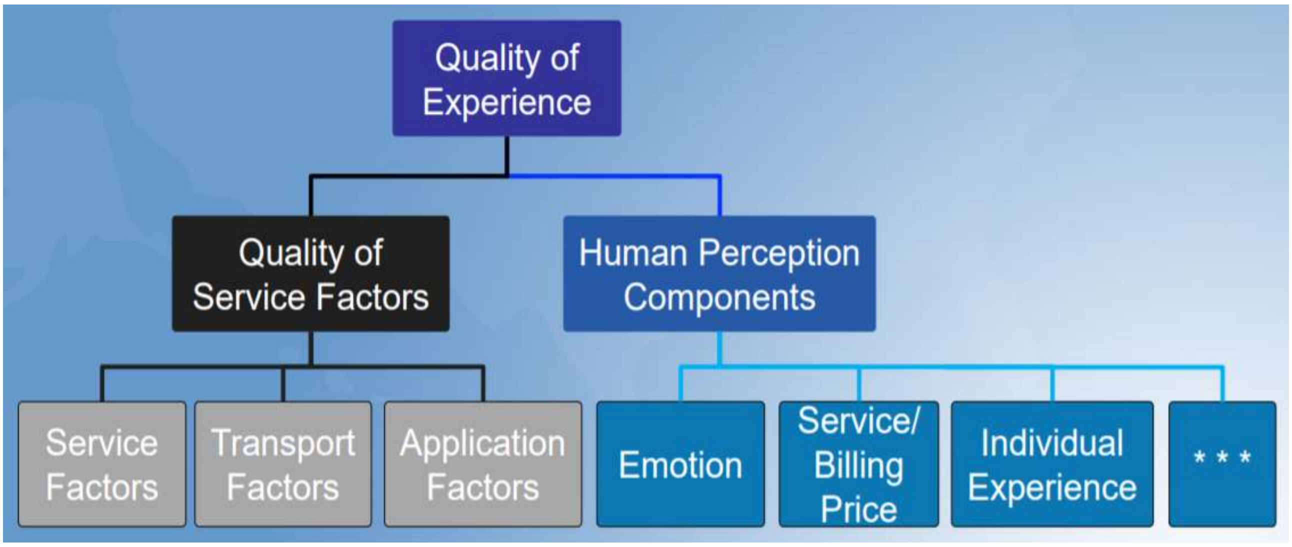


圖 11 ITU-T Approach[36]

就本計畫而言，串流視訊品質感受評估量測法 (Perceptual Evaluation of Streaming Video Quality, PEVQ-S)可視為一參考指標，原因為 PEVQ-S 是目前業界與學界衡量較客觀的基準[37]，且可相容 ITU-TJ.247 和 P.910，使其成為標準並應用於各項業界產品，其主要 KPI 項目包含：MOS 動機參數、失真指標、延遲、亮度、對比、峰值信噪比(Peak signal-to-noise ratio, PSNR)、抖動、模糊等[38]。而 PEVQ-S、J.247、P.910 皆為完整衡量影像的評測框架，J.247 主要提供了多媒體串流、有線或其他網絡中的視頻電話、有線及無線

[36]Federal Communications Commission (June 9, 2016). FCC Technological Advisory Council. Retrieved from https://transition.fcc.gov/bureaus/oet/tac/tacdocs/
meeting6916/TAC-Presentations6-9-16.pdf
[37]PTICOM, *PEVQ-S - the New Measurement Standard for Video Streaming Quality* (Feb. 2015), http://www.pevq.com/nhsei8geh98e4thi87etidowne4ihuestli8es878/SpecSheet_PEVQ-S_2015_v1-2.pdf (last visited June 30, 2018).
[38]PTICOM, *PEVQ – the Standard for Perceptual Evaluation of Video Quality*, http://www.pevq.com/pevq.html (last visited June 30, 2018).

網絡中視訊品質[39]，P.910 則提供多媒體視頻會議，遠端醫療視訊，性能水平衡量，PEVQ-S 則含蓋了以上的評定目標[40]。

ITU 顧問 Milan Jankovic 博士（2016）在歐洲電子通訊市場國際管制會議（International Regulatory Conference for Europe Regulating Electronic Communication Market）演講中強調，國家管制機關應備妥通知消費者及服務提供者有關其權利的指導準則，包括服務提供者應公告其服務的品質指標（quality parameters）以及最低服務品質（minimal QoS）相關資訊，透過零售端、網站上或資訊通路中將該訊息明示於用戶契約當中（Bennett, 2015）。下圖 12 是目前具有 QoS 服務的規管架構，從標準規範（Standards）、執照規範（License Regulation）、量測方法技術規範（KPI Measurement Techniques）、監測調查（Monitoring Survey），到最後規範的執行（Enforcement），包括規範公告、宣傳、處罰與爭議處理。

[39]ITU, Series J: Cable Networks and Transmission of Television, Sound Programme and Other Multimedia Signals (Measurement of the Quality of Service): Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference, ITU-T J.247 (08/2008), https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-J.247-200808-I!!SOFT-ZST-E&type=items (last visited June 30, 2018).

[40]ITU, P 系列：電話傳輸質量、電話裝置和本地線路網絡質量的客觀和主觀評定方法：多媒體應用的主觀性視頻質量評價方法，ITU-T P.910 (04/2008), https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.910-200804-I!!PDF-C&type=items (last visited June 30, 2018).

英文	中文	英文	中文	英文	中文
license regulation	許可證規定	monitoring survey	監測調查	publication	出版物
industry guidelines	產業準則	enforcement	強制	penalty	罰款
KPI measurement techniques	測量技術	regulatory notice	監管通知	dispute	爭議

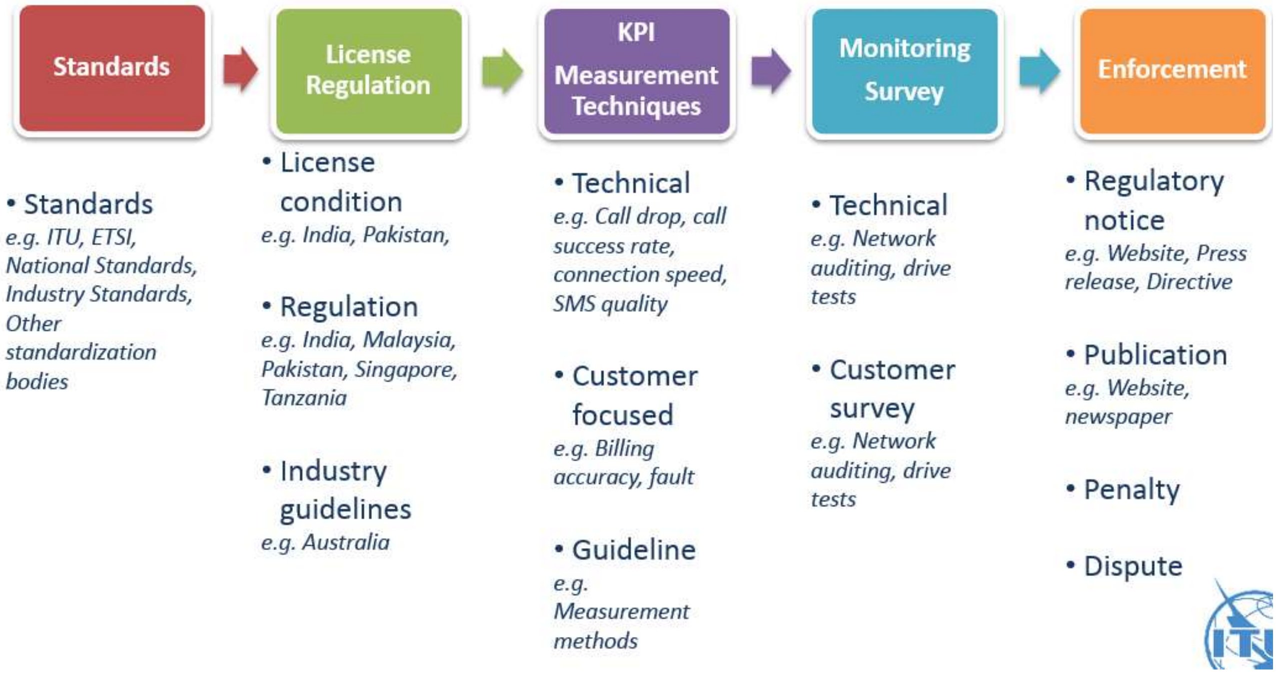


圖 12 QoS 規管架構[41]

針對線上影音服務提供商和最終用戶的視頻流，在 Shen et al.（2012）論文中，提出了一種 QoE 評估模型來預測終端用戶對考慮不同視頻內容類型的視頻流服務的感知，這種名為「視頻平均意見評分」(VMOS）模型的 QoE 模型直接關注最終用戶的感受。按此，終端用戶可以直觀地感受到的關鍵效能指標（KPI）被映射到 QoE 得分而不考慮網絡參數，VMOS 模型在視頻品質 QoE 評估方面的出色表現已通過大量的主觀平均意見評分（MOS）測試進行驗證，其中包括 180 個視頻樣本，有效投票數為 1280。VMOS 得分與 MOS 之間的 Pearson 相關係數高達 0.925，這表明該模型能夠與主觀測試幾乎相同的準確度，評估用戶對視頻品質的感知。

[41]ITU, International Regulatory Conference for Europe Regulating Electronic Communication Market, Milan JANKOVIC 2016; https://pdfs.semanticscholar.org/presentation/afb2/f1802870e134caf697947ae2c5ee22a8fda9.pdf

二、 美國

ITU 所定義的 QoS 主要是以電話系統為主，並認為「服務性能的集體效應，決定了服務用戶的滿意度」（“……the collective effect of service performance which determines the degree of satisfaction of a user of the service.”）[42]，而根據網際網路工程任務小組（Internet Engineering Task Force, IETF）提供更具體的 QoS 定義，係指「網路在傳輸數據時要滿足的服務要求。」（“……a set of service requirements to be met by the network while transporting a flow.”）（Villagra, 2017）。

典型的測量指標包括：傳輸量、延遲、抖動、位元誤碼率、可用性和封包遺失，通常在服務級別協議（Service Level Agreement, SLA）中規定。ISP 業者通常提供 1-3 層。更高的服務層通常也由 ISP 業者、應用業者，或由 ISP 業者提供給用戶 1-3 層介面。不同的應用內容對於性能因素而言有著不同的敏感度，這些因素有助於 QoE 的應用。從終端用戶應用來看，QoS 指標交易互相抵銷，並應從改善用戶體驗的情況下解釋。

目前國際組織（如 ITU）或其他產業組織（如 3GPP）對於 QoS 各有其量測方法，美國聯邦通訊委員會（Federal Communications Commission, FCC）則在 2015 年制定網路中立性管制規則（2015 Open Internet Order）中明訂五項原則，其中針對「透明性原則」制定「公開網路透明性規則要件指導原則」（Open Internet Transparency Guidance），藉以釐清業者應如何揭露網路管理相關資訊，以滿足透明性要件（詳見圖 13）。

[42]ITU-T Rec. E.800, Terms and Definitions Related to Quality of Service and Network Performance Including Dependability, 1994, revised in 2008.

英文	中文	英文	中文
commission driven metrics	委任推動指標	private initiatives	私人倡議

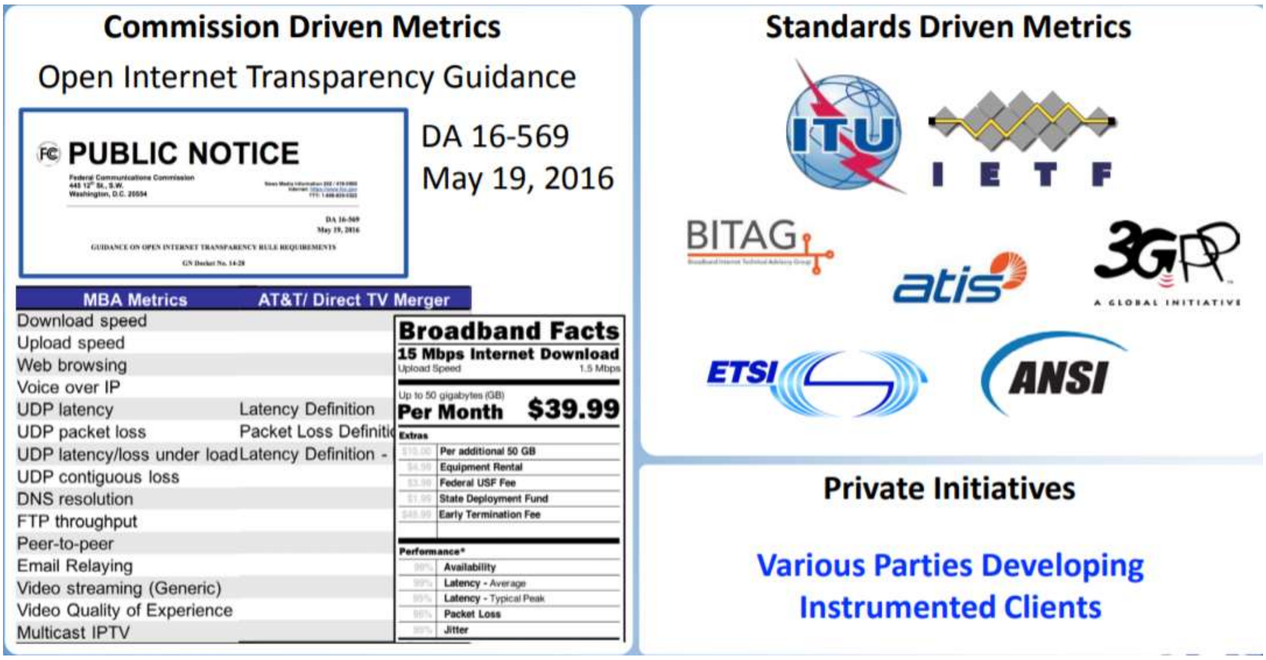


圖 13 QoS 的品質量測方法[43]

很多因素在這個主觀評價中有了一定的作用，包括通常用來測量 QoS 的傳輸量、延遲、抖動、位元誤碼率和封包遺失，而內容或應用的來源及發送路徑也會影響網路性能的感知。此外，用戶的網路、設備、設備配置、用戶介面設計、正在運行的應用程式、寬頻層及服務環境都扮演著重要角色。可靠的 QoE 測量方法需要上述數據計算「實際」的 QoE 測量，添加上述數據，可消除很多可能導致錯誤認知和測量的因素（Villagra, 2017, p.19）。

[43] Federal Communications Commission (June 9, 2016). FCC Technological Advisory Council. Retrieved from https://transition.fcc.gov/bureaus/oet/tac/tacdocs/meeting6916/TAC-Presentations6-9-16.pdf

倘若服務場域是網際網路，則與 QoS/QoE 有關之網路（流量）管理（network traffic management）則可能受到一些限制，因為各國規管機關並無全面性管控的權能（Webach, 2009）[44]，也因此如欲實施線上服務的 QoS/QoE 將更形複雜，例如：是否所有傳輸內容（資料、聲音、圖片及影像）都應一視同仁？ISP 是否應支持與其相互競爭的產品？用戶端價格應否反映在服務成本或其價值上？使用者端的 QoS 該如何衡量？改善某使用者 QoS 是否不可損及他人的 QoS？是否有方法減少服務受阻的情形？（IEEE, 2010）以下表 2 是針對網路電話（VoIP）、視訊會議（telepresence）以及一般電力系統控制（Ordinary Power System Control）之應用服務比較其 QoS 之網路管理條件：

表 2 按不同網路應用服務之網路管理條件[45]

Application	Bandwidth（Mb/s）	Acceptable Packet Loss	Target Latency（milliseconds）	Target Jitter（milliseconds）
VoIP	1-5	Up to 1%	150	50
Telepresence	8-10	Up to 0.05%	150	30
Ordinary Power System Control	Negligible	Generally much greater than 1%	2000-6000	Not applicable

美國聯邦通訊委員會（Federal Communications Commission, FCC）在 2015 年制定網路中立性管制規則（2015 Open Internet Order），明訂五項原則，其中針對「透明性原則」制定「公開網路透明性規則要件指導原則」（Guidance on Open Internet Transparency Rule Requirements，下稱指導原則，參見【附錄一】）[46]，藉以釐清業者應如何揭露網路管理相關資訊，以滿足透明性要件。**然而 2015 年網路中立性規則已於 2018 年 1 月 4 日由 FCC 撤銷，並已於同年 6 月 11 日正式生效。以下訊息僅供參考。**

[44]Webach (2009). "…we don't have a regulatory structure for that new, converged, broadband Internet infrastructure."
[45]Kostas et al (1998), Szigeti & Hattingh (2004).
[46] FCC, Guidance on Open Internet Transparency Rule Requirements, DA-16-569, GN Docket No.14-28, released on May 19, 2016.

指導原則將行動寬頻與固網寬頻區分開來，並分別要求服務等級（service tiers）。在行動方面要求業者必須揭露個別技術（例如 3G 或 4G）的寬頻性能；在固網方面則除要求揭露個別技術（例如 DSL、有線纜線、光纖或衛星）外，還需揭露服務等級（例如傳輸速率下載 50Mbps/上傳 10Mpbs）。

2015 年網路中立性管制規則要求所有寬頻業者應揭露個別服務之**預期與實際下載/上傳速率、延遲以及封包遺失**等數據，且最好能提供預期與實際的比較表格。指導原則則規定應揭露之實際網路性能指標（actual network performance matrics），在測量傳輸速率上以中間值（median speed）或特定範圍（如 25%~75%區間）為準，但若是固定寬頻 DSL 或行動寬頻則建議可採實質變動速率（substantial variation in speed）。在延遲率上則與傳輸速率類似，但必須揭露決定服務端點比率的資訊，讓消費者得以決定是否使用該服務（make informed choices）同時也讓內容應用服務開發者能開發、行銷並維繫其服務提供。在封包遺失方面則是揭露平均封包遺失率。指導原則也認為對於尖峰時段（peak usage period）必須根據本地所在時區，業者可對於如何決定網路利用尖峰時段保留適當的彈性。

在 2015 年網路中立性法規中要求 ISP 必須向其報告封包遺失率，這項規定可讓消費者了解其 ISP 所提供之網路接取服務是否符合預期，而「傳輸速度」經常被作為測量 QoE 的主要指標。但端對端供應鏈經常是複雜且難以管理，使用者經驗通常不會因其接取服務提供者之影響，而是在供應鏈當中的其他業者，而這些影響 QoE 的因素經常難以被證實。但有論者指出 FCC 僅著重在平均封包流失率，這不是很好的量測方法，因為傳輸可能性的數據經常難以顯現，因為能看到的就是封包大量丟失的情形；另外，網路管理經常面臨兩難：當你要維持低封包遺失率，換來的卻是網路傳遞遲延，造成極差消費者體驗的反差效果（Geddes, 2015）。

相同的，在網路中立性法規上僅同意 ISP 業者在「合理的網路管理」（reasonable network management）情形下得以進行差別待遇。例如在網路傳輸過程同時遇到低頻寬消耗應用（例如 VoIP）與高頻寬消耗應用（例如 BitTorrent 的 P2P 檔案交換或 Video Streaming），ISP 業者若能在此兩項應用當中調整進出 ISP 網路以及進出用戶端網路的封包傳遞時間序，這樣提升某一應用 QoE 的同時也不至於損害另一個應用的 QoE，這樣的網路管理屬於合理（Bennett, 2015）。

使用者體驗經常難以客觀衡量，特別在行動服務上，更涉及應用程式種類、手持設備、地點、實施日期以及採用服務費率之不同，皆可能影響 QoS/QoE 的結果，甚至數位匯流影音服務會根據不同的種類（例如：Skype、YouTube、Netflix）、根據解析度之不同採取的串流模式，以及根據流量或壅塞程度而就近選擇內容傳遞網路（CDN）也會有所不同。行動應用甚至會因透過 WiFi、企業大規模的 BYOD（Bring Your Own Device）、應用系統以及螢幕大小之不同，其 QoS/QoE 情境種類之多、衡量方式更形複雜（Sandvine, 2017）。從企業角度而言，其著眼點在於建構使用者之滿意程度（user satisfaction），包括技術層面的 QoS，以及使用者為中心的品質感知程度（Quality of Perception）再到 QoE（如圖 14）。

英文	中文	英文	中文	英文	中文
usability	可用性	enjoyment	享受	perceived	感知
effectiveness	效用	appeal	吸引力	engagement	需償付的款項

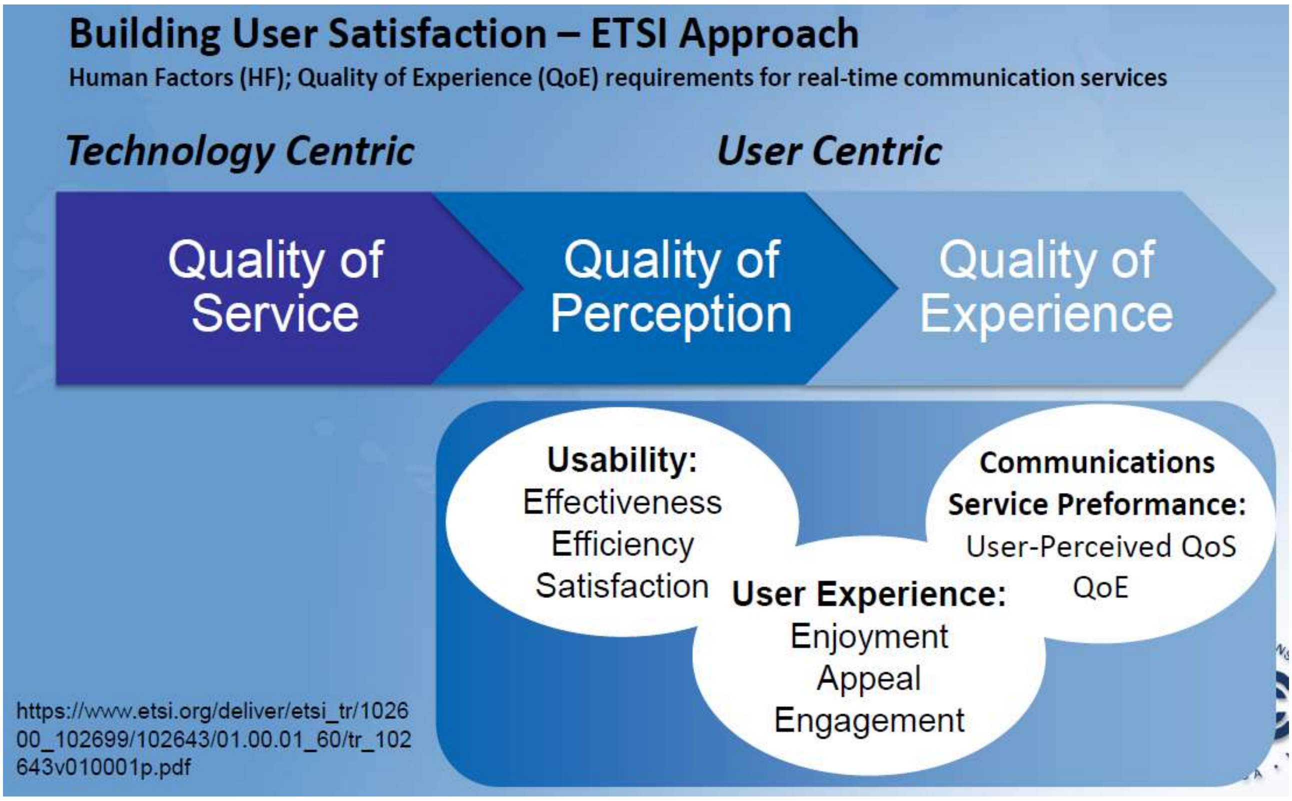


圖 14 QoS 與 QoE 不同的著眼點[47]

在不同的服務，例如瀏覽網頁、線上遊戲或特定內容串流服務，有些應用服務業者會與 ISP 業者間簽訂網路互連契約（interconnection agreements）以便讓其服務能具有較佳的成效，但這些應用業者是否「應該」支付費用給 ISP 以換取較佳的服務成效，這牽扯到後面所談的第三段有關網路中立性問題。

[47] Federal Communications Commission (June 9, 2016). FCC Technological Advisory Council. Retrieved from https://transition.fcc.gov/bureaus/oet/tac/tacdocs/meeting6916/TAC-Presentations6-9-16.pdf

三、 加拿大

根據加拿大廣播電視和電信委員會（Canadian Radio-television and Telecommunications Commission, CRTC）與大型 ISP 業者在 2016 年 3 月至 4 月之間所共同執行的一項開創全加拿大寬頻用戶社群的計畫，藉由有意願衡量其寬頻網路服務效能的用戶下載 Whiteboxes 程式以獲取資料，以了解實際上網速率以及 ISP 廣告宣傳速率的差異性。結果顯示，大多數 ISP 實際所提供寬頻服務之速率皆高於其廣告宣傳之速率，而封包遺失的情況也相當低。因此，CRTC 將其寬頻速率測試結果公告於眾，讓消費者能自行檢驗其寬頻連線程度以及費用。這項計畫結果有助於 CRTC 增進其寬頻政策的制定（ITU, 2017, p75）。2016 年 12 月 CRTC 在 Telecom Regulatory Policy 2016-496 決議中認定寬頻網路接取應被視為普及服務內容，其中固網寬頻速率應達到下行 50Mbps、上行 10Mbps 的水準，並提供五年 7.5 億加幣的政府補助以提升網路建設（Chhabra, 2018）。

2018 年 7 月 13 日 CRTC 確立了初步的基本網路服務品質規範，根據 Telecom Decision CRTC 2018-241 決議（參見【附錄二】）[48]，當網路接取服務在尖峰時段所量測的來回通訊遲延（round-trip latency）在 50 毫秒以內、封包丟失（packet loss）比率在 0.25%，即可被認定為高品質的寬頻服務；其中尖峰時段雖無清楚定義，但一般認為是非假日當地時間晚上 7 點至晚上 11 點。此項決議亦建立一套網路抖動率（Jitter, 封包延遲變異）的品質衡量方法（quality metrics）。目前這項衡量方法已公告徵求公眾意見並已於 2018 年 8 月 13 日截止（Chhabra, 2018）。

[48]Telecom Decision CRTC 2018-241, CISC Network Working Group – Non-consensus report on quality of service metrics to define high-quality fixed broadband Internet access service, https://crtc.gc.ca/eng/archive/2018/2018-241.pdf

實際上 CRTC 並未對線上提供廣播電視服務加以規管，過去曾於 2013-2014 在一項「Let's Talk TV」公眾意見諮詢（CRTC 2013-563）上討論未來電視系統的規管方向，特別是數位匯流影音線上影音，但是以 Netflix 為首的線上影音業者認為加拿大廣電法並未授權 CRTC 有管制線上影音的權力而拒絕遵守（Zboralska & Davis, 2017, p.16; Vlessing, 2014）。

CRTC 政策上有要求廣播電視業者若其內容已於受管制播放系統(例如傳統電視台）上播放時，其於非受管制平臺上播放（例如電視台的網站上）必須有字幕處理[49]。對於 QoS 及 QoE 暫時無明確的法律、法規或政策以處理數位匯流影音等服務，但對廣播電視部分，加拿大確實有簡易的規範以確保服務質量。加拿大多數政策都是為了保持提供商和許可問題之間的競爭力，例如在 Broadcasting Regulatory Policy 2014-459 決議，其宗旨是爲了管控包括視聽衆及加拿大各基金會的實質利益。其中提到的質量要求與視聽衆的利益息息相關。該決議第 2 條明確說明實質利益包括：一、觀眾或聽眾通過加拿大節目的質量與數量的增加直接受益，以及二、創作者透過加拿大節目製作、發行和推廣獲得多支持而獲益[50]。由此可見，加拿大並目前並沒有非常著重在影音服務的質量上，但是卻非常著重與觀衆，聽衆以及創作者的利益。

CRTC 在 1982 年起即要求所有聯邦管制的電信業者必須每季提交一份 16 項符合可接受程度服務（acceptable level of service）的檢驗報告，每項檢驗（indicator）內容必須符合法規標準，倘若有未符合者，業者必須向 CRTC 解釋原因並提供用戶相當之賠償。這些標準可供 CRTC 確認當服務品質發生問題時該如何因應，告知電信業者在必要時應採取矯正措施，而電信業者則透過自我申報與客訴處理管道確認自身服務品質的問題[51]。

[49]Canadian Hard of Hearing Association – Broadcast Accessibility Hub, Internet accessibility, https://chha.ca/baf/internet.php.
[50]Broadcasting Regulatory Policy CRTC 2014 -459, Simplified approach to tangible benefits and determining the value of the transaction, https://crtc.gc.ca/eng/archive/2014/2014-459.htm.
[51]Reports on Quality of Service Indicators and Reporting Letters (8660), https://crtc.gc.ca/eng/publications/reports/8660/8660.htm.

四、 英國

今日絕大多數的 ISP 對於網路管理係採取公平使用原則（fair use），亦即限制超量使用或不公平的使用。目前網路流量快速成長的部分是數位匯流影音，倘若 ISP 未能對網路流量採取適當管理，即可能面臨網路壅塞並且減損消費者體驗。因此，對於不同應用類型的流量（音訊、視訊、資料、訊息等）在網路的各個構成（接取、核心或轉訊）採取網路管理，將可確保網路的 QoS 與 QoE（ITU, 2017, p.38）。

英國通訊管理局（Ofcom）在 2011 年委託 Technologia 執行一項有關於流量管制與體驗品質（Traffic Management and Quality of Experience）的研究案（參閱【附錄三】），指出因應網路容量壓力而採取的流量管制措施，最後多會採取擴充容量方式處理，雖然並非所有 ISP 業者在擴充容量上皆面臨相同的成本結構，但可預見網路管理將在不同網路型態朝向不平等方向發展，有些 ISP 會擴充容量，但有些則會透過更多網路管理措施處理。目前的網路管理方式大多集中在避免少數超量的使用者減損大多數使用者的服務體驗，不過，網路管理由於可能會損害消費者權益而經常與網路中立性原則有所扞格，但有認為網路管理能能控制消費者的體驗並減少難以預測的網路壅塞情形。藉由將不同資料種類進行差異化處理，網路管理允許應用服務客製化處理，讓多數需要較佳 QoS 應用獲得較好的網路服務品質。因此，網路管理將會越來越講究確定性與透明性（Klein et al., 2011）。

通常網路管理本身是對網路運作的「介入」（intervention），實施網路管理需考量流量種類、服務費率、網路使用量上限或其他限制，通常有變更封包優先權（packet prioritization）或改變頻寬容量設定（bandwidth allocation）等。大多數的介入行為是透過檢測 IP 封包表頭並根據所傳輸的網路加以標示包括狀態檢查（stateful inspection or shallow inspection）、深度檢查（deep inspection）以及啟發式檢查（heuristic inspection），如下圖 15。

英文	中文	英文	中文	英文	中文
user profiles	用戶檔案	shallow inspection	淺層檢查	admission control	准入控制
packet monitoring	封包監控	heuristic	探索式	priority control	優先控制
intervention	介入	signalling	信號	bandwidth	頻寬

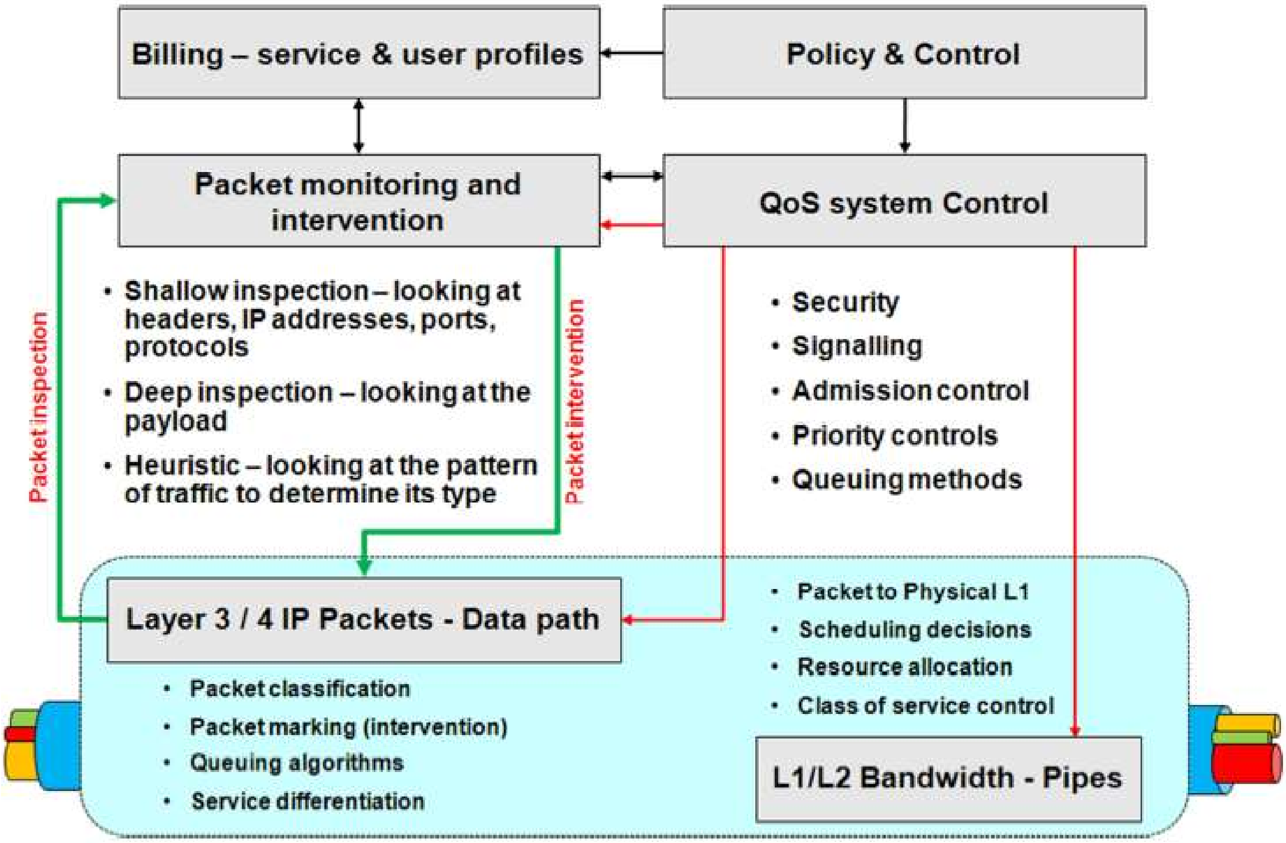


圖 15 網路流量管理架構[52]

另外，區分接取通道（partitioning of access pipes）或使用 CDN 皆會不同程度地影響網路流量，也會被視為是一種網路管理。有些時候這些議題可透過市場競爭解決，但有時候則需要政府介入管制（ITU, 2017, p.39）。

一般來說，當流量需求遠低於網路所設定之容量，則無需介入管理；當流量需求稍高但低於網路容量時，則需要簡單的網路管理，包括頻寬容量設定變更或封包優先權；但倘若流量需求已達到所設定之容量上限且產生網路壅塞的可能性遽增時，則網路管理必須積極介入以提升 QoE，如下圖 16。

[52]Klein, J., Freeman, J., Morland, R. & Revell, S. (Apr. 2011). Traffic Management and Quality of Experience. Project commissioned by Ofcom, conducted by Technologia. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0028/63955/traffic_management.pdf

英文	中文	英文	中文	英文	中文
allocation	分配	unconstrained demand	無限制的需求	prioritisation	優先

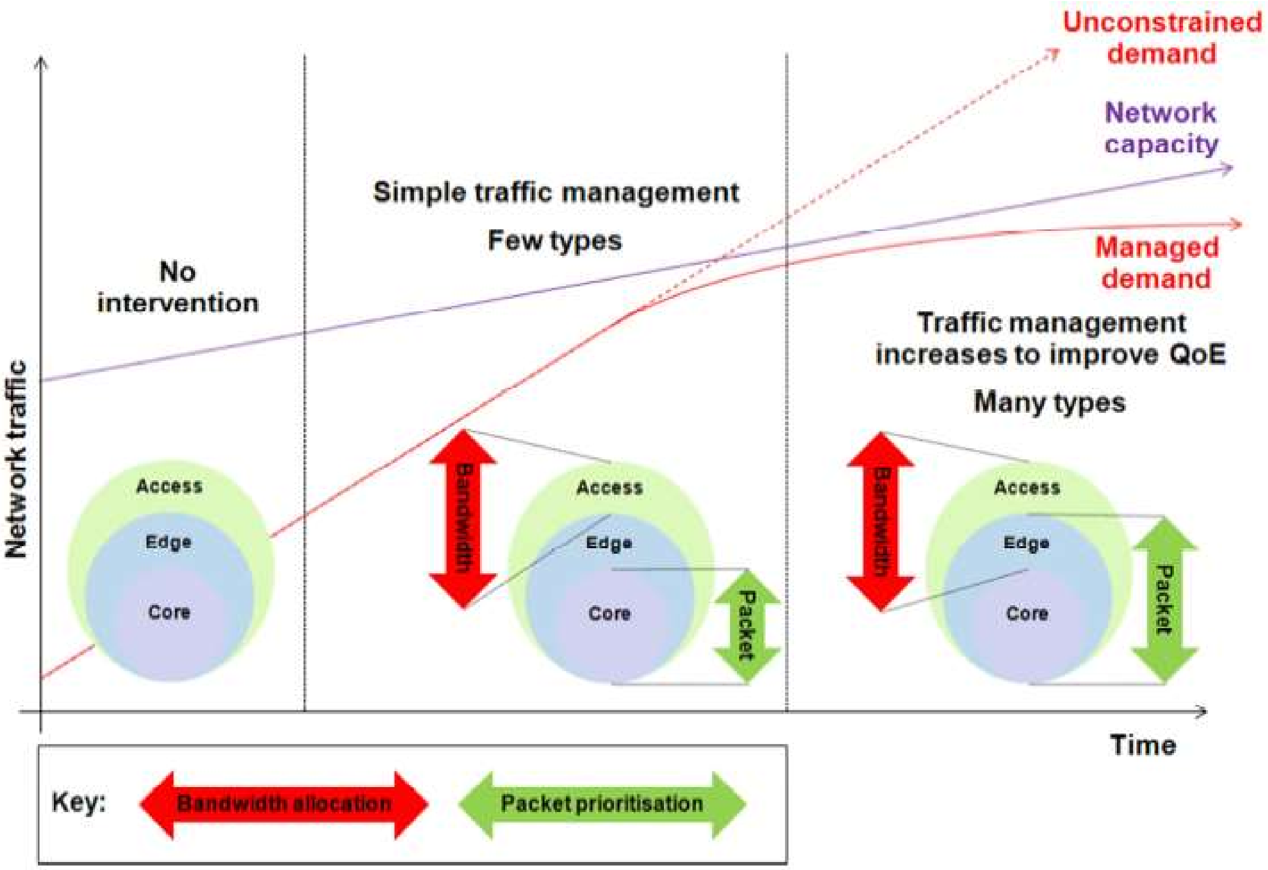


圖 16 網路管理作為網路容量調節功能[53]

2013 年 O fcom 出具官方報告「量測行動語音與數據之實際體驗品質」（Measuring mobile voice and data quality of experience），必須蒐集電信業者特定資訊以辨識競爭網路之效能、特定地理訊息、消費者實際使用情形，以及網路效能定期檢測等資訊（參閱【附錄四】）。

[53]Klein, J., Freeman, J., Morland, R. & Revell, S. (Apr. 2011). Traffic Management and Quality of Experience. Project commissioned by Ofcom, conducted by Technologia. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0028/63955/traffic_management.pdf

未來網路管理可能會因應網路壅塞情形而制定一套機制確保高品質的QoS，最大的改變可能是封包檢測會從原先於核心網路移出，且逐漸演變為讓品質更好且以使用者為中心的網路管理（ITU, 2017, p.42）。目前，英國消費者可自行上網了解其 ISP 的網路管理政策[54]。2015 年 Ofcom 分別委託 Predictable Network Solutions 一項關於網路管理檢測方式與工具之研究[55]（參閱【附錄五】），以及委託 Actual Experience 執行網路服務品質經驗調查（Investigation of Internet Quality of Experience）[56]（參閱【附錄六】），皆有提供部分網路品質量測方法，但 Ofcom 尚未採納成為具體政策。

另外根據 Ofcom 2016 年「讓通訊在所有人之間運作」（Making communications work for everyone）報告[57]指出，Ofcom 對於 Pay TV 的影音質量管控進行討論。Ofcom 係爲達成以下幾項目的：

（一）投資效益：Ofcom 正努力確保維持和加強有效投資和創新。

（二）競爭：Ofcom 正處於探索未來網路競爭政策之關鍵時刻。

（三）放鬆管制：Ofcom 正調查是否對面臨「瓶頸」的網路及服務放鬆其管制。

---

[54] Ofcom, What is Internet Traffic Management? http://webarchive.nationalarchives.gov.uk/20150106103712/http://consumers.ofcom.org.uk/internet/internet-traffic-management/

[55] Predictable Network Solutions, A Study of Traffic Management Detection Methods & Tools, prepared for Ofcom, June 2015, https://www.ofcom.org.uk/__data/assets/pdf_file/0024/71682/traffic-management-detection.pdf

[56] Actual Experience, Investigation of Internet Quality of Experience for Ofcom, July, 23, 2015, https://www.ofcom.org.uk/__data/assets/pdf_file/0030/56388/qoe-analysis.pdf

[57] Ofcom, Making communications work for everyone: Initial conclusions from the Strategic Review of Digital Communications, Feb, 25, 2016, https://www.ofcom.org.uk/__data/assets/pdf_file/0016/50416/dcr-statement.pdf

目前看來，Ofcom 規範重點在於 Pay TV 上。根據 Ofcom 的報告，對於 Pay TV 的監控包括下列項目：

（一）傳統和 Pay TV 服務的選擇、可用性和價格（包括促銷折扣）。

（二）Pay TV 平臺隨選觀看點播內容之可用性。

（三）更好地了解消費者參與程度，包括消費者對 Pay TV 服務的意識、消費者以在傳統付費電視提供商、OTT 提供商和免費觀看電視之間的轉換。

（四）Pay TV 提供商用戶和收入數據。

（五）Pay TV 提供商內容權利和批發安排的詳情。

（六）Pay TV 的技術和服務創新。

除了 OFCOM 所設定的條規以外，各大數位製作夥伴關係廣播公司包含 BBC, BT Sport, Channel 4, Channel 5, ITV, Sky, STV 和 TG4 共同達成了協議，列出了電視節目傳遞之技術規範（Digital Production Partnership Broadcasters, 2017）。該協商內包含了三大部分：第一部，圖像和聲音質量和質量控制要求；第二部，節目文件交付的附加技術要求；第三部，針對各大廣播公司特有的特定要求。

對於視頻技術要求，規範中列出的要求包含了視頻格式、信號參數、視頻陣容、視頻源頭、HD 和 UHD 採集的電影要求、後期製作、圖像縱橫比、以及檔案材料等規範。以下表 3 是基本視頻格式規範：

表 3 基本視頻格式規範[58]

	像素和比例	每秒幀數	顏色二次採樣	色彩空間
超高清（UHD）	3840 x 2160;16:9	2160p/50 or 2160p/25	at a ratio of 4:2:0 or 4:2:2	ITU-R BT.2100
高清（HD）	1920 x 1080;16:9	1080i/25	at a ratio of 4:2:2	ITU-R BT.709
標準（SD）	702 x 576;16:9	576i/25	at a ratio of 4:2:2	ITU-R BT.601

[58]Digital Production Partnership Broadcasters (2017). Technical Specification for the Delivery Of Television Programmes

對於音頻技術要求，規範中列出了對對話、響度、計量要求、立體聲音頻要求、環繞聲要求、環繞聲混音要求、杜比元數據設置，以及聲音與視覺同步等的要求規範。

五、 法國

歐洲聯盟（European Union）於 2015 年 11 月通過公開網路接取規則（Regulation（EU）2015/2120，即歐洲網路中立性規則），以確保法人與個人平等獲取網路服務，並協調跨國界規則以建立統一的歐洲市場（Holznagel & Hartmann, 2016）。根據歐洲電子通信監管機構（Body of European Regulators for Electronic Communications）所提供的有關 QoS 結構的報告，VoIP、IPTV、VoD 等服務都是屬於專業服務以及品質保證之服務（BEREC, 2011）。

專業服務和一般網路服務都有不同程度的服務保證。如今，從 Best Effort 的企業網路 VPN 到具有保證 QoS 的 IPTV 和 VoIP，專業服務擁有大範圍的服務品質技術，但開放網路服務在使用服務品質技術方面受到限制，IETF 對服務品質結構所設下的標準相信可以在未來為網際網路提供有保證的 QoS。如下圖 17。

英文	中文	英文	中文	英文	中文
facilities based service provisioning	基礎設施的服務提供	corporate	共同的	throttling	節流
enhanced traffic management	增強流量管理	facilities	設備	guarantees	保障

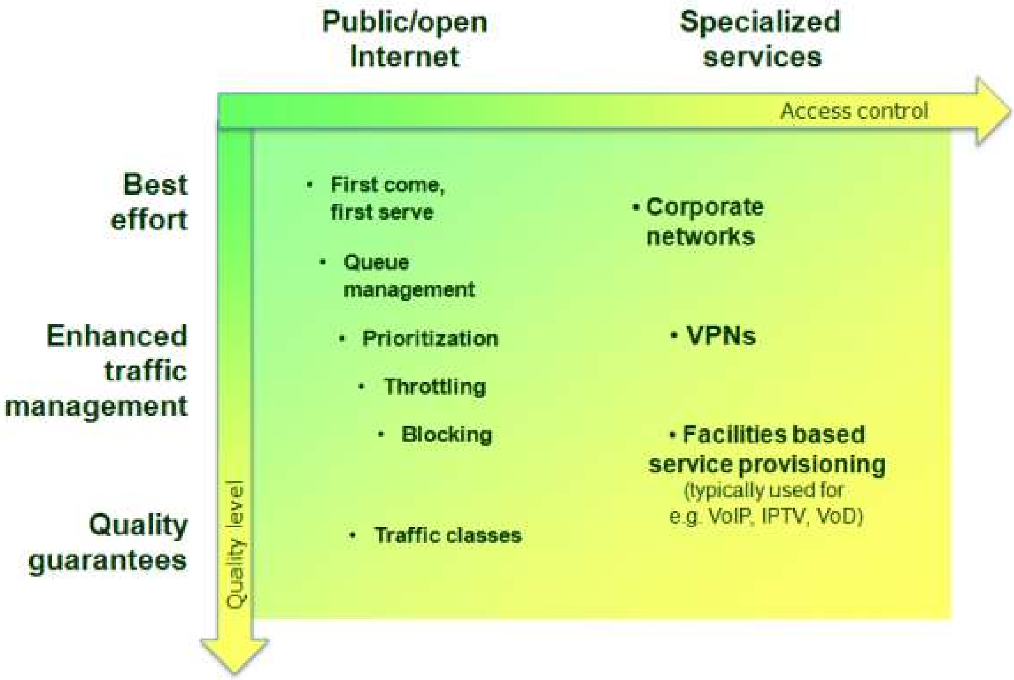


圖 17 歐盟所定義的專業服務類型[59]

關於這些體系結構能否實現，最具挑戰的是提供者之間的 IP 互連（IP Interconnection）（BEREC, 2011, p.22）。有鑑於只能為各個服務指定 QoS，包括線上服務業者控制之外的許多參數，在普及服務指令（Universal Service Directive）第 22（3）條規定應解釋為限於網路性能概念，因為所定義的 QoS 對各個 end-to-end 服務列表都有不同的 QoS 等級要求。另外，它也需要干預不受服務業者控制的區域。

因為本質上即具備服務提供品質的契約條款，所以似乎沒有必要考慮將「QoS 最低要求」（minimum QoS）應用於專業服務上。此處的「QoS 最低要求」可理解為：

（一）足夠的網際網路接取服務性能之水準。

（二）不存在選擇及/或濫用降速（throttling）或阻礙傳輸的作為。

[59]BEREC (2011). A framework for QoS in the scope of Net Neutrality.

特別當網路接取服務不正當降低其水準以增進專業服務的發展，更應有 QoS 最低要求。早在 2011 年法國通訊傳播監管機關 Arcep 即認為，為避免服務水準因網路傳輸問題而下降，主管機關可以設定 minimum QoS 標準（Arcep, 2017, p.60）。

2017 年 Arcep 公告「法國網路現況報告」(The State of Internet in France)，其中第 2 章提到量測與公告網路接取服務品質指標已成為各國監理機關的主要作為。由於市場已呈現競爭，使用者可以透過 Arcep 所公告之資料對業者產生壓力（Arcep, 2017, p.10）。除了接取服務價格的考量外，提供給用戶的服務品質即成為選擇其服務與否的重點，因此用戶是否能得知其接取服務品質的清楚資訊甚為重要，例如一般家中使用 Wi-Fi 品質必須清楚明示，或者 IPTV 服務測試結果也不應忽視，而消費者不清楚 IP 互連因涉及多重利害關係人，也應提出說明。但通常消費者也會因為過多資訊而迷失焦點，因此在技術層面之外，應當將資訊簡化讓消費者易於瞭解（Arcep, 2017, p.27）(參閱【附錄七】)。

2018 年新的「法國網路現況報告」指出，過去由 GESTE 實驗室（由內容與服務提供者組成的產業團體）量測的網路服務品質因所量測網路種類、量測方式、地點等諸多原因導致結果並不理想，也因此如何創立由實際使用者所進行的量測工具（crowdsourcing tools）以便能確認各種連線技術，藉由社群力量將量測結果更為準確，已成為目前 Arcep 的政策目標（Arcep, 2018, p.10, 12）。目前 Arcep 正擬定行動準則（Code of Conduct）以便引進新的量測工具並設定最佳典範（best practices），將有助於獲取值得信賴的量測方法如下圖 18（Arcep, 2018, p.14）(參閱【附錄八】)。

英文	中文	英文	中文	英文	中文
co-construction	共建	characterise	表徵	establish transparency criteria	建立透明度標準
controlled	受控	combatting fraud	打擊欺詐	statistical representativeness	統計代表性
comparison	對照				

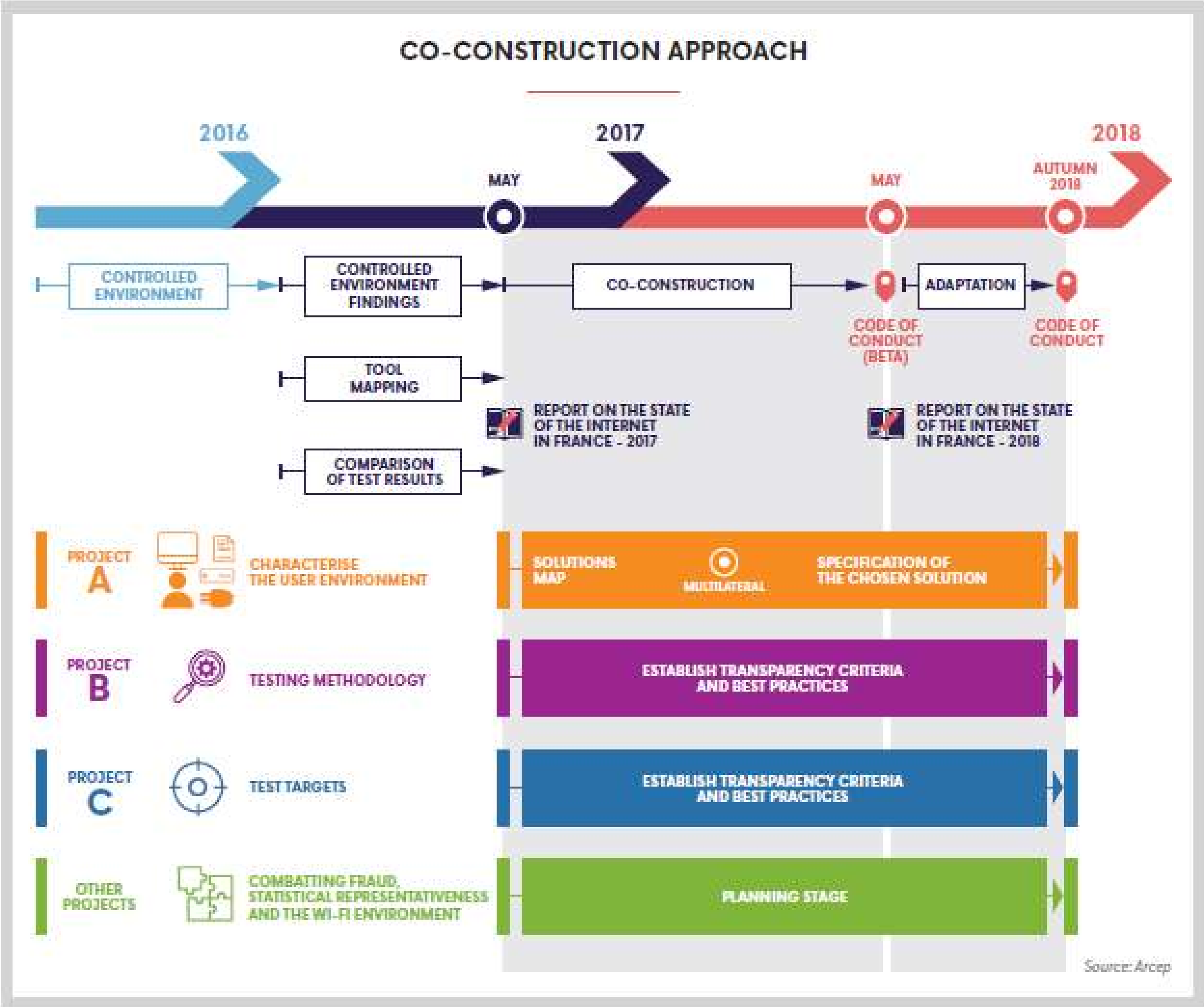


圖 18 Arcep 催生網路量測行動準則[60]

[60]Electronic Communications and Postal Regulatory Authority (Arcep)(June 2018). The State of Internet in France. France.

六、新加坡的標準實務準則（Code of Practice）

新加坡影音服務的 QoS 測量方式，資通訊與媒體發展局（Infocomm Media Development Authority, IMDA）（參閱【附錄九】）已於 2015 年 5 月 4 日制定廣播電視標準實務守則（Code of Practice for Television Broadcast Standards）。此守則目的在確保新加坡全國之許可電視持有者能達到高標準技術質量，以及廣播電視服務的可靠性要求[61]。

實務守則中所稱服務區塊包括了廣播電視服務、有線電視服務以及線上影音平臺。其中第四章說明了影音服務品質標準。規定如下：

執照持有者具有遵守指定需求的義務

（一）提供無線廣播電視服務、有線電視服務或 IPTV 服務者需確保

1. 依 ITU-R BT.500 所描述，「直播」節目的圖像和音頻應達到 ITU-R-5-Points 質量等級量表之等級五。
2. 依 ITU-R BT.500 所描述，錄製節目的圖像和音頻應達到 ITU-R-5-Points 質量等級量表之等級四。

（二）持照者應確保所傳送節目的圖像和音頻是同步的

依照守則的第四章第二節第三項，所有的執照持有者有義務遵守的條規如下：

1. 持照者應向 IMDA 提供其節目的書面說明，以確保其服務達到所需的圖像和音頻質量標準。
2. 持照者應對其電視服務的技術質量進行例行評估。
3. 持照者應儘速處理並提供適當的途徑，以便解決觀眾投訴或與圖像和音頻質量有關之回應。

[61] Broadcasting Act (Chapter 28), Code of Practice for Television Broadcast Standards, https://www.imda.gov.sg/-/media/imda/files/regulation-licensing-and-consultations/codes-of-practice-and-guidelines/acts-codes/07-annex-b--code-of-practice-for-television-broadcast-standards.pdf?la=en

4. 持照者應按照 IMDA 規定的格式向 IMDA 提交季度報告，以便在即墨前一個月內交上過去三個月內所收到的與圖像和音頻質量相關的投訴。如果季內未收到任何投訴，持照者仍應提交季度報告並相應註明。
5. 持照者應錄製其服務廣播的所有節目，並自節目播出之日起保留其錄影爲期四週。
6. 當觀眾提出與圖片和音頻質量有關的投訴時，IMDA 可自行決定對持照者所提供的服務展開調查。進行調查之時，持照者應按照 IMDA 要求無條件提供有關節目或頻道的錄製。這些記錄應在進行所有編碼之後構成可供傳輸的資訊。

實務守則在第四章第二節第四項記載說，對於新聞插頁、現狀或過往資訊而言，畫質較差的圖片和音頻質量可被允許，但前提為其不可進一步再提高技術品質或該低品質屬於節目編輯目的之一部。

七、 小結

以上各國政府對於網路 QoS 有一定的規範準則（例如法國已有執行網路服務品質量測但效果不彰而有待日後採取「由上而下」方式調整），對於線上影音平臺服務品質量測目前皆無具體的法規政策支持，**本研究認為可能是因為線上影音平臺服務在固定與行動寬頻技術仍在持續發展、線上服務內容多樣以及網路接取生態複雜等各種因素，使得各國並未針對線上影音平臺服務制定具體的服務品質量測標準。**以下謹將前述美國、加拿大、英國、法國與新加坡等國的法規政策，整理如下表 4。

表 4 各國線上影音服務品質量測政策法規比較

國家	政策法規或措施
美國	1. 並無針對線上影音服務品質制定量測規範，但 2015 年 FCC 制定網路中立性管制規則的「透明性原則」有要求業者應揭露網路管理相關資訊。但該項規則已於 2018 年遭 FCC 決議撤銷並生效。 2. 2015 年網路中立性規則要求所有寬頻業者應揭露個別服務之預期與實際下載/上傳速率、延遲以及封包遺失等數據。
加拿大	1. CRTC 有要求受管制廣播電視節目部分應有字幕處理，並確保一定的服務質量（CRTC 2014-459 決議），但對於線上影音平臺的 QoS/QoE 暫無明確法律或其他規範。 2. 2018 年 7 月 13 日 CRTC 2018-241 決議，透過封包延遲變異比率的品質衡量方法，當網路接取服務在尖峰時段所量測的來回通訊遲延（round-trip latency）在 50 毫秒以內、封包遺失比率在 0.25%，即可被認定為高品質的寬頻服務。這項衡量方法目前僅止於公告徵求公眾意見階段，尚未形成具體法規。
英國	1. Ofcom 僅對於 Pay TV 的影音質量有管控規範，並無針對線上影音平臺服務品質制定具體法律或其他規範。 2. ISP 對網路管理採取公平使用原則（fair use），可限制超量使用或不公平使用網路行為，確保網路 QoS。消費者可以自行上網了解 ISP 的網路管理政策。
法國	1. 2011 年 Arcep 認為為避免服務水準因網路傳輸問題而下降，可以設定 QoS 最低要求標準，包括足夠的網路接取服務性能水準以及業者不得有濫用降速或阻礙傳輸等行為。 2. 2017 年 Arcep 報告認為監理機關應重視網路服務品質量測，讓用戶易於瞭解服務品質資訊；但 2018 年報告指出因量測的網路種類、方式、地點等諸多原因導致結果並不理想，認為應可採取

	「由下而上」（bottom up approach）方式藉由實際使用者自採量測工具確認各種連線技術，其結果將更為準確。目前仍在制定行動準則（Code of Conduct）以獲取值得信賴的量測方法。
新加坡	1. 並無針對線上影音平臺服務品質制定量測規範。 2. IMDA 於 2015 年 5 月制定廣播電視標準實務準則（Code of Practice）確保業者服務的技術質量與可靠性。但此規範僅及於無線電視、有線電視以及 IPTV，並不及於線上影音平臺服務。

(資料來源:本研究整理)

第三節 網路中立性對數位匯流影音服務品質監理之影響

網路中立性的規範在已開發或開發中國家當中被認為屬於網路 QoS 的一環。各國看待網路中立性，所採取的觀點大不相同，例如印度於 2016 年制定反價格歧視的規定、巴西則早在 2014 年制定「網際網路公民權利法案」(Marco Civil da Internet）直接將網路中立性加以明文化。目前全球並無明確獲最佳的方案因應網路中立性，各國根據其內國環境所採取的態度各有不同，約略可分為三種：第一類是採取謹慎觀察態度，這類國家並未採取任何特定措施以因應網路中立性，認為既有規範即為已足；第二類是採取低度管制態度(light-handed approach)，例如資訊揭露與透明性原則、降低轉換障礙、最低服務品質（minimum QoS）而與既有規範之間有所微調，但不至於禁止特定行為；第三類則是採取特定管制措施禁止 ISP 執行特定行為，通常是根據合理網路管理實務作為(ITU, 2017, p.103-104)。以下本研究從美國與歐盟的規管制度經歷，進行說明。

一、 美國網路中立性規範

2010 年 FCC 網路中立性管制規則

在 2010 年 Comcast v. FCC 案（Comcast Corp. v. FCC, 600 F.3d 642（2010））聯邦巡迴上訴法院否決 FCC 得禁止 Comcast 公司阻止其客戶使用點對點網路應用程式，認為 FCC 並無所謂的附屬管轄權（ancillary jurisdiction），而 1934 年通訊法第二章 Title II 之管轄權又僅限於電信事業，無法擴張到寬頻服務（資訊服務）。隨後 FCC 於 2010 年 12 月公布維護開放網際網路（Preserving the Open Internet）管制規則。該法規歸納出以下三項措施（葉志良，2016，p.168-169）：

（一） 透明性（Transparency）：要求固網與行動寬頻業者必須揭露合理的網路管理資訊、網路性能表現與服務條款。在資訊內容揭露上，包括需公開網路管理政策、相關費率與限制條款等資訊，但對於安全管理或營業秘密則可不予公開；所需揭露的具體資訊類型，如網路壅塞之管理、對於網路接取設備或應用程式的限制、網路安全要求等。

（二） 禁止封鎖（No Blocking）：除基於合理的網路管理（reasonable network management）以外，固網寬頻業者不得封鎖合法的內容、應用、服務與對網路本身無害的設備；行動寬頻業者亦不得封鎖合法網站或阻礙與其語音或影像電話具有競爭性的應用或服務。其中「合理的網路管理」，定義為「寬頻服務業者考量特定網路架構與寬頻接取技術，對其網路進行適當的管理，則有助於達成合法網路管理目的。」FCC 對於「合理的網路管理」要求寬頻業者所實施之合理作為，僅能包括：減少或減輕網路壅塞的影響，或處理服務品質的問題；處理對用戶有害或不需要的流量；防止非法內容的傳輸；防止以非法方式傳輸內容。

（三） 禁止不合理差別待遇（No Unreasonable Discrimination）：規範僅固網寬頻業者不得以不合理方式對用戶之網路傳輸流量有差別待遇，且合理的網路管理並不構成不合理差別待遇。FCC 表示該原則可促成限制有害行為以及允許有利的差異處理之間的適當平衡。雖業者依據合理的網路管理即符合「合理」的差別待遇，但 FCC 也要求業者必須揭露管理作為（transparency）、賦予終端使用者有一定的控制權（end-user control）、可給予與使用或應用上無關的差別待遇（use-agnostic discrimination or application-agnostic discrimination）、標準作業程序（standard practices）等。

2015 年 FCC 網路中立性管制規則

2014 年巡迴上訴法院宣判廢棄 2010 年網路中立性管制規則中禁止封鎖與禁止差別待遇兩項措施，但仍維持透明性。為使管轄權問題獲得一次性解決，FCC 決定制定新規定，將寬頻服務拉回 Title II 管理。2015 年 2 月 26 日 FCC 公告新版「保護與促進開放的網際網路」（Protecting and Promoting the Open Internet Order）管制規則，將網路接取業者（IAP）歸屬為通訊法 Title II 所規範的電信事業，原則上電信業者不得對網路傳輸內容有不合理的差別待遇行為。新法規當中前四項原則屬於修訂與新增，第五項原則因未被宣告無效，故仍繼續維持：

（一） 禁止封鎖（No Blocking）：本條規範為不得任意封鎖使用者、網路服務業者合法接取網路的權利。

（二） 禁止降速（No Throttling）：此項原則原本為「禁止不合理差別待遇」。此次修訂為不得任意降低合法使用者的網路速度。

（三） 禁止付費取得優先傳輸之權利（No Paid Prioritization）：本項為 2015 年新增，意指 IAP 業者不得以付費優惠其用戶，使其享有「快車道」（fast lane）的差別待遇，可免於網路壅塞。

（四） 禁止針對網路行為有不合理的干擾及降低標準（No unreasonable interference or unreasonable disadvantage standard for Internet conduct.）：本項規範任何從事提供寬頻網路接取的業者，不得不合理干擾或歧視標準；歧視標準的意義在於不得干擾終端用戶選擇連接寬頻網路的服務業者以及設備；或者使用合法的網路內容及應用程序，服務或設備的權利。而合理的網路管理不被視為違反本規則。所謂「合理的網路管理」，依據規範定義「網路管理是一種以技術面的理由進行的網路管理，不包括其他商業上的理由。如果是基於合法的網路管理目的，並考量寬頻網路服務的網路架構和技術來認定網路管理之合理性。」

（五） 透明度原則（Transparency）：此項原則為 2010 年即存在之條款，並未被法院宣告無效，故仍繼續維持。其內容為要求網路服務業者必須公開其網路管理政策及相關資訊，包含管理參數與網路服務效能等。

2016 年 USTA 判決確認 2015 年 FCC 法規

2015 年網路中立性法規再次於哥倫比亞特區聯邦巡迴上訴法院受到電信業者的挑戰。2016 年由與 2014 年 Verizon 案同樣的多數法官，做成 USTA v. FCC 本案判決。雖巡迴上訴法院支持 FCC 網路中立性法規將寬頻服務納入電信服務 Title II 管理，但網路中立性議題的爭議卻未停歇。

在 USTA 判決中，法院認定 FCC 在 2015 年網路中立性管制規則將寬頻業者定位為電信業者，並將寬頻業者之間或與其他網路間之互連亦涵蓋在電信服務的定義中是可以的，除非 FCC 會恣意濫用裁量權或其他不符合法規之情事，否則法院基本上會尊重 FCC 對於事實的認定與其對政策決定所擬定之審查基準。

比較有趣的分析在於「禁止付費優先權」議題上。少數見解 Williams 法官認為，在電信管制上 FCC 早已承認合理的費率差別待遇，僅不正當或不合理的差別待遇在禁止之列，禁止合理的費率差別待遇是極為不尋常的。同時 Williams 法官亦質疑禁止付費優先權但卻不管制快取服務（caching services）與內容傳遞網路（Content Delivery Networks）的理由何在，以及在未執行市場調查的前提下關於付費優先權與創新正向循環兩者之間如何認定其因果關係[62]。Williams 法官不僅質疑管制的好處，更認為付費優先權會對有效利用網路資源與寬頻建設帶來更多的好處，例如會產生寬頻網路投資意願、改善使用者經驗並增加需求。最後 Williams 法官認為，即使 FCC 主張都是正確的情況下，對寬頻網路的最低品質要求（minimum QoS）是可解決付費優先權所宣稱缺點的另一可能選項[63]。

2017 年 FCC「恢復網路自由」廢除網路中立性規範

2017 年 1 月共和黨籍川普總統就任，旋即表態會廢除過去歐巴馬總統所主張的網路中立性管制規則。新任 FCC 主席 Ajit Pai 過去也曾於 2012 年受歐巴馬總統任命為 FCC 委員，當初大力反對 2015 年網路中立性管制規則，其就任後旋於 4 月 26 日正式展開廢除網路中立性規範。2017 年 12 月 FCC 表決通過廢除 2015 年網路中立性管制，並於其後 2018 年通過的「恢復網路自由」（Restoring Internet Freedom）規則中，除正式廢除三項具體明確的規則以及禁止寬頻網路服務提供者不合理的妨礙或不利於消費者或應用服務提供者的一般規定外，更明確其聯邦法規優於州法（Preemption）的立場，意謂各州不得自行制定與「恢復網路自由」規則不一致的網路中立性規範。此作法引起各州反彈，超過二十州對於 FCC「恢復網路自由」規則提起訴訟。

---

[62]所謂創新正向循環是指網路創新會提高需求，增加寬頻建設的投資，並激發新的創新。Williams 法官檢視本案證據，並未發現付費優先會危及寬頻建設的相關證明。

[63]對寬頻業者的最低品質要求，可以解決 ICP 對付費優先權所可能帶來對非付費優先其他業者所為劣質服務的憂心。

二、 歐盟網路中立性規範

2015 年以前

2009 年歐盟執委會通過電信改革法案（Telecom Package），其中「精進管制指令」（Better Regulation Directive 2009/140/EC）中的「網路中立性宣言」（Declaration on Net Neutrality）強調增進消費者資訊透明度並對業者不當網路流量管理賦予管制權能，維持網路開放性與中立性；另在「公民權利指令」（Citizens' Rights Directive 2009/136/EC）落實加強透明度的要求、建立管制機關管制權能等規範。綜言之，2009 年改革法案在網路中立性議題上，強調「增強資訊透明」與「服務品質監督」兩個面向，由於網路中立性牽涉不僅是 ISP 網路管理而已，更涉及消費者網路傳輸資訊監測、隱私權與個資保護等問題。其實歐盟的網路中立性規則和美國 FCC 所提出的「合理的網路管理」有相當類似的規範內容，強調業者必須以明確方式通知終端使用者關於管理網路的各種措施。（葉志良，2015，p.181-182）。

2015 年以後

歐洲議會於 2015 年 11 月通過「開放網路接取規章」（Regulation（EU）2015/2120），其中網路中立性規範已於 2016 年 4 月 30 日正式實施。開放網路接取規章通過後，對於歐盟網路中立性政策走向以及歐盟「數位單一市場」（Digital Single Market）而言，無非是重要的里程碑。2016 年 8 月 BEREC 制定網路中立性各國主管機關執行準則（Guidelines on the Implementation by National Regulators of European Net Neutrality Rules）[64]，可讓各國對於網路中立性之執法有所依據。其主要內容包括：

[64]BEREC (2016), Guidelines on the Implementation by National Regulators of European Net Neutrality Rules, para. 8-9, http://berec.europa.eu/eng/document_register/subject_matter/berec/download/0/6160-berec-guidelines-on-the-implementation-b_0.pdf

（一）保障使用者權利：使用者有權接取和傳送資訊與內容、使用或提供應用及服務，並使用自己選擇的終端設備。

（二）網路流量管制：ISP 業者可以執行合理網路管理措施，但是否「合理」，必須是該措施應透明公開、非歧視性且符合比例原則，且不應基於商業考量而是客觀判斷後根據特定流量類別提供不同的流量管理（服務品質）要件。另外各會員國在檢討流量管理是否具有適當性時，必須考慮以下情形：

1. 當該項流量管理機制在實施後得實現某項合法目標時，則該項流量管理機制具備適當性。合法目標包含促進網路資源有效利用或是提升整體網路傳輸品質等。
2. 當該項流量管理機制在實施後有充分證據顯示其不會帶來不適當之影響時，則該項流量管理機制具有適當性。
3. 當該流量管理機制為促進機制之適當性所必需時，則該管理機制具有適當性。
4. 當在可用網路資源中已無其他可能造成衝突或具有相同效果的流量管理替代方案時，則該項管理機制具有適當性。
5. ISP 所提供適當的流量管理辦法，以平衡在不同流量類別中的競爭需求，以及不同群體間的競爭利益。

在整體寬頻服務品質管理上，當 ISP 業者針對特定流量類別需求提供不同服務品質時，則可客觀指出該流量類別中所容納的應用服務本身具有特殊的服務品質需求，例如該流量類別中本身包含具即時（real-time）屬性的應用程式，在封包傳送的過程中必須維持低延遲（latency）的傳送品質等。而針對不同的應用程式，亦可能在跳動及干擾情形（jitter）、封包遺失情形（packet lose）以及頻寬需求上有特殊的要求。

（三）禁止行為：ISP 業者對於特定內容、應用或服務及其特定種類，除有必要外（包括維護網路或使用者設備之完整性或處理網路壅塞情形），不得採取封鎖、降速、更改、限制、干擾、降級或歧視等作為。

（四）專業服務：在透過網路接取服務進行應用程式傳遞之外，可另外允許業者提供具有特殊服務品質規格的傳輸服務，此服務稱為「專業服務」（specialised services），意指必須透過特別網路優化才能提供符合特定等級品質需求的服務，必須是網際網路接取以外的服務。另外歐盟為避免「專業服務」與網路接取服務可能因運作於同一網路基礎上而造成原有網路接取服務使用者權益受到損害，或出現業者刻意規避開放網路接取規章所規範之情事，設定以下三項限制：

1. 欲提供「專業服務」之 ISP 業者本身必須在網路接取服務以外俱備充足網路容量以提供「專業服務」。
2. 「專業服務」不得取代網際網路接取服務。
3. 業者在提供「專業服務」時，不得損害 ISP 終端用戶使用權益（availability）及服務品質。

該「開放網路接取規章」在 ISP 應告知消費者網路管理措施方面，以及消費者轉換其他業者無須受罰的規定，遠較 2009 年改革法案來的進步。

第四節 服務品質與網路中立性的監理挑戰

從政府管制觀點來看，QoS 管理面臨許多挑戰。對 QoS 進行區別管理（differentiated management）不僅對網路業者、內容/應用業者，甚至消費者或其他終端使用者皆具有潛在利益。然而，按網路中立性規則來看（按：美國 FCC 已於 2018 年 1 月撤銷該規則），許多線上服務（包括 Video Streaming）根本不需要提供 QoS 保證。要在 QoS 方法上謀求一個平衡點並不容易，**以事業（例如電信業者）而言，提供 QoS 服務意味著需要更多的投資與營運成本，以滿足網路上 QoS 功能要求；但就消費者而言，QoS 服務在某種意涵上是符合消費者權益保障的措施**。

美國、歐盟及部分國家過去或現在皆制定網路中立性相關規範，但由於施行時間尚短，目前難以論斷這些規則的成效為何。例如在網路中立性的基礎上，線上影音服務屬性是全球性的，但**網路中立性卻是國內性或區域性的規範**。網路中立性規範也經常被視為促進應用與服務的主要驅使動力，但這是源自於網際網路創始之初採取「盡量傳送」（best effort）原則下的衍生論述，在面臨電信網路與服務朝向以網際網路為基礎的匯流時代，在此新興的全 IP 網路環境下，端對端的 QoS 仍有其必要（ITU, 2017, p.109）。

在此環境下，所有服務是（或將是）在 IP 網路上傳遞，但並非所有 IP 網路都是開放式網際網路。網路中立性專指網路接取服務端 IAP（例如公開網路接取）的議題，但並非是所有的 IP 網路。例如，電信業者在全 IP 網路上提供具有 QoS 保證的語音服務以作為傳統 PSTN 網路的替代服務，但並非指在開放式網際網路上提供服務，因此與網路中立性無關；但線上語音服務（例如 Skype, Viber）透過 IAP 端提供，是在開放式網際網路上，屬於網路中立性內涵。因此，**在單一市場上讓所有電信服務皆具備網路中立服務，<u>使其成為具備 QoS 保證的資料服務</u>**（例如全球資訊網、電子郵件、VoIP、Facebook、BitTorrent 等），包括具有 QoS 保證的 VoIP、IPTV、商業服務（例如 VPN）、IoT（例如智慧城市）等，這些服務都同樣透過相同的 IP 固定網路、行動網路、核心網路以及轉訊網路等，從企業界角度來看，其結果是極好的（ITU, 2017, p.110），但同樣的，業者提供這類 QoS 服務的成本也相對很高。

2017 年 11 月由行政院審查通過報請立法院審議之「數位通訊傳播法草案」（下稱草案）第二章「維護數位通訊傳播流通」列有四個條文，確保使用者平等近用各種數位通訊傳播服務的自由，維護公平競爭環境，同時使不同資訊能在網路上自由流通，藉以保障言論自由，與國外網路中立性原則的討論相似。

草案中與網路中立性原則有關之內容，諸如：第 6 條第 1 項「**使用者選擇使用數位通訊傳播服務及其設備之自由應受保障**」、第 6 條第 2 項「數位通訊傳播服務提供者對於數位通訊傳播網路**通訊協定或流量管理應以促進網路傳輸與接取之最佳化為原則**......**不得附加任何顯失公平之限制**」、第 7 條「數位通訊傳播服務提供者**應合理使用網路資源**......不得以其他技術或非技術之障礙干擾使用者之選擇」、第 8 條「數位通訊傳播服務提供者得**自由選擇傳輸技術或規格**」以及第 9 條「數位通訊傳播服務提供者提供接取服務時，**應以適當方式對使用者揭露其網路流量管理措施**」，上述規定係要求 ISP 對於「網路流量管理」需以促進傳輸與接取之最佳化（相當於本研究所稱 QoS/QoE）為目的，且不得附加不當限制（包括無故降速、阻礙特定內容），並執行資訊透明化措施，草案精神是以確立消費者與服務平臺之間的民事法律關係。

除此之外，草案第三章「數位通訊傳播服務提供者之責任」也針對消費者權益列出規範。例如：第 10 條要求業者應以適當方法公開揭露資訊，因此若 QoS/QoE 成為服務契約之義務時，業者就必須揭露該網路管理訊息；第 11 條規定若服務使用條款有變更，例如業者變更網路流量管理方式並重大影響使用者權益時，即應依使用者提供之聯絡資訊通知之；第 12 條業者對於所提供之服務應具備約定服務所需之網路品質，例如提供服務所稱傳輸速度倘若在實際使用時顯不符合者，業者應負舉證責任以及因此所造成之損害賠償。因此，當消費者反應影音服務體驗不佳或產生爭議時，例如因網路壅塞造成長時間觀看影音內容不順暢，按上述規定，業者必須負擔舉證之責且須對此情況負損害賠償之責，至於究竟是因平臺系統本身問題抑或網路傳輸之問題，則屬於平臺與 ISP 之間內部問題，消費者僅須提出損害結果即已足。

綜言之，如何在數位影音平臺量測、消費者權益與網路中立性政策之間尋求平衡，本研究認為**政府目前尚無需針對數位影音平台制定規範強制要求其應符合一定品質的網路傳輸（QoS）與消費者體驗品質（QoE）**。目前立法院審議之「數位通訊傳播法草案」相關規定，雖然業者必須擔負起因網路品質不佳對消費者造成損害之賠償責任，但對於業者應促進網路傳輸最佳化、不得附加顯失公平之限制，以及必須揭露網路流量管理措施等義務，其規範目的僅在於釐清消費者與影音平台之間的民事法律關係，而非政府要求業者必須達到 QoS/QoE 之保證。

第五節 線上影音平臺服務法規建議研析

國外大型線上影音服務平臺，如 Netflix、愛奇藝，近年來陸續來台探路，在與國內本地線上影音平臺相互競爭之下，由於品牌知名度、串流技術和先行者的豐富經驗，讓本地線上影音業者面臨外來線上影音在價值創造和競爭之間對本國產業造成極度不公平競爭，包括賦稅不均、寬頻建設投資、影音服務對傳統媒體衝擊等，其中又以網路侵權為線上影音產業最大的挑戰與困境，如前所述。

從政府管制觀點來看，政府如欲對線上影音服務進行 QoS/QoE 規範或管理，將可能面臨許多不確定因素，特別是國際上採行網路中立性原則的「禁止不合理差別待遇」的認定。由於網路中立性屬於國內性規範，但此原則源自於對於網際網路「盡量傳送」(best effort) 原則下的「非歧視性」的規範論點，本研究認為在面臨以網際網路為基礎的數位匯流時代，其實政府無需為目前線上影音服務強制要求必須提供 QoS 保證，若有業者認為服務應具備 QoS，甚至對 QoS 進行區別管理，這對消費者或其他終端使用者而言皆有利益。透過網路中立性的研究，本研究認為 QoS/QoE 本身就是產業會主動關注、自我要求的事情，無需政府以法律法規來形成業者的一般義務。

透過上述的產業現況與國際監理政策研析，本研究發現國際間尚無對於線上影音平臺服務提出具體的量測方法以及法規政策，其原因可能是寬頻技術仍在持續發展當中(特別是行動寬頻)、線上影音平臺服務內容多樣化，以及網路接取生態複雜等各種因素，使得各國尚未針對線上影音平臺服務制定具體的服務品質量測標準。以服務提供者而言，提供 QoS 服務意味著需要更多的投資與營運成本，以滿足網路上 QoS 功能要求；但就消費者而言，QoS 服務在某種意涵上是符合消費者權益保障的措施。

如何在數位影音平臺量測、消費者權益與網路中立性政策之間尋求平衡，本研究認為**政府目前尚無需針對數位影音平台制定規範強制要求其應符合一定品質的網路傳輸（QoS）與消費者體驗品質（QoE）**。目前立法院審議之「數位通訊傳播法草案」相關規定，雖然業者必須擔負起因網路品質不佳對消費者造成損害之賠償責任，但對於業者應促進網路傳輸最佳化、不得附加顯失公平之限制，以及必須揭露網路流量管理措施等義務，其規範目的僅在於釐清消費者與影音平台之間的民事法律關係，而非政府要求業者必須達到 QoS/QoE 之保證。

如何提升通傳產業之數位匯流影音品質，本研究有以下三點建議提供未來相關政策法規之修法方向與建議：

一、具體落實「資訊透明化」原則：線上影音服務平臺應明確向消費者揭露服務品質（QoS），並明白說明各種網路流量管理情境與標準流程（SOP），對於此類「資訊透明」義務之違反，按照數位通訊傳播法草案中違反此民事法律上之權利義務，應依循一般民事法律處理。

二、承諾服務品質形同消保法中「廣告」之義務。線上影音服務平臺若涉及提供未具備所約定服務所需之實際體驗品質（QoE），則應將約定服務品質納入消費者保護法對於「廣告」的定義，違反者除依消保法規定處罰外，尚有公平交易法關於「不實廣告」之不公平競爭的處罰規定可資適用。

三、避免以政府監理方式要求線上影音平臺服務遵守服務品質規範。如今網際網路發展已是全球化、跨疆域性的發展，應避免以政府監理方式要求線上影音平臺服務遵守服務品質規範。較好的做法是，讓線上影音平臺服務業者自行向使用者約定服務品質或體驗，業者可以參考國際標準或實務做法提出適當的服務水準。對於影響消費者群較大之國內外平臺業者，例如國外 YouTube、Netflix、愛奇藝等，政府或可要求業者落地登記來進行服務管理，以確保國內消費者權益。

網際網路發展至今，大多數的人都受惠於其自由開放的特性，然而近來網路假新聞、仇恨言論、隱私風險、使用者個資濫用等事件頻傳，今日的網路已非當時想要實現的開放、安全、有建設性的平臺。2018 年 11 月發明全球資訊網（World Wide Web）英國科學家柏納李爵士（Sir Tim Bernes-Lee）在里斯本網路高峰會中提出「網路契約」(Contract for the Web)，並主張九大原則[65]，為政府、企業與公民提供守護開放網路必須堅守的核心價值，其中要求政府應保護符合公共利益的事情、保護網際網路、鼓勵創新與多樣性。

[65] Contract for the Web: Core Principles, https://fortheweb.webfoundation.org/principles.

因此，在維護網路開放的核心價值下，政府對於線上影音服務應否執行品質量測規範，其主要目的為何？是否符合消費者利益？今日的線上影音平臺與傳統有線電視、IPTV 在訊息傳遞範圍上有相當大的差異，由於地域性的關係，使得政府基於保護消費者權益可以對境內有線電視、IPTV 進行市場參進、產業結構、經營行為等進行管制，然而對於在網際網路上提供線上影音平臺所面對的市場已是全球性的，應可讓業者自行與使用者締約，政府的角色即在維護消費者權益，並促使產業在競爭環境下健全發展。

第三章 國際間影音服務品質量測方法研析

第一節 前言

新興的線上影音平臺服務(指透過網際網路的傳輸提供線上影音觀看)已成為最受歡迎且最耗頻寬的網路應用，根據 Cisco 的視覺化網路指數預估報告顯示，全球消費者上網流量至 2021 年時將有 80%來自於線上影音平臺[66]。在這類服務中，自適性串流（Adaptive Streaming or Adaptive Bitrate Streaming）的技術被廣泛地與內容傳遞網路（Content Delivery Network; CDN）一同佈署以向用戶傳遞多媒體資源。對於影音平臺服務提供商、網路營運商以及用戶來說，都有服務品質量測的需求，特別是量化的用戶體驗品質（Quality of Experience; QoE），以利評估和改進系統效能以及用戶滿意度。

受到視訊品質、播放器行為和螢幕尺寸等影響，傳統的網路性能量測(例如吞吐量、延遲和抖動等）沒有辦法完整評估用戶實際的體驗品質。因此，「數位匯流影音平臺服務品質量測之量測方法委託研究案」(以下簡稱：本研究案）擬針對有線廣播電視系統、電信事業固網與行動寬頻等通傳事業經營者透過行動寬頻或自建 WiFi 網路提供的影音服務(可能是自有影音平臺或第三方影音平臺，如 YouTube）建立一套可佈建於用戶端的客觀 QoE 模型與量測方法，以便評估用戶對於影音服務品質的真實感受。據此，本節的研究目的包含：

一、 探討客觀評估線上影音服務體驗品質的可行性與對應的評估指標；

二、 依據評估指標提出具體的量測方法；

三、 實作量測工具進行實驗驗證所提的評估指標與量測方法。

[66]Cisco, “Cisco Visual Networking Index: Forecast and Methodology, 2016 – 2021,” Sept. 15, 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html

為達前述目的，研究案初期首要任務即是彙整國際間對於線上影音服務品質的量測方法。本研究報告的內容即為研究既有的相關文獻資料後，客觀探討各種（類）線上影音服務品質的量測方法，經分析其優缺點後提出（建議）適用的量測線上影音服務 QoE 的實施方法，作為本研究案後續實際開發線上影音服務 QoE 量測工具的依循。

（一） 服務品質定義

有鑑於服務品質（Quality of Service; QoS）一詞經常被不正確地使用，波蘭的 Gozdecki 等人[67]依循[68]中定義的 QoS 通用模型（General Model），並基於國際電信聯盟（ITU）、歐洲電信標準協會（ETSI）和網際網路工程任務小組（IETF）的定義，對於 QoS 相關的術語進行釐清，其中 ITU 與 ETSI 對於"QoS"的定義如下[69][70]：

"the collective effect of service performance which determine the degree of satisfaction of a user of the service"（決定服務用戶滿意度的服務性能的集體效應）

故其主要強調使用者對於服務品質的滿意程度，並獨立定義"網路性能（Network Performance; NP）"以涵蓋技術面向的品質如下：

"the ability of a network or network portion to provide the functions related to communications between users"（一個網路提供用戶之間通訊相關功能的能力）

[67] J. Gozdecki, A. Jajszczyk, and R. Stankiewicz, "Quality of Service Terminology in IP networks," IEEE Comm. Magazine, March 2003.
[68] W. C. Hardy, QoS Measurement and Evaluation of Telecommunications Quality of Service, Wiley, 2001.vgt
[69] ETSI, "Network Aspects (NA); General Aspects of Quality of Service (QoS) and Network Performance (NP)," Tech. rep. ETR003, 2nd ed., Oct. 1994.
[70] ITU-T Rec. E.800, "Terms and Definitions Related to Quality of Service and Network Performance Including Dependability," Aug. 1993.

至於 IETF 則將"QoS"定義如下[71]：

"A set of service requirements to be met by the network while transporting a flow"（當網路傳輸流量時需要被滿足的一組服務要求）

其並沒有探討使用者的感受，而是以網路本身為考量點，比較近似於 ITU 與 ETSI 對於 NP 的定義。

為了避免混淆，在本報告中，乃將 QoS 視為網路傳輸性能的指標，至於用戶主觀地感受到的體驗品質則另以 QoE 來表示：

1. QoS 主要從網路的角度進行評價，評價對象為乘載該項服務的網路品質。
2. QoE 主要是由終端用戶所評價，評價對象為該項服務的體驗品質。

兩者之間有明顯的差距傳，傳統的 QoS 指標主要規範提供傳輸網路的業者，對大多數基於網際網路協定（Internet Protocol; IP）的服務而言，影響 QoS 的主要可量測的量化參數包含吞吐量（throughput）、傳輸延遲（delay or latency）、傳輸延遲變異（packet delay variation or jitter）與丟包率（packet loss rate）。至於 QoE 則受到許多主觀因素影響，其中包含許多與網路傳輸性能無關的因素。

[71] E. Crawley et al., "A Framework for QoS-Based Routing in the Internet," IETF RFC 2386, Aug. 1998.

第二節 串流影音服務

一、 服務簡介

近期所發展的**線上影音服務**意指透過開放式網際網路(不具有 QoS 保證的網路環境)，直接對用戶提供各種影音內容的服務，其中影音內容通常以**串流**（streaming）或**漸進式下載**（progressive download）方式經由網際網路,再透過用戶端的行動寬頻或固網整合 WiFi 網路傳送至使用者的電視、電腦、智慧型手機或平板電腦等各種終端設備。

與傳統的「預先下載完整的影音檔案才能收看」服務不同，用戶端可以一邊下載一邊觀看(只需等待相對短暫的初始片段下載時間,就可以持續收看完整的影音內容)，享受一經需求即可快速觀看的便利。此截然不同的服務模式直接影響服務品質的需求，因此在本報告中分別以「**串流影音服務**」與「**串流影音平臺**」來代表此類影音媒體經由分批傳輸的模式傳送至用戶端的服務與服務的提供者，以便與傳統服務作一個區隔。基於智慧型個人行動裝置的快速普及與無線網路技術的成熟，完整的串流影音服務也包含了跨裝置連續播的加值服務。對於用戶端的需求則是應具備支援的瀏覽器或特定的播放器/應用程式（App）。

串流影音服務又可進一步被區分為兩類主要應用：

（一） 直播（live）

現場事件透過攝影錄音成為影音內容，即時進行壓縮編碼處理後，立刻經由伺服器在網路上傳送至播放器。典型應用為：新聞、遊戲、社群、實況轉播、視訊會議、保全監控等。

（二） 點播（on demand）

預先錄製的影音內容經過壓縮編碼處理而存放於伺服器端，當用戶端提出收看要求後才透過網路傳送到播放器。典型應用為：點播影音平臺（娛樂）、教育、社群等。

其中，**本研究案主要以第二類應用作為探討的對象。**

發展至今，串流影音服務的營運模式包含：廣告、贊助、付費訂閱、授權、週邊商品與大數據運用等。用戶數與營收數字都顯示人們已開始習慣利用串流影音平臺的方式觀看節目[72]，相較於傳統的頻道電視，用戶可以直接搜尋想要的內容，也可以從用戶的習慣與分群後由平臺推薦適合的內容。免費或付費的平臺都有不少選擇，主要是以內容差異化區隔/佔領市場。舉例而言，各平臺提供的影片類型不盡相同（與版權有關），幾乎沒有一個平臺可以同時提供齊全的美、日、韓、陸、台劇等，用戶通常是依據喜好的內容類型來選擇平臺。當年看似毀了影音產業的網路科技，造就出的串流影音平臺卻成為這些產業的新契機，若沒有這些平臺的興起，整個影音產業仍在數位化下載的風暴中遭受侵權行為摧殘。

台灣常見的影音平臺包含：YouTube、Netflix、FOX+、Amazon Prime Video、CATCHPLAY、Dailymotion、愛奇藝、friDay 影音、CameraBay、CHOCO TV、LiTV、KKTV、LINE TV、Facebook、Google Play 電影、GagaOOLala、酷瞧、Hami Video（原中華影視）、myVideo、Vidol 影音、四季線上影視等。根據台灣最大網路電視平臺 OVO（由 Ovomedia（展雋創意）於 2013 年創立）公布的業界首份數據報告，影音付費平臺最受歡迎的前三名分別是愛奇藝（韓/陸劇）、LiTV（體育賽事/本土頻道）與 friDay 影音（電影）[73]。足見消費者終究還是以內容為導向來選擇平臺，但優質內容也要有相等品質的觀看環境。與利用網路營運商或多系統業者的專用網路提供服務的 IPTV（Internet Protocol Television）平臺相較，**串流影音服務係基於開放式網際網路作為傳輸，使得提供服務的傳輸品質較難維持**。Akamai Technologies, Inc.研究觀眾對緩衝處理及低品質影音的身體及情緒反應，結果顯示不論免費或訂閱模式都會因此流失觀眾[74]。

[72]柯思瑪, "你付費看影集了嗎？串流影音平臺火力全開！," 數位時代, Nov. 24, 2017. [Online]. Available: https://www.bnext.com.tw/article/46954/video-stream-netflix
[73]TechNews, "OVO 月觀看突破 700 萬次，公布台灣首份 OTT 網路電視數據報告," Sept. 15, 2017. [Online]. Available: http://technews.tw/2017/09/15/ovo-taiwan-internet-tv-report/
[74]Akamai,"The Science Behind How Our Bodies React to Video Quality," April, 2017. [Online]. Available: https://content.akamai.com/gl-en-pg9246-sensum-whitepaper.html

二、 典型的傳輸系統架構

目前串流影音服務業者多透過 CDN 將內容傳送至消費者，以提供較為穩定的服務品質。典型的串流影音服務傳輸架構簡化如圖 19，資料的流向程序一般是由左端的用戶終端（Users' Terminal）向右端的串流影音平臺（Streaming Service Provider）發出請求，經平臺決策後將該請求重導至最適合的 CDN 節點（CDN Edge Server），由該 CDN 節點向原用戶傳送影音串流。

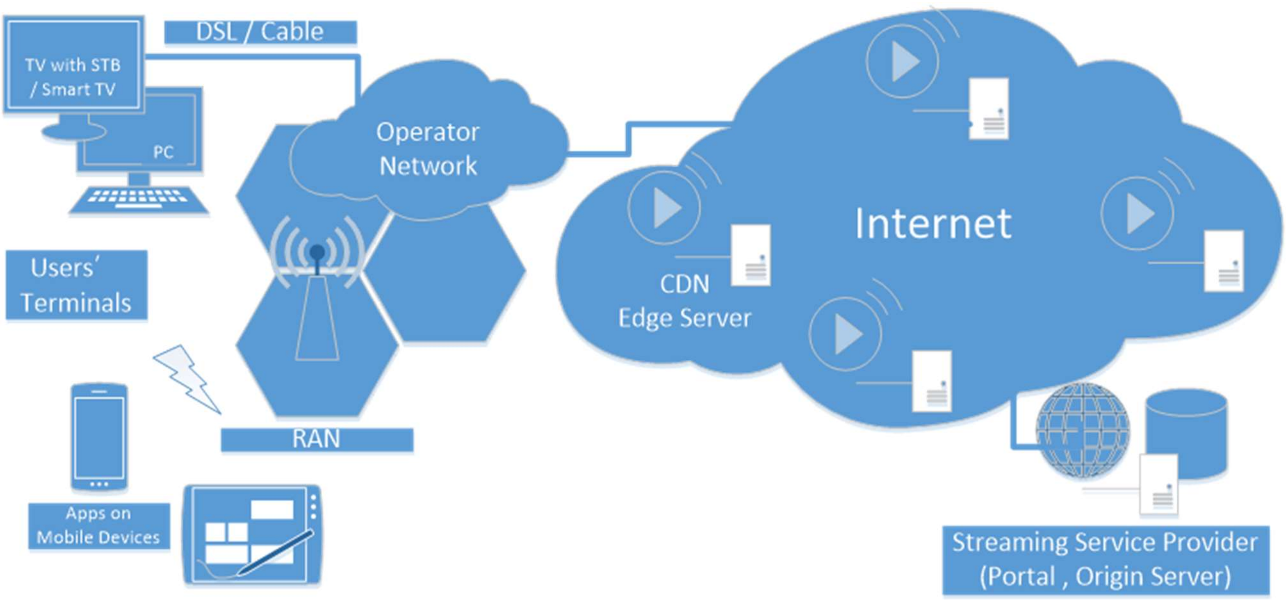


圖 19 典型的串流影音服務傳輸架構(資料來源:本研究整理)

其中 CDN 節點會在多個地點，多個不同的網路上擺放，而節點之間會互相傳輸內容。在 CDN 服務的使用下，可提供內容彙集及快速存取服務，對用戶的下載進行最佳化，以期提高用戶的體驗品質。

然而，串流媒體為了配合有限網路頻寬的限制（圖 19 中 Operator Network 與 Internet 的綜合效應），通常都將影音內容進行高程度的資料壓縮（源編碼）處理，使得畫質受到影響，失真現象相對嚴重。若網路頻寬仍不足以負荷，就會在播放時造成遲滯（lag）與卡頓（stalling）（又被稱為重新緩衝（rebuffering））現象。這些影響都有可能引發收看者負面的觀看體驗。為了降低後者的影響，串流影音平臺多採用自適性串流（Adaptive Streaming）技術因應。目前常見的自適性串流技術（包含 HTTP Live Streaming（HLS）[75]、Microsoft Smooth Streaming （MSS）[76]、HTTP Dynamic Streaming（HDS）[77]、與 Dynamic Adaptive Streaming over HTTP（DASH）[78]等）是藉由將多位元率編碼的影音內容切割成較小的片段並透過 HTTP 方式傳輸（可以穿過允許 HTTP 協定通過的防火牆/代理伺服器），以支援由解碼端驅動的自適性速率機制，播放器可依當前條件（例如頻寬和 CPU 使用率）動態地切換適合的位元率方案，無縫適應不斷變化的網路條件，避免卡頓（重新緩衝）。儘管如此，有限網路頻寬下勉強以高壓縮率（避免卡頓）的影音內容（可能反應在低解析度、低影格率與低量化位元）呈現給用戶，收看者的觀看體驗可能依舊是負面的。

[75]Apple Inc., “HTTP live streaming technical overview 2013,” [Online]. Available: https://developer.apple.com/library/ios/documentation/networkinginternet/conceptual/streamingmediaguide/Introduction/Introduction.html
[76]A. Zambelli, “Smooth streaming technical overview,” [Online]. Available: http://www.iis.net/learn/media/on-demand-smooth-streaming/smoothstreamingtechnical-overview
[77]Adobe Systems Inc., “HTTP dynamic streaming 2013,” [Online].Available: http://www.adobe.com/products/hds-dynamic-streaming.html
[78]DASH Industry Forum, “For promotion of MPEG-DASH 2013,” [Online]. Available: http://dashif.org

第三節 串流影音服務之量化品質量測概觀

一、 問題描述

隨者視音訊串流技術的成熟，串流影音服務逐漸普及，再加上高解析度的螢幕製作成本迅速的下降（720p 的解析度已成為行動裝置的基本配備），民眾的收視習慣隨之改變，然而民眾使用收費服務時所產生的消費糾紛也隨之而來。定性而言，串流影音服務的「高品質」意謂著「低延遲、流暢穩定、高畫質與高傳真」；然而，**影音服務的品質需要被量化**，具體的量化數據才有助於資料蒐集、處理、統計與分析，如何具體地衡量消費者實際體驗到的影音服務品質卻仍待研究。

影響用戶對於影音服務體驗品質的層面很廣，各用戶多樣化的使用情境也是一個因素，至於偏技術面的因素通常涉及影音平臺效能、傳輸網路服務品質及終端設備性能等。其中又以視訊服務品質主導影音服務體驗品質，視訊服務品質的好壞，取決於空間品質（與視訊源本身的品質、螢幕尺寸等有關）與時間品質（與視訊播放過程的順暢度有關）的綜合影響，前者主要由內容提供者採用的編碼技術規格及用戶端的螢幕尺寸決定，後者則與網路服務品質及播放器的設計相關。

根據 2017 Q1 由 Mux, Inc.委託調查 1,035 名美國消費者的線上影音觀看體驗的結果[79]，當用戶感受影音服務體驗品質不佳時，通常第一個歸咎於所屬的網際網路服務提供商（Internet Service Provider; ISP）（相關問卷調查結果彙整如圖 20）。因此，對於經營圖 19 中 Operator Network 的有線廣播電視系統、電信事業固網與行動寬頻等通傳事業經營者而言，如何隨時掌握用戶透過所經營網路觀看線上影音服務的體驗品質、進而優化架構以改善品質，成為營運與管理的重要課題，也是本研究案的研究動機。

[79]Mux, "2017 VIDEO STREAMING PERCEPTIONS REPORT," 2017 Q1. [Online]. Available: https://static.mux.com/downloads/2017-Video-Streaming-Perceptions-Report.pdf

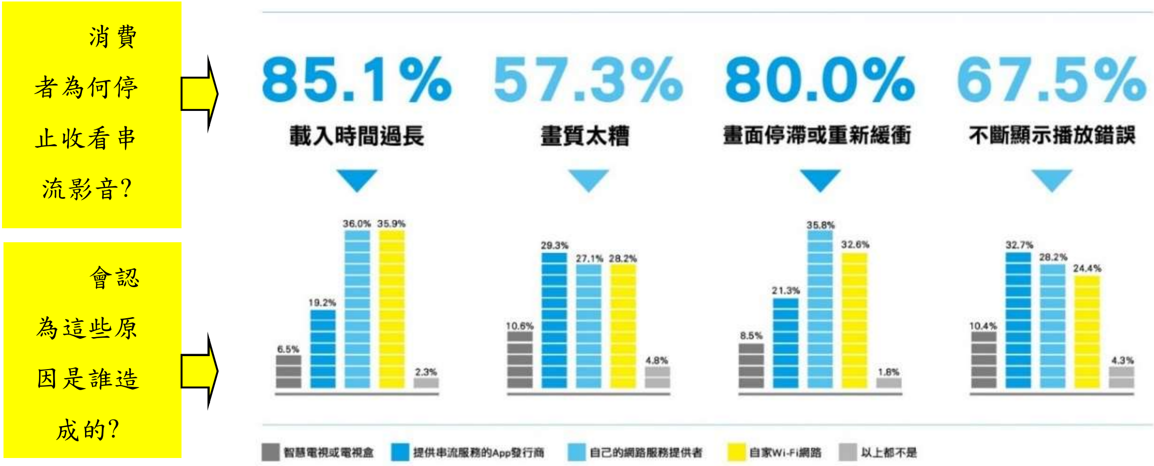


圖 20 消費者的線上影音觀看體驗調查結果[80]

二、 量測方法的演進

線上影音服務從 Server 端到用戶端的傳輸路徑簡化模型如圖 21 所示，包括：編碼器、傳輸網絡、解碼器與顯示器。其中任一處都可能引入會影響用戶體驗品質的失真現象，例如編碼器參數將影響視訊品質、傳輸網路影響播放的流暢度、播放器與顯示器影響最終視訊呈現的效果等。

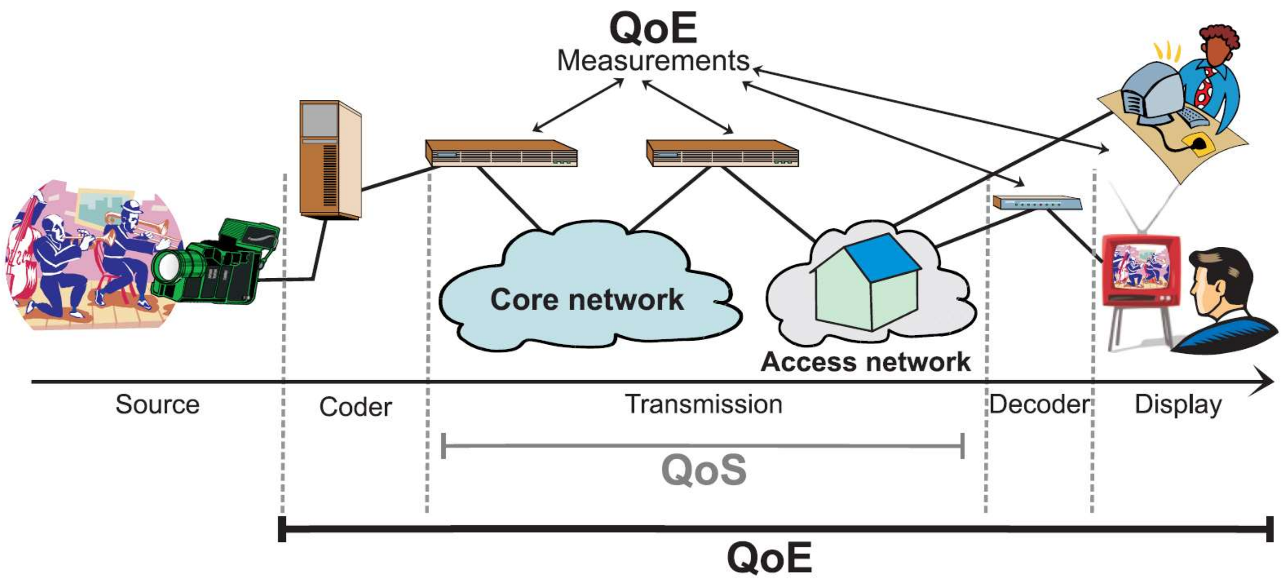


圖 21 傳輸路徑簡化模型[81]

[80] 柯思瑪, “你付費看影集了嗎？串流影音平臺火力全開！,” 數位時代, Nov. 24, 2017. [Online]. Available: https://www.bnext.com.tw/article/46954/video-stream-netflix

[81] R. Serral-Gracià et al., “An Overview of Quality of Experience Measurement Challenges for Video Applications in IP Networks,” *Proc. International Conference on Wired/Wireless Internet Communications,* Luleå, Sweden, June 1-3, 2010, pp. 252-263.

為了量化量測服務品質，就需要定義出可反映視訊服務品質的指標與量測方法。就量測方法的發展演進來看，大致可以分成四個階段：QoS 監控（QoS Monitoring）、主觀測試（Subjective Test）、客觀模型（Objective Quality Model） 與資料驅動型分析（Data-driven Analysis），如圖 22 所示。而這些不同視訊服務品質量測方法的比較則摘要如表 5。

表 5 不同視訊服務品質量測方法的比較[82]

	Direct measure of QoE	Objective or Subjective	Real-time	Wide application	Cost
QoS monitoring	NO	Objective	YES	Wide	Not sure
Subjective test	YES	Subjective	NO	Limited	High
Objective quality model	NO	Objective	YES/NO	Limited	Low
Data-driven analysis	YES	Objective	YES	Wide	Not sure

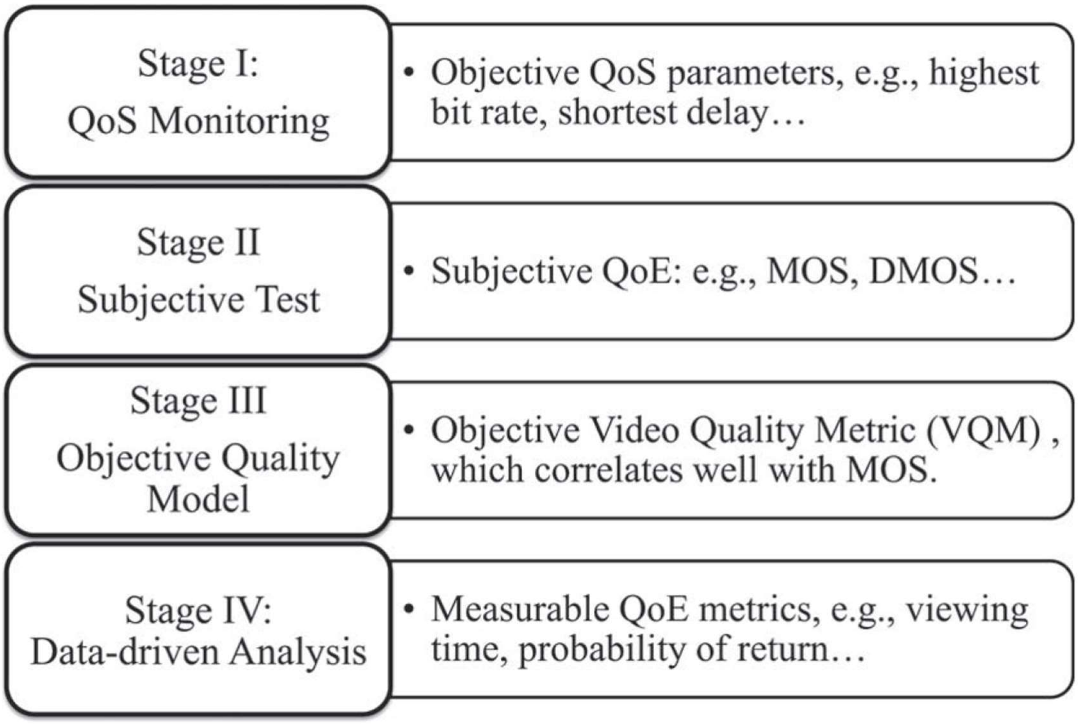


圖 22 視訊服務品質量測方法的演進[83]

[82]陈楚雄，柯江毅與覃道满，"视频业务体验评估和优化提升探讨 （Discussion on Video Service Experience Evaluation and Optimization）," 邮电设计技术, no. 2, pp. 17-23, 2017.
[83]Y. Chen, K. Wu, and Q. Zhang, " From QoS to QoE: A Tutorial on Video Quality Assessment," IEEE Comm. Surveys Tut., vol. 17, no. 2, pp. 1126-1165, 2015.

本研究報告後續主要探討前三階段的量測方法（即 QoS 監控、主觀測試與客觀模型），適用於評估每一次個別影音服務的品質。至於第四階段新興的資料驅動型分析方法乃仰賴於大規模（尺度）用戶數據的取得，例如：觀看時間 （viewing time）、回閱率（return rate）、棄閱率（abandoned view ratio）、某時間區間內的觀看次數（number of view）、視訊評第（video rating）等與用戶行為相關的參數，適用於平臺業者從更高層面分析用戶的長期體驗品質，則不在本研究報告探討的範圍，相關的文獻資料可以參考[84],[85],[86],[87]。

三、 QoS－基於網路性能指標的量測

此類視訊服務品質量測方法主要是透過量測網路性能指標 QoS，再由 QoS 量測結果經由某種“QoS-to-QoE 對應函式”預估 QoE。其中，QoS 量測參數一般包含傳輸延遲、傳輸延遲變異、丟包率與上/下行吞吐量等[88]。網路品質量測的發展已有一段歷史，故傳統 QoS 參數的量測方法發展較為成熟，業界已有普遍習知的共通作法，例如[89] [90]中提出了相似的 QoS 參數量測方法與建議。唯一需要注意的，在本研究案中，量測路徑的另一端預設不可控制的 CDN 節點，將使得某些需要端與端配合的 QoS 參數量測方法不適用，僅能考慮在此實際條件下可行的量測方法。

---

[84]Q. Wang et al., “Data Analysis on Video Streaming QoE over Mobile Network,” EURASIP Journal on Wireless Communications and Networking, July 2018.
[85]A. Balachandran, “Developing a Predictive Model of Quality of Experience for Internet Video,” ACM SIGCOMM Computer Communication Review, vol. 43, no. 4, October 2013, pp. 339-350.
[86]J. Jiang et al., “A Practical Prediction System for Video QoE Optimization,” Proc. USENIX Symposium on Networked Systems Design and Implementation, Santa Clara, CA, March 16-18, 2016, pp. 137–150.
[87]Y. Sun et al., “Improving video bitrate selection and adaptation with data-driven throughput prediction,” Proc. ACM SIGCOMM, Florianopolis, Brazil, Aug. 22-26, 2016, pp. 272–285.
[88]D. Paxson et al., "Framework for IP Performance Metrics,” IETF RFC 2330, May 1998.
[89]SamKows, “SamKnows Test Methology (Document Reference: SQ301-005-EN),” Feb. 2015.
[90]BEREC, “Net Neutrality Regulatory Assessment Methodology,” BoR (17) 178, May 2017.

至於 QoS 與 QoE 之間的對應仍是個研究議題[91,92,93,94]。一方面是因為不同應用有不同的考量，在考量一般用戶端播放器皆具備適當緩衝器情況下，相對於語音服務主要訴求為低傳輸延遲與抖動，串流影音點播服務更需要穩定的傳輸速率才可以讓使用者順暢的播放影片。另一方面則是先天上欲單純地以純粹客觀的 QoS 預估完全主觀的 QoE 就是不可能達成的，從圖 21 可明顯看出，雖然 QoS 對 QoE 有（部份）影響，但這些 QoS 量化參數並不足以直接反應用戶對於影音品質的真實感受。因此，本研究案的研究方向乃是朝向結合 QoS 量測參數與其它參數來預估 QoE，或是當服務品質不佳時（預估 QoE 低於某個門檻）由 QoS 參數量測結果來進行問題診斷。

四、 QoE－基於使用者體驗/感受的量測

此類視訊服務品質量測方法主要是將使用者的體驗/感受對應到一量化指標，依據是否由真人評定 QoE，又可分為主觀的（subjective）QoE 與客觀的（objective） QoE 兩類。

[91]D. Pal and V. Vanijja, “Effect of Network QoS on User QoE for a Mobile Video Streaming Service Using H.265/VP9 Codec,” Procedia Computer Science, vol. 111, Aug. 2017, pp. 214-222.
[92]T. Wang, A. Pervez and H. Zou, “VQM-based QoS/QoE Mapping for Streaming Video,” Proc. IEEE International Conference on Broadband Network and Multimedia Technology, Beijing, China, Oct. 26-28, 2010, pp. 807-812.
[93]H.J. Kim and S.G. Choi, “A Study on a QoS/QoE Correlation Model for QoE Evaluation on IPTV Service,” Proc. IEEE International Conference on Advanced Communication Technology, Phoenix Park, South Korea, Feb. 7-10, 2010, pp. 1377-1382.
[94]H.J. Kim et al., “The QoE Evaluation Method through the QoS-QoE Correlation Model,” Proc. IEEE International Conference on Networked Computing and Advanced Information Management, Gyeongju, South Korea, Sept. 2-4, 2008, pp. 719-725.

(一) 主觀的 QoE

在大多數影音串流應用中，由於人是影音服務最終的消費者，因此由人所進行主觀的 QoE 評估（部份文獻中也稱之為 “user QoE”）是最直接與可靠的方法。此類主觀的 QoE 評估方法蒐集觀眾主觀的感受（perception）並予以量化及統計，例如表 所列之平均意見分數（Mean Opinion Score; MOS）[95],[96]就是常見的方式，由真人根據真實感受（非常差、差、普通、好、非常好）給予 1 分至 5 分的絕對分數。

在進行主觀的 QoE 評估時，觀察使用者的行為與反應可提供數據以便研究人類在評估體驗品質時的行為模式，而主觀的 QoE 評估結果也可進一步作為驗證並比較其它客觀的 QoE 評估方法的性能。然而這種主觀的用戶研究並不方便，需要實際的人員在實驗室環境下參與測試，耗時且昂貴。最重要的是，它們並不適用於隨時地（或週期性地）量測應用。

表 6 平均意見分數 （Mean Opinion Score; MOS）[95]

MOS	Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

[95]ITU-T Rec. P.910, Subjective video quality assessment methods for multimedia applications, 2008.
[96]ITU-T Rec. P.913, Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment, 2014.

(二) 客觀的 QoE

為了避免進行主觀測試，而希望能客觀且廣泛地量測 QoE，研究人員通常會基於人因工程發展所謂的客觀的 QoE 模型。這類客觀 QoE 模型將 QoE 視為 QoS 參數、應用性能指標（Application Performance Metrics; APM）和其它外部因素的某種函式關係[97,98]，以便透過可量測到的參數來計算（預測）QoE。這些可量測到的參數可能包含網路品質參數、影格率（frame rate）、串流位元速率（bit rate）、解析度（resolution）、初始緩衝時間、卡頓時間、卡頓次數或頻率與播放器事件日誌等，實際可量測到的參數與架構有關。理想的客觀 QoE 模型應該要能得到與主觀測試結果（作為 ground truth）高度正相關的 QoE 預測值。

基於量測過程是否需要參考原始視訊，客觀 QoE 模型又可被分類為完整參考（Full Reference; FR）模型[99,100]、簡化參考（Reduced Reference; RR）模型[101,102]和無參考（No Reference; NR）模型[103,104]。FR 能夠取得完整的原始檔案並與經過傳輸後的檔案進行比較；RR 僅能取得少數或者片段的原始數據；NR 則是無法取得任何原始數據。本研究案所考量的隨時 QoE 評估應用時，預設是無法取得原始視訊數據，這意味著必須使用 NR 模型，其應用場合較不受限制，但準確性將受到影響。

---

[97] Y. Chen, K. Wu, and Q. Zhang, “ From QoS to QoE: A Tutorial on Video Quality Assessment,” IEEE Comm. Surveys Tut., vol. 17, no. 2, pp. 1126-1165, 2015.
[98] ITU-T Rec. J.343, Hybrid perceptual bitstream models for objective video quality measurements, 2014.
[99] TU-T Rec. J.343.5, Hybrid-FRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and encrypted bitstream data, 2014.
[100] ITU-T Rec. J.343.6, Hybrid-FR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and non-encrypted bitstream data, 2014.
[101] ITU-T Rec. J.343.3, Hybrid-RRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and encrypted bitstream data, 2014.
[102] ITU-T Rec. J.343.4, Hybrid-RR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and non-encrypted bitstream data, 2014.
[103] ITU-T Rec. J.343.1, Hybrid-NRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of encrypted bitstream data, 2014.
[104] ITU-T Rec. J.343.2, Hybrid-NR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of non-encrypted bitstream data, 2014.

第四節 QoS 量測方法研析

一、 常用網路性能指標之量測

（一） 傳輸延遲（Delay or Latency）

1. 定義：鏈路上的單向封包傳輸延遲（單位: ms）[105]。
2. 分析：值越小表示性能越佳，當此指標性能不佳時可能造成網頁回應慢、初始緩衝時間長、無法滿足即時應用的要求等影響。
3. 量測方法：如圖 23 所示，來源端與目的端需維持時間同步，由來源端傳送一個攜帶發送時間（T_0）戳記的封包，目的端記錄接收到該封包的時間（T_1），即可計算傳送延遲。可以傳輸控制協定(Transmission Control Protocol, TCP)或使用者資料包協定(User Datagram Protocol, UDP)方式進行。

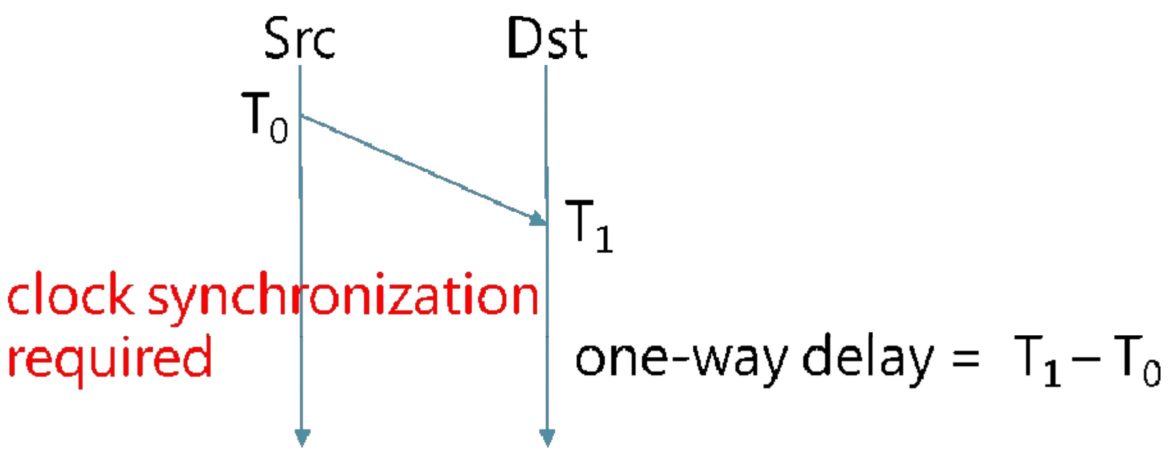


圖 23 傳輸延遲量測示意圖(資料來源:本研究整理)

[105] G. Almes, S. Kalidindi and M. Zekauskas, “A One-way Delay Metric for IPPM,” IETF RFC 2679, Sept. 1999.

（二） 傳輸延遲變異（IP Packet Delay Variation; IPDV）

1. 定義：鏈路上的單向封包傳輸延遲變異（單位: ms）變異[106]（多數人稱之為 “Jitter / 抖動”[107,108,109]）。
2. 分析：值越小表示性能越佳，當此指標性能不佳時可能造成偶發性遲滯與卡頓、無法滿足即時應用的要求等影響。
3. 量測方法：如圖 24 所示，來源端與目的端需維持時間同步，由來源端傳送第一個攜帶發送時間（T0）戳記的封包，目的端記錄接收到第一個封包的時間（T1），由來源端傳送第二個攜帶發送時間（T2）戳記的封包，目的端記錄接收到第二個封包的時間（T3），即可計算傳送延遲變異。可以 TCP 或 UDP 方式進行。

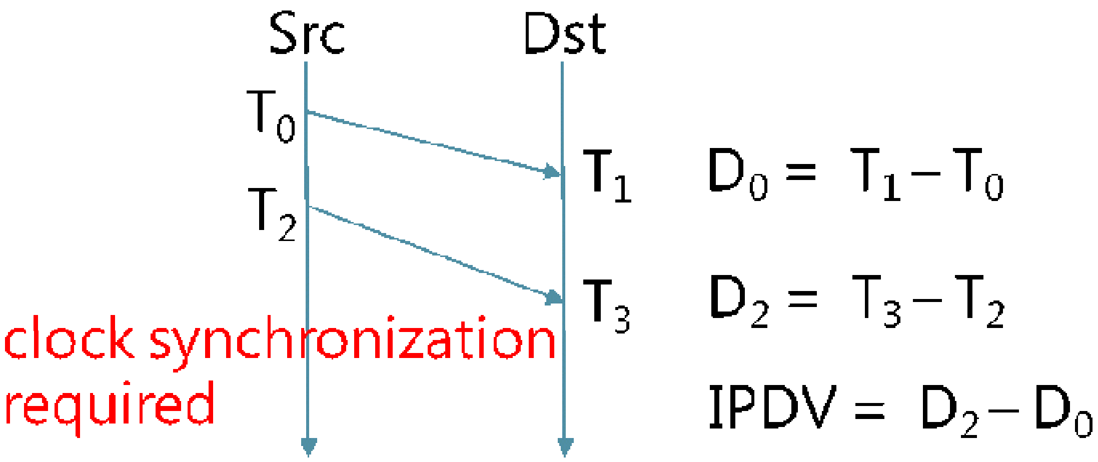


圖 24 傳輸延遲變異量測示意圖(資料來源:本研究整理)

---

[106]C. Demichelis and P. Chimento, “IP Packet Delay Variation Metric for IP Performance Metrics (IPPM),” IETF RFC 3393, Nov. 2002.
[107]S. Poretsky et al., “Terminology for Benchmarking Network-layer Traffic Control Mechanisms,” IETF RFC 4689, Oct. 2006.
[108]V. Jacobson, K. Nichols and K. Poduri, “An Expedited Forwarding PHB,” IETF RFC 2598, June 1999.
[109]H. Schulzrinne et al., “RTP: A transport protocol for real-time applications,” IETF RFC 1889, Jan. 1996.

（三） 往返時間延遲（Round-trip Time; RTT）

1. 定義：鏈路上的封包往返時間延遲（單位: ms）[110]。
2. 分析：值越小表示性能越佳，當此指標性能不佳時可能造成網頁回應慢、初始緩衝時間長、無法滿足即時應用的要求等影響。
3. 量測方法：如圖 25 所示，由來源端傳送一個封包（並記錄本地發送時間 T_0），並等待目的端回應，當接收到回應封包時記錄接收的時間(T_3)，即可計算往返時間延遲。可以 TCP、UDP 與網際網路控制訊息協定(Internet Control Message Protocol, ICMP)方式進行，**尤其適合無法與目的端維持時間同步時的應用。**

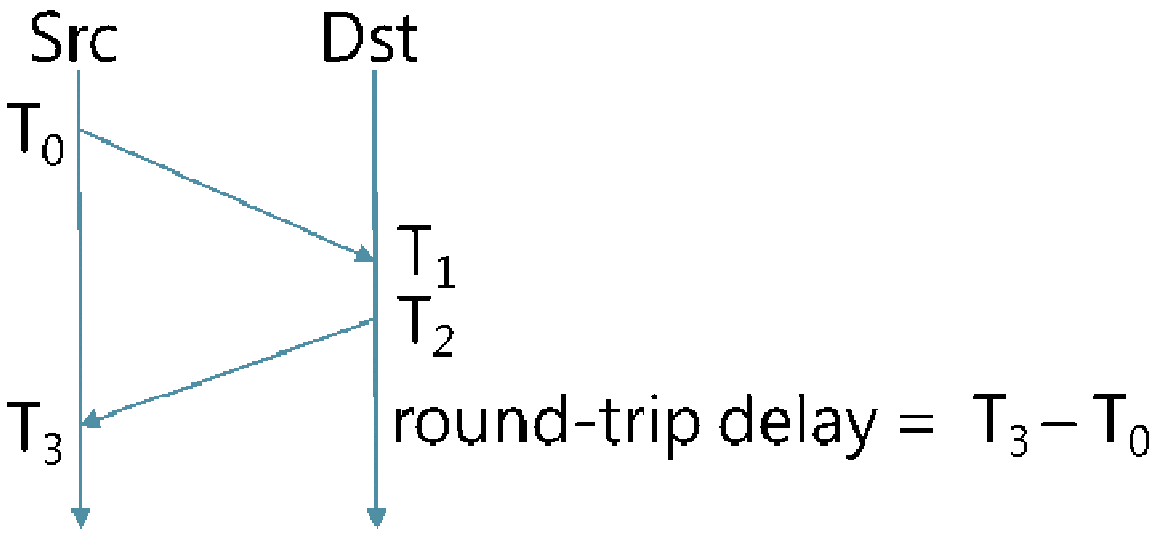


圖 25 往返時間延遲量測示意圖(資料來源:本研究整理)

---

[110] G. Almes, S. Kalidindi and M. Zekauskas, “A Round-trip Delay Metric for IPPM,” IETF RFC 2681, Sept. 1999.

（四） 丟包率（Packet Loss Rate）

1. 定義：鏈路上的單向封包傳送失敗率[111]/往返封包傳送失敗率[112]（單位:%）。
2. 分析：值越小表示性能越佳，當此指標性能不佳時可能造成網頁無回應、初始緩衝時間長、偶發性遲滯與卡頓、無法滿足即時應用的要求等影響。
3. 量測方法：如圖 26 所示，在前述傳輸延遲量測時記錄封包傳送失敗的次數。

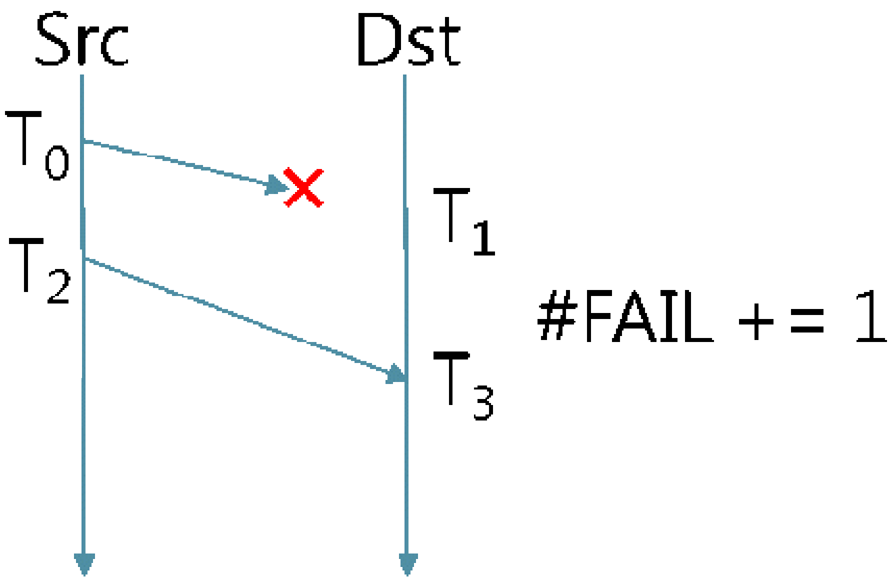


圖 26 丟包率量測示意圖(資料來源:本研究整理)

[111]G. Almes, S. Kalidindi and M. Zekauskas, “A One-way Packet Loss Metric for IPPM,” IETF RFC 2680, Sept. 1999.
[112]A. Morton, “Round-Trip Packet Loss Metrics,” IETF RFC 6673, Aug. 2012.

（五） 上/下行吞吐量（Downlink/Uplink Throughput）

1. 定義：鏈路的上/下行吞吐量（單位: Mbps）[113,114]。
2. 分析：值越大表示性能越佳，當此指標性能不佳時可能造成初始緩衝時間長、經常性遲滯與卡頓等影響。
3. 量測方法：上/下行吞吐量分別測試（不同時），某單向測試時依指定方向傳輸預存的檔案，該檔案的大小假設為 X_{DL}（或 X_{UL}）MBytes。測試時同時建立多個（建議 3~5 個）HTTP 連線進行傳輸（避免受到流速控制（flow control）機制影響），並以預指定的時間間隔取樣傳輸封包，並記錄取樣之（瞬時）傳輸速率、累計資料傳輸總量及所需之傳輸時間後，即可計算吞吐量=傳輸總量/傳輸時間

一般而言，建議採取固定傳輸時間，適用於所有寬頻速率。若有傳輸用量的限制，才考慮固定傳輸總量。後者雖適合節省頻寬使用，但對於較高速率的鏈路可能會有較大的量測誤差，對於較低速率的鏈路則又可能會造成測試時間過長等影響。折衷的解決方案是可以考慮設計適當的演算法以自適性調整傳輸總量，當量測結果接近收斂時即可結束量測。

(1) 下行吞吐量：如圖 27 所示。

(2) 上行吞吐量：如圖 28 所示。

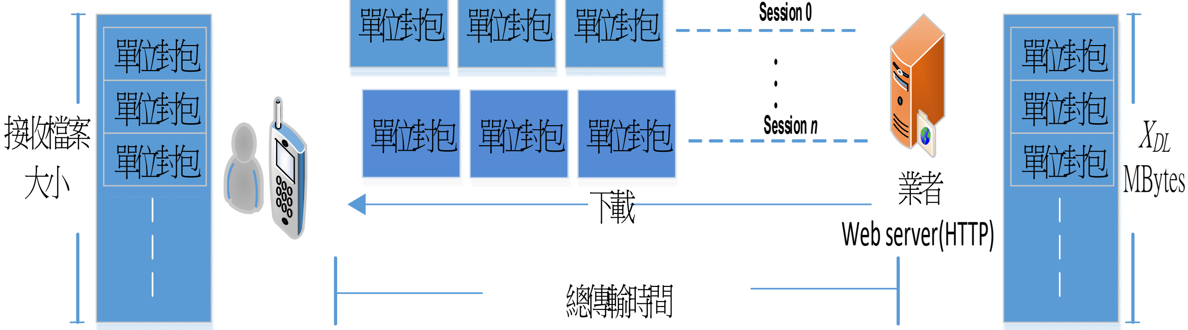


圖 27 下行吞吐量量測示意圖(資料來源:本研究整理)

[113]S. Bradner and J. McQuaid, “Benchmarking Methodology for Network Interconnect Devices,” IETF RFC 2544, Mar. 1999.

[114]B. Constantine et al., “Framework for TCP Throughput Testing,” IETF RFC 6349, Aug. 2011.

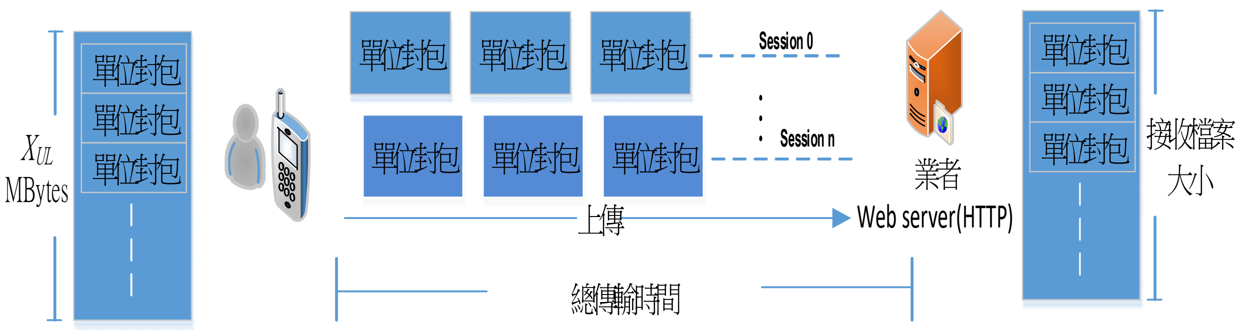


圖 28 上行吞吐量量測示意圖(資料來源:本研究整理)

二、 QoS 與 QoE 之對應關係

基於前一節中所量測到的各種網路品質參數，經由某種“QoS-to-QoE 對應函式”可用於預估 QoE。在“QoS-to-QoE”對應這部份可分為兩類作法，個別描述如下。

（一） QoS-QoE 關係模型（Correlation Model）或映射（Mapping）

如第三章第三節所提，QoS 與 QoE 之間的對應仍是個研究議題。典型上，此類方法假設某些 QoS 參數對於 QoE 有影響，並以主觀 QoE 的值對該欲考量的 QoS 參數進行迴歸（regression）分析（線性、非線性或對數等），以決定最符合的映射函式與係數值（即能與主觀 QoE 具有最接近的預測結果）。

相關文獻資料可以參考[115],[116],[117]。由於其作法類似，在此我們僅以[118]的研究結果作為例子說明。在該文獻中探討在行動串流視訊服務下五種網路 QoS 參數對 QoE 的影響，這五種參數包含了常用的丟包率、抖動與吞吐量等三項，再加上兩種作者相信對 QoE 也有影響的參數：可變初始延遲（variable initial delay） 與緩衝延遲（buffering delay）。為探討個別參數對於 QoE 間的影響，該文獻採用的是對個別單一參數進行多種非線性迴歸（nonlinear regression）分析，並定義一決策指標用於評估映射誤差，以便找出最佳映射函式。其研究結果顯示，丟包率、抖動、可變初始延遲與緩衝延遲這四項參數與 QoE 間的最佳映射函式是雙因子指數（two-factor exponential）函式

$$\mathrm{MOS} = a \times e^{(b \times \mathrm{QoS})} + c \times e^{(d \times \mathrm{QoS})}$$

其中 a, b, c 與 d 是係數，可由迴歸分析決定其值，如圖 29 所示為丟包率與 MOS 值變化的關係。而吞吐量與 QoE 間的最佳映射函式則是對數 （logarithmic） 函式

$$\mathrm{MOS} = a \times \log(\mathrm{QOS}) + b$$

其中 a 與 b 是係數，可由迴歸分析決定其值，如圖 30 所示為吞吐量與 MOS 值變化的關係。

---

[115]T. Wang, A. Pervez and H. Zou, "VQM-based QoS/QoE Mapping for Streaming Video," *Proc. IEEE International Conference on Broadband Network and Multimedia Technology,* Beijing, China, Oct. 26-28, 2010, pp. 807-812.

[116]H.J. Kim and S.G. Choi, "A Study on a QoS/QoE Correlation Model for QoE Evaluation on IPTV Service," *Proc. IEEE International Conference on Advanced Communication Technology,* Phoenix Park, South Korea, Feb. 7-10, 2010, pp. 1377-1382.

[117]H.J. Kim et al., "The QoE Evaluation Method through the QoS-QoE Correlation Model," *Proc. IEEE International Conference on Networked Computing and Advanced Information Management,* Gyeongju, South Korea, Sept. 2-4, 2008, pp. 719-725.

[118]D. Pal and V. Vanijja, "Effect of Network QoS on User QoE for a Mobile Video Streaming Service Using H.265/VP9 Codec," *Procedia Computer Science,* vol. 111, Aug. 2017, pp. 214-222.

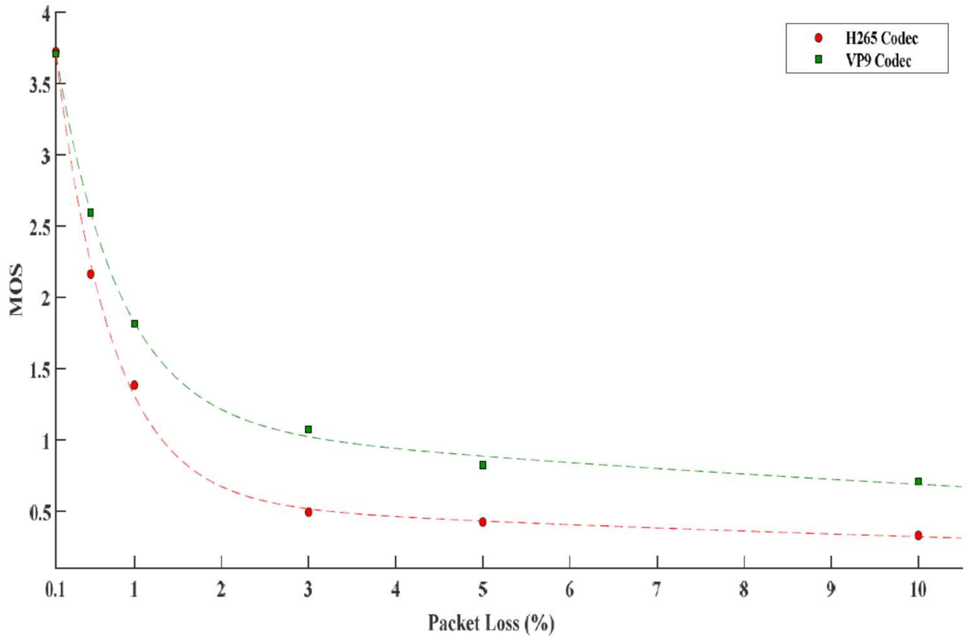


圖 29 丟包率與 MOS 值變化的關係[52]

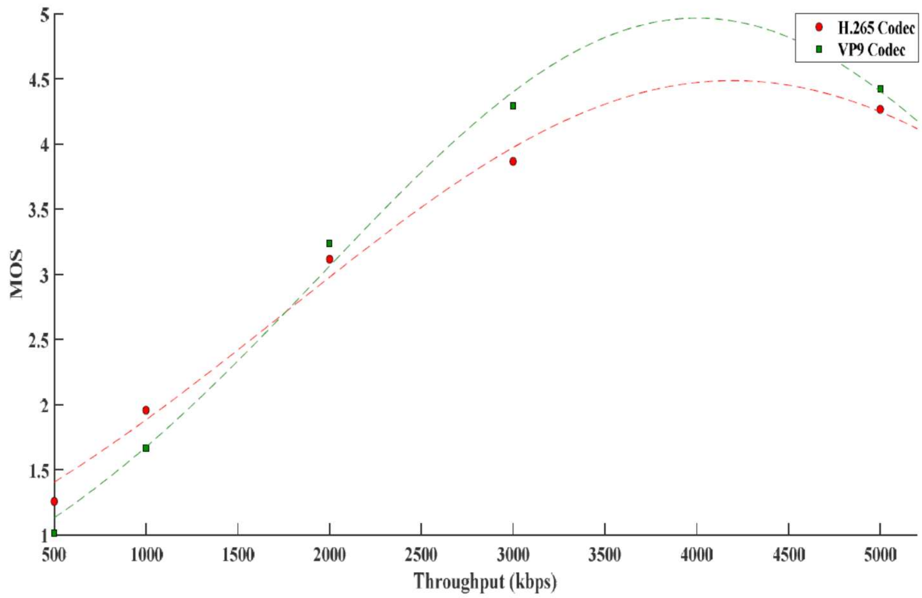


圖 30 吞吐量與 MOS 值變化的關係[52]

（二） 以 QoS 參數與緩衝模型進行資料串流品質模擬測試

此類方法乃基於以 QoS 參數（例如瞬時傳輸速率與吞吐量）進行資料串流品質模擬測試，若將線上影音內容（即視音訊壓縮資料）統一視為資料串流，則透過適當的緩衝模型（如圖 31）測量該鏈路上的資料串流品質即可用以代表該鏈路所能提供影音服務的品質。

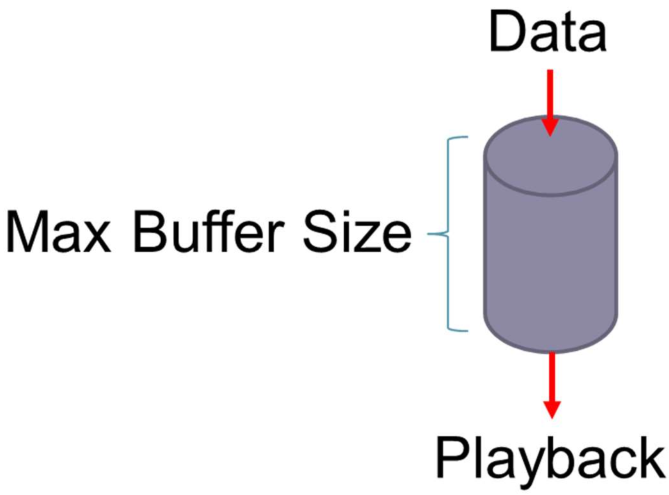


圖 31 簡化的緩衝模型(資料來源:本研究整理)

在此我們僅以[119]所提的模擬測試方法作為例子說明，此方法以播放器取用緩衝區內的資料播放作為前提，基於瞬時傳輸速率的變化來模擬觀看特定位元率的內容時可能遭遇的事件，當資料傳輸率低於影音播放位元率時則表示緩衝資料量遞減，當緩衝耗盡時則表示卡頓事件（如圖 32。此類量測方法的優點在於無須實作視音訊串流與編解碼技術，且只要測量一次鏈路上的資料串流品質即可用以推算（預估）各種影音解析度下預期的影音服務穩定度，例如可以預估最佳可靠（指不發生卡頓現象）的影音解析度。

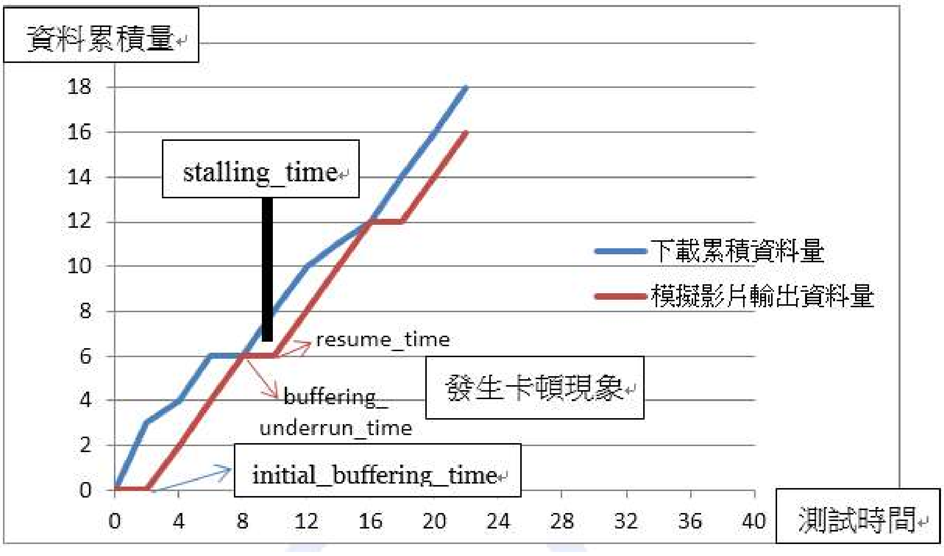


圖 32 發生卡頓現象之資料串流品質模擬示意圖(資料來源:本研究整理)

[119]SamKows, “SamKnows Test Methology (Document Reference: SQ301-005-EN),” Feb. 2015.

第五節 QoE 量測方法研析

一、 主觀量測－使用者評分

主觀量測的方式乃直接由使用者給予評分，其測試流程整理如圖 33，重點在於測試環境的規劃籌備、影片的選取（需涵蓋空間/時間/亮度/主題各維度的複雜度）、測試人員的挑選（需涵蓋目標年齡族群、性別與人數）與異常數據 （outlier）的剔除，詳細的測試建議與規範可以參考[120],[121],[122]等標準。

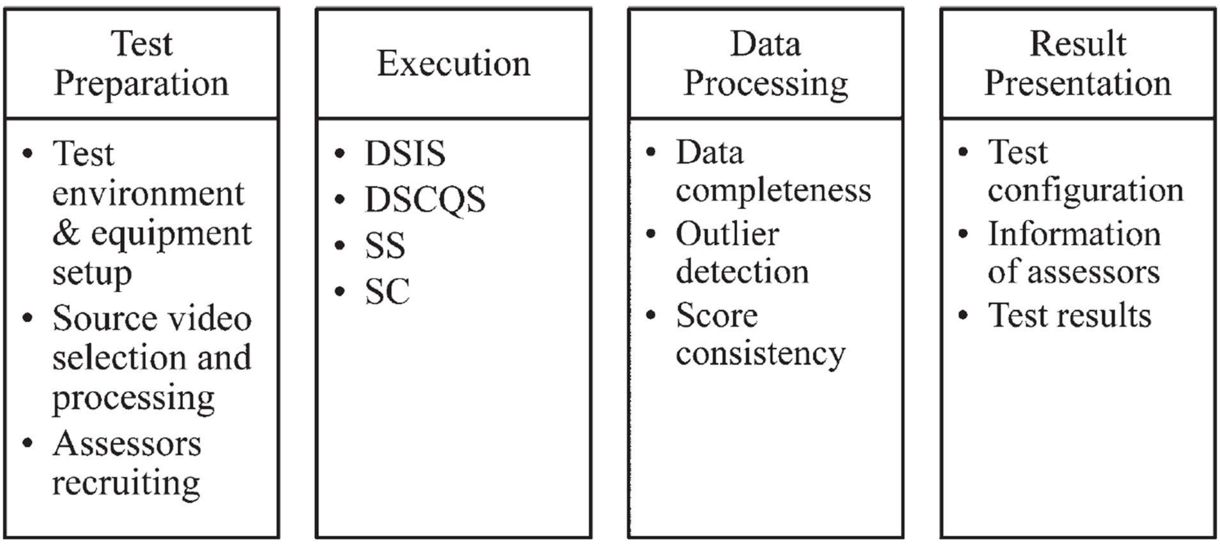


圖 33 主觀測試之流程[83]

---

[120]ITU-T Rec. P.910, Subjective video quality assessment methods for multimedia applications, 2008.
[121]ITU-T Rec. P.913, Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment, 2014.
[122]ITU-R Rec. BT.500-10, Methodology for Subjective Assessments of the Quality of Television Picture, 2000.

過去已有大量的主觀 QoE 研究探討不同類型的品質因素對於人類感知的影響，其中，一些與 HTTP 自適應串流應用相關的研究結果調查可以參考[123],[124],[125],[126],[127],[128]。這些研究中獲得的主要結論整理如下：

（一） 視訊品質、卡頓（stalling）時間與卡頓頻率是決定整體 QoE 的關鍵因素。

（二） 高解析度影音搭配感人的故事情節能提升觀眾參與度 10%以上。

（三） 當緩衝處理開始即產生負面反應：

1. 觀眾快樂情緒降低 14%
2. 觀眾負面情緒（厭惡及悲傷）平均增加 8%
3. 觀眾感到驚訝的情緒增加 27%
4. 觀眾注意力降低 3%，專注力降低 8%

（四） 很短（低於 400 ms）的卡頓可能不會被察覺，因此對 QoE 幾乎沒有影響。

（五） 明顯可見的卡頓事件會嚴重降低 QoE。

（六） 相同總卡頓時間下，多次較短的卡頓比起少許較長的卡頓更令人不悅。

（七） 較高品質的視訊發生卡頓事件時所造成的 QoE 下降會比低品質的視訊來得嚴重。

（八） 相較於卡頓，觀眾對於初始緩衝的接收度比較寬容。

---

[123]Z. Duanmu et al., “A Quality-of-Experience Index for Streaming Video,” IEEE Journal on Selected Topics in Signal Processing, vo. 11, no. 1, Feb. 2017.

[124]M. Seufert et al., “A survey on quality of experience of HTTP adaptive streaming,” IEEE Comm. Surveys Tut., vol. 17, no. 1, pp. 469–492, Sep. 2014.

[125]M.-N. Garcia et al., “Quality of experience and HTTP adaptive streaming: A review of subjective studies,” Proc. IEEE Int. Conf. Quality Multimedia Exp., Sep. 2014, pp. 141–146.

[126] Akamai, “The Science Behind How Our Bodies React to Video Quality,” April, 2017. [Online]. Available: https://content.akamai.com/gl-en-pg9246-sensum-whitepaper.html

[127]S.S. Krishnan and R.K. Sitaraman, “Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs,” Proc. Internet Measurement Conference (IMC), Boston, MA, USA, Nov. 14-16, 2012.

[128]M.N. Garcia, D. Dytko and A. Raake, “Quality Impact due to Initial Loading, Stalling, and Video Bitrate in Progressive Download Video Services,” Proc. IEEE International Workshop on Quality of Multimedia Experience, Singapore , Sept. 18-20, 2014, pp. 129-134.

（九）感知的遲滯現象造成過去的負面觀看體驗經常會影響未來與整體的 QoE。

（十）時近效應（recency effect）造成靠近視訊尾段的卡頓對於 QoE 有更為明顯的影響。

（十一）受測觀眾經常被視訊內容本身吸引，而非品質的變化。

由於一般使用者在進行測試時容易被影片本身的內容所吸引，反而忽略影片本身品質的變化。因此在進行測試之前，須提醒受測者在測試過程中盡量不要被影片內容吸引，而是要觀察影片品質的變化。

二、 客觀量測－QoE 模型評分

欲建立客觀 QoE 模型，典型的作法是依據先期的主觀 QoE 研究確定影響使用者感受的關鍵因素，再以這些關鍵因素的量化度量指標的函式建立一合理的客觀 QoE 模型，當中保留部份係數以提供自由度（例如權重係數等），最後再以主觀 QoE 的值對該客觀 QoE 模型進行迴歸分析（線性、非線性或對數等），以決定最佳的係數值（即能與主觀 QoE 具有最接近的預測結果）。最後，為了支持模型的合理性，需要再次透過人員實驗進行可靠性驗證。以下就數個客觀 QoE 模型案例進行說明。

（一）Ketykó's Model

Ketykó 等人[129]考量視訊內容（content; c）、畫面品質（picture quality; PQ）、聲音品質（sound quality; SQ）、興趣匹配（matching to interests; i）、流暢度（fluidness; f）與載入速率（loading speed; l）這些度量指標，並給予個別度量指標適當的加權係數（weight），提出了一種混合（Hybrid）QoE 模型：

$$\mathrm{QoE} = 0.213 \cdot c + 0.175 \cdot PQ + 0.170 \cdot SQ + 0.160 \cdot i + 0.153 \cdot f + 0.131 \cdot l$$

[129] I. Ketykó et al., "QoE measurement of mobile YouTube video streaming," *Proc. Workshop on Mobile video delivery,* Firenze, Italy, Oct. 25, 2010, pp. 27-32.

其中個別的加權係數乃是直接對使用者詢問個別度量指標對於 QoE 的影響程度（1～5），再經過統計後取得的平均值。此處“混合（Hybrid）”指的是包含主觀與客觀的因素，在該模型中，視訊內容（c）與興趣匹配（i）兩項是完全主觀的因素（即必須由使用者評定），至於其餘四項則屬於可經由其它 QoS 參數計算得到的客觀因素（即可經由量測數據計算）。比較各項加權係數，可發現該模型強調視訊內容（c）本身對於 QoE 具有重要的影響，然而這項度量卻又無法被客觀地量測，因此在應用上有其侷限。

（二）Mok’s Model

Mok 等人[130]則在僅考量影響視訊串流時間品質（temporal quality）的應用性能指標（包含初始緩衝時間（initial buffering time; T_{init}）、平均再緩衝時間（mean rebuffering duration; T_{rebuf}）與再緩衝頻率（rebuffering frequency; f_{rebuf}））下，提出一較簡單的客觀 QoE 模型如下：

$$\mathrm{QoE} = 4.23 - 0.0672 \cdot L_{ti} - 0.742 \cdot L_{fr} - 0.106 \cdot L_{tr}$$

其中 L_{ti}、L_{fr}與 L_{tr}分別對應到 T_{init}、f_{rebuf}與 T_{rebuf} 等級（分三級）的量化分數，該對應關係如表 7。該模型並未考慮視訊串流空間品質（spatial quality）（例如解析度等）的因素，且特別強調再緩衝頻率對 QoE 的影響。Gómez[131]等人採用該模型在 Android 平臺上開發針對 YouTube 串流視訊的 QoE 評估工具，並配合使用者回報的主觀 QoE 來驗證客觀 QoE 模型預估結果的準確性。驗證的結果是客觀 QoE 模型偏向於過份悲觀，平均而言所預測的 QoE 較使用者 QoE 低了 20%。Gómez 等人解釋可能的原因是原模型是在有線環境下建立，而在後來實驗的無線應用情境下，使用者可能對於連線品質本身就有較寬容的接受度。

[130] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, “Measuring the quality of experience of HTTP video streaming,” Proc. IFIP/IEEE International Symposium on Integrated Network Management, Dublin, Ireland, May 23-27, 2011.

[131] G. Gómez et al., “YouTube QoE evaluation tool for Android wireless terminals,” EURASIP Journal on Wireless Communications and Networking, Dec. 2014.

表 7 應用性能指標的等級與量化分數對應關係[132]

L_{ti}	T_{init}	L_{rf}	f_{rebuf}	L_{tr}	T_{rebuf}
1	0 to 1 s	1	0 to 0.02	1	0 to 5 s
2	1 to 5 s	2	0.02 to 0.15	2	5 to 10 s
3	>5 s	3	>0.15	3	>10 s

（三） Khan's Model

Khan 等人[133]考量在 UMTS 網路應用中的發送端位元率（sender bitrate; SBR）、誤塊率（block error rate; BLER）、平均叢發長度（mean burst length; MBL） 與內容形態（content type; CT）這些度量指標，經由非線性迴歸分析得到下列客觀 QoE 模型：

$$QoE = \frac{\alpha + \beta \times \ln(SBR) + CT \times (\gamma + \delta \times \ln(SBR))}{1 + (\eta \times (BLER) + \sigma(BLER)^2) \times MBL}$$

其中 α, β, γ, δ, η 與 σ 是係數，而 CT 則是依據視訊的空間與時間複雜度來分別。

[132] Vlessing, E. （Sep. 23, 2014）. Netflix offers peek at biz as standoff continues with CRTC. Retrieved from http://mediaincanada.com/2014/09/23/as-crtc-netflix-stand-off-continues-u-s-streamer-offers-sneak-peek-at-business/.
[133]A. Khan, L. Sun and E. Ifeachor, "QoE Prediction Model and its Application in Video Quality Adaptation Over UMTS Networks," *IEEE Transactions on Multimedia,* vol. 14, no. 2, pp. 431–442, April 2012.

（四）User-Viewing Activities Model

Mok 等人[134]觀察到當人感受到視訊品質不佳時經常會觸發一些使用者對播放器介面進行的操作活動（User-Viewing Activities; UVA），例如當遭遇卡頓時通常會觸發使用者按下暫停播放鍵、縮小畫面尺寸或是切換較低解析度（預期心態是這些動作可能有助於加快緩衝速率與縮短緩衝時間），完整考量的 UVA 如表 8 所示，經分析後發現部份 UVA 可能只是隨機的使用者操作行為，並不一定與視訊品質感受有關。其中，兩個關鍵的指標：暫停動作事件次數與縮小畫面尺寸動作事件次數被加入與初始緩衝時間、再緩衝時間與再緩衝頻率等應用性能指標一起進行順序變項邏輯迴歸（ordinal logistic regression）分析，其研究結果顯示 UVA 的加入有助於改善 QoE 預估準確度。

表 8 使用者的操作活動與其涵意[135]

Activities	Tech.	$\Delta \bar{r}^j$	Explicit Meaning	Possible Implicit Meaning
Pause	+	−2.37***	Stop playing the video playback for a short period of time.	More time is needed to buffer the video data.
Resume	−	0.26	Continue playing the paused video playback.	Reach the tolerance limit.
Refresh	+	−1.79	Reload the page and video.	The playback quality is unacceptable, and reloading may help.
Switch to a lower video quality	+	−1.79***	Watch the video with lower picture quality.	Scarify the spatial qualify for the playback smoothness.
Switch to a higher video quality	−	2.26***	Watch the video with better picture quality.	The user thinks the current speed is fast enough for watching with a better quality.
Play with full screen	○	2.42***	Watch the video with a larger size.	The playback is enjoyable.
Return to the normal-size screen	○	−0.74*	Watch the video with a smaller size.	The impairments are annoying.
Forward time shift	−	0.53	Watch the content after the current video position	The video content is not interesting.
Backward time shift	+	−0.58	Replay the content before the current video position.	Replaying the buffered video can result in a smoother playback.
Lost of window focus	+	−0.47	The browser window is covered or minimized.	The user may not be watching the video.
Frequent mouse movement	○	n/a	The user moves the mouse over the screen quickly.	The impairments are annoying.
Infrequent mouse movement	○	n/a	The user does not use the mouse.	The user is enjoying the video.

[134] R. K. P. Mok et al., "Inferring the QoE of HTTP video streaming from user-viewing activities," *Proc. ACM SIGCOMM workshop on Measurements up the stack,* Toronto, Ontario, Canada, Aug. 19, 2011, pp. 31-36.

[135] Zboralska, E. & Davis, C. （2017）. Transnational over-the-top video distribution as a business and policy disruptor: The case of Netflix in Canada. The Journal of Media Innovation, 4（1）, 4-25.

（五） Shen’s Model

Shen 等人[136]考量峰值信雜比（Peak Signal to Noise Ratio; PSNR）的 NR 版本[137]、區塊（block）與模糊（blur）三項指標，再依據不同的內容類別（稍微移動、劇烈移動與豐富色彩情境等）與解析度（176x144、352x288、720x480 等）提出不同的客觀 QoE 模型。

（六） U-vMOS Model

Huawei[138]提出了 U-vMOS（User, Unified, Ubiquitous-Mean Opinion Score for Video）作為視訊品質量測的指標，用於視訊體驗與網路最佳化。其始源可回推至 2012 年由該公司的 iLab 部門進行人因工程實驗，使用眼動儀觀察人們在觀看視訊時的反應，利用收集到的數據建立 U-vMOS 的評估模型（如圖 34 所示）。該模型主要考量視訊品質（Video Quality，以 sQuality 表示）、互動體驗 （Interactive Experience，以 sInteraction 表示）與觀看體驗（Viewing Experience，以 sView 表示）三項主要指標，其個別又與其它參數有關（如圖 35 至圖 37 所示）。由於該模型考量的參數範圍廣，舉例而言螢幕尺寸從 4.5”到 100”、解析度從 360p 到 8K，因此 Huawei 宣稱 U-vMOS 適用於廣泛的視訊服務應用（行動裝置、個人電腦與電視機等）。

[136]Y. Shen et al. “QoE-based Evaluation Model on Video Streaming Service Quality,” Proc. IEEE Globecom Workshop, Anaheim, CA, USA, Dec. 3-7, 2012.
[137]A. Eden, “No-Reference Estimation of the Coding PSNR for H.264-coded Sequence,” *IEEE Trans. on Consumer Electronics,* vol. 53, no. 2, pp. 667-674, May 2007.
[138]Huawei, Video Experience-based Bearer Network (White Paper), Aug. 31, 2016.

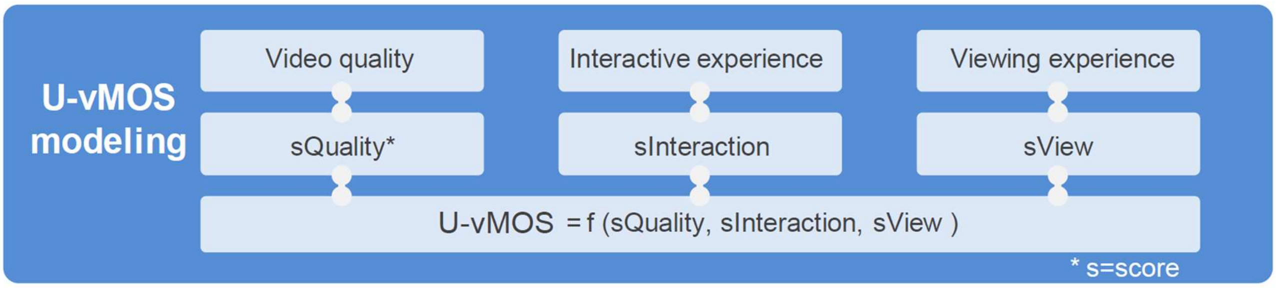


圖 34 U-vMOS 的評估模型[139]

圖 35 sQuality 相關參數

圖 36 sInteraction 相關參數[139]

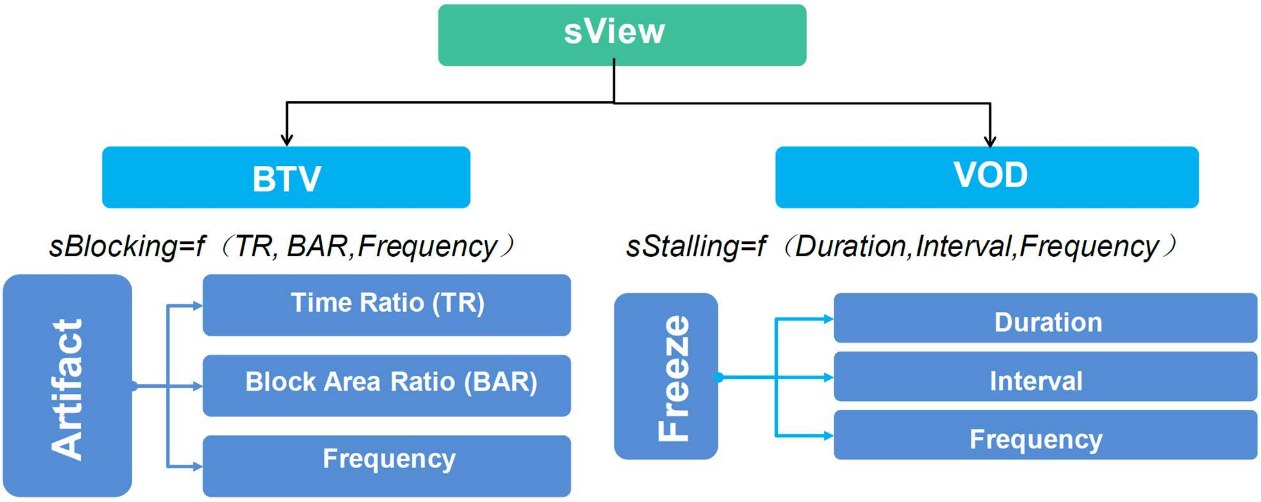


圖 37 sView 相關參數[139]

[139] Huawei, Video Experience-based Bearer Network （White Paper）, Aug. 31, 2016.

（七）5.2.7 Mobile U-vMOS Model

專注考量在小尺寸螢幕上觀看點播視訊，Huawei 的 mLab 部門與牛津大學 （Oxford Univ.）及北京大學（Peking Univ.）合作研究此類應用中影響用戶實際體驗的因素，其發現影響整體用戶體驗最關鍵的三個參數分別是視訊解析度 （以 sQuality 表示）、初始緩衝時間（以 sLoading 表示）與卡頓時間（以 sStalling 表示），其中第一項參數主要反映視訊源本身的空間品質，後二項參數則反映了網路傳輸對時間品質造成的影響，並據此提出一適用於小尺寸螢幕應用的視訊品質量測指標 Mobile U-vMOS [140,141,142]：

$$\text{vMOS} = \text{sQuality} \times \left[\frac{1}{2.5} \times （0.23 \times \text{sLoading} + 0.27 \times \text{sStalling}）\right]$$

其中，sQuality 與實際解析度間的對應關係如表 9 所列，而 sLoading 及 sStalling 個別與初始緩衝時間及卡頓時間比（Stalling Ratio）的對應關係如表 10 所列。

表 9 sQuality 取值列表 [142]

分辨率	sQuality 分值	分辨率	sQuality 分值
5K	5.0	720P	4.0
4K	4.9	480P	3.6
2K	4.8	360P	2.8
1080P	4.5		

表 10 sLoading 與 sStalling 取值列表 [142]

分值	初始緩沖（Sloading）/ms	卡頓占比（sStalling）/%
5	100	0
4	1000	5
3	3000	10
2	5000	15
1	10000	30

[140]Huawei, Mobile Bearer Network Requirements for Mobile Video Services (White Paper), 2016.
[141]Huawei, Mobile Video Service Performance Study (White Paper), June 2015.
[142]陈楚雄, 柯江毅與覃道満, “视频业务体验评估和优化提升探讨 (Discussion on Video Service Experience Evaluation and Optimization),” 邮电设计技术, no. 2, pp. 17-23, 2017.

與第五節之二（六）介紹的 U-vMOS 模型相較，Mobile U-vMOS 模型分別以 sLoading 與 sStalling 取代了複雜的 sInteraction 與 sView，並將 sQuality 化簡為只考慮解析度因素（前提是小尺寸螢幕應用中尺寸差異不大），故可將 Mobile U-vMOS 模型視為是 U-vMOS 模型的一個精簡子集，專為行動裝置應用而設計。

（八） RST-V Model

ITU-T J.343.1 標準[143]考量在有加密的視訊位元流（encrypted bitstream）限制前提下，對於 HDTV 與基於 IP 封包傳送的視訊服務之客觀品質量測方法。其建議了一種「混合無參考加密」（Hybrid No-Reference Encrypted; Hybrid-NRe） 參考模型以適用此類應用的客觀品質量測，如圖 38 與圖 39 所示，此處所謂的混合（Hybrid）是指同時參考視訊位元流數據與經解碼後的視訊資料，而加密（Encrypted)則是指該視訊位元流數據的本體(payload）是有經過加密（亦即無法利用 payload 資訊），僅能使用標頭（header）資訊。

[143] ITU-T Rec. J.343.1, Hybrid-NRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of encrypted bitstream data, 2014.

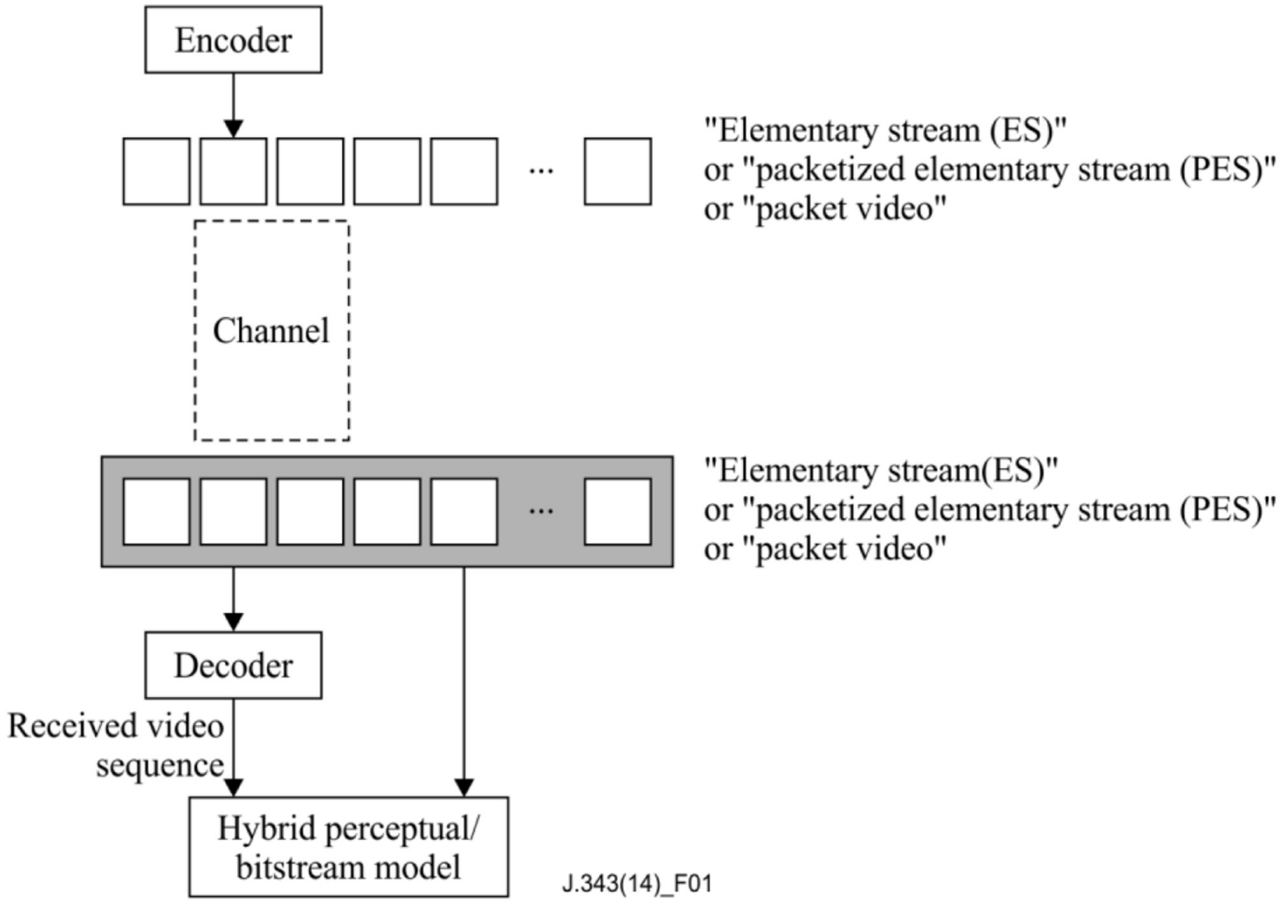


圖 38 Hybrid-NRe / Hybrid-NR 參考模型核心觀念方塊圖 [98]

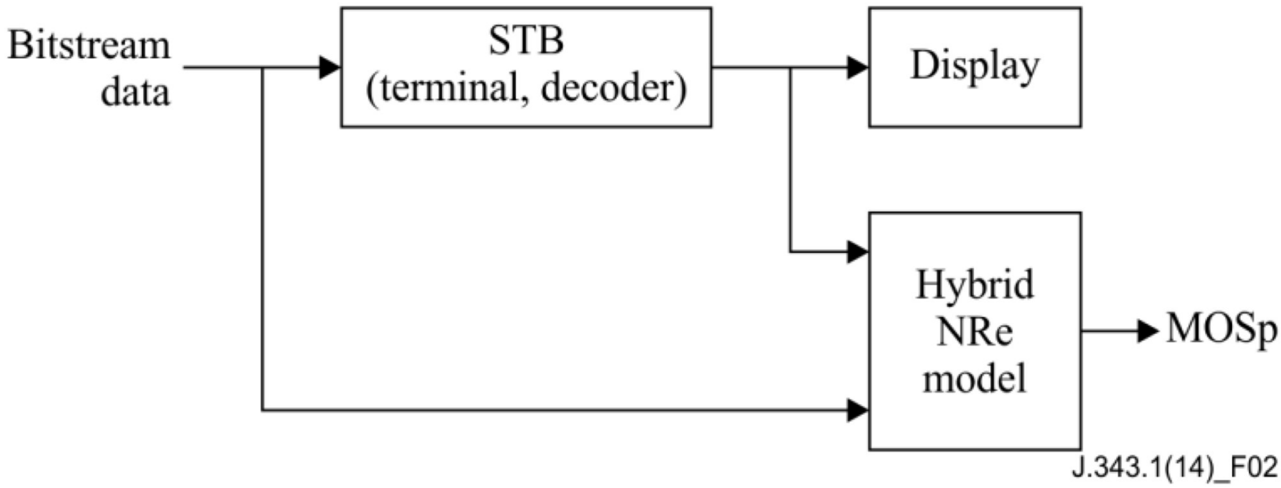


圖 39 Hybrid-NRe 參考模型方塊圖 [99]

RST-V Model（Annex A of [144]）便是遵循 ITU-T J.343.1 標準規範的 Hybrid-NRe 參考模型所設計出來的一個客觀 QoE 模型，其由視訊位元流數據的標頭資訊抽取視訊影格高度（video frame height）、影格編號（frame number）、影格相對時間（frame relative time）、即時串流協定封包序號（RTP sequential number）、RTP 時間戳記（timestamp）與 UDP 資料長度（length）等參數，另由解碼後的視訊資料以視訊處理演算法取得視訊解析度（video frame resolution）、移動統計值（motion statistics）、影格間差異統計值（interframe difference statistics）、空時複雜度統計值（spatio-temporal complexity statistics）、影格顯示時間 （frame display time）與場景切換統計值（scene change statistics）等，再以這些參數個別以演算法估計視訊編碼品質（coding quality）（以 q_cod 表示，其值位於[0, 1]區間）與傳輸品質（transmission quality）（以 q_trans 表示，其值位於[0, 1]區間），最後以兩者的乘積並調整至[1, 5]區間作為視訊品質的量測結果：

$$\mathrm{MOS} = 4 \times \mathrm{q\_cod} \times \mathrm{q\_trans} + 1$$

相關的 RST-V Model 測試驗證結果可以參考由視訊品質專家群組（Video Quality Experts Group; VQEG）所發表的最終測試報告[82]。

[144]ITU-T Rec. J.343.1, Hybrid-NRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of encrypted bitstream data, 2014.

（九）YHyNRe Model

如同第五節之二（八）介紹的 RST-V Model，YHyNRe Model（Annex B of[45]）也是遵循 ITU-T J.343.1 標準規範的 Hybrid-NRe 參考模型所設計出來的一個客觀 QoE 模型，其主要由封包總數（total number of packets）與丟包數（number of packet loss）透過查表方式快速計算出一初始的視訊品質指標（video quality metrics; VQM），再經由將傳輸錯誤導致的各種不同的失真效應納入考量以修正視訊品質指標結果，這些效應包含嚴重傳輸錯誤造成的綠區塊（green block）、停滯（freeze）、區塊（blcok）、模糊（blur）與重覆區塊（repeating block）等，乃經由視訊處理演算法偵測。最後，依據某些效應落在特定範圍的條件成立時再次對視訊品質指標結果進行後處理（post-processing）修正。相關的 YHyNRe Model 測試驗證結果可以參考由 VQEG 所發表的最終測試報告[145]。

（十）YHyNR Model

有別於 ITU-T J.343.1，ITU-T J.343.2 標準[146]則是考量在無加密的視訊位元流前提下，對於 HDTV 與基於 IP 封包傳送的視訊服務之客觀品質量測方法。其建議了一種「混合無參考」Hybrid No-Reference; Hybrid-NR）參考模型以適用此類應用的客觀品質量測，如圖 38 與圖 40 所示，並假設視訊位元流數據的標頭（header）與本體（payload）資訊都可以取得使用。

[145]Video Quality Experts Group, Hybrid Perceptual/Bitstream Validation Test Final Report, 2014.
[146]ITU-T Rec. J.343.2, Hybrid-NR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of non-encrypted bitstream data, 2014.

YHyNR Model 便是遵循 ITU-T J.343.2 標準規範的 Hybrid-NR 參考模型所設計出來的一個客觀 QoE 模型，其方法與第五節之二（九）介紹的 YHyNRe Model 類似，最大差別在於 YHyNR Model 使用量化參數（quantization parameter）與錯誤面積（error area）來計算初始的視訊品質指標，因為在無加密的前提下預設可以由視訊位元流數據的 payload 取得這些參數。相關的 YHyNR Model 測試驗證結果可以參考由 VQEG 所發表的最終測試報告[147]。

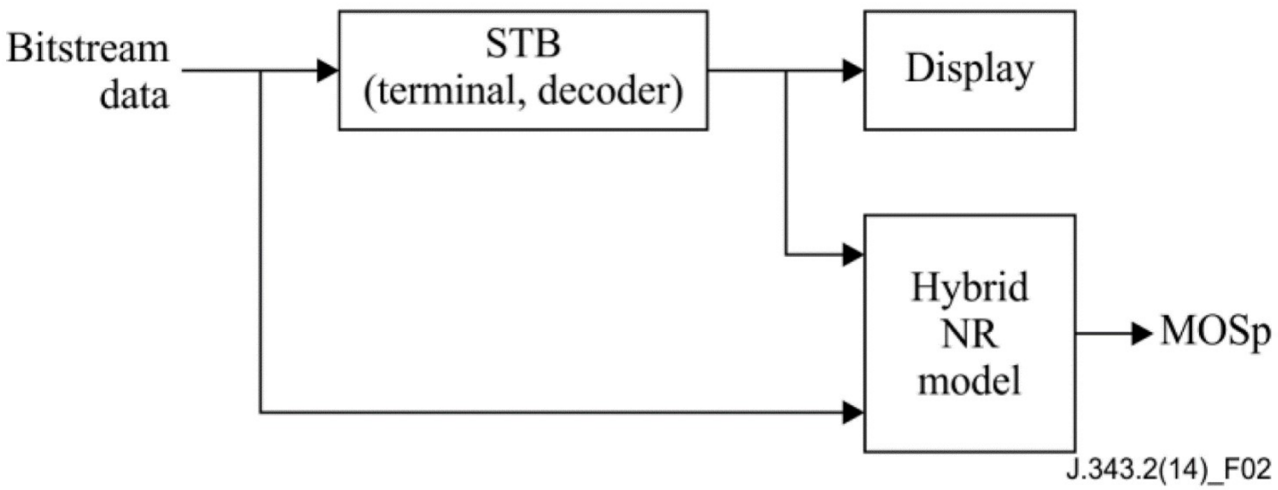


圖 40 Hybrid-NR 參考模型方塊圖 [100]

[147]Video Quality Experts Group, Hybrid Perceptual/Bitstream Validation Test Final Report, 2014.

第六節 量測方法綜合分析與建議

一、 量測方法之優缺點比較分析

根據[148]的分析，在設計 QoE 量測方法時，必須考慮許多因素。其中在資料採集（data acquisition）部份至少要考量三個問題：

（一） 必須採集哪些（which）資料 – 由所選擇的度量指標（metrics）決定。例如客觀 QoE 量測方法所採集的資料都必須是可客觀量測的度量指標，並且可根據過往主觀 QoE 研究中獲得的結論來選擇待測度量指標。

（二） 在何處（where）進行資料採集 – 由傳輸架構中能受量測端控制的節點層級決定將。例如在用戶端應用層或網路層、或是在路由器端。明顯地，在最接近用戶端的應用程式進行資料採集，可取得與用戶實際感受最接近的度量指標結果或是事件紀錄。

（三） 如何（how）執行資料採集 – 取決於“where”問題的答案，可能依據封包紀錄（標頭或主體）蒐集網路流量資料或甚至設計應用程式根據影音解碼資訊取得用戶感受到的延遲、卡頓等事件參數。

[148]R. Serral-Gracià et al., “An Overview of Quality of Experience Measurement Challenges for Video Applications in IP Networks,” Proc. International Conference on Wired/Wireless Internet Communications, Luleå, Sweden, June 1-3, 2010, pp. 252-263.

第四節中所介紹的 QoS 量測方法，主要是在用戶端應用層或傳輸層（where） 依據封包紀錄（how）量測網路品質指標（which），再經過映射函式間接地預估 QoE（主要偏時間品質的部份），屬於量測方法初始發展階段的作法。優點是只要網路品質指標定義明確，保證可以透過正確的量測方法達成量測目的，且量測結果是完全客觀不受使用者主觀意識的影響。缺點則是完全沒有考慮視訊源本身品質與終端播放器與顯示器對 QoE 的影響，終究是不足以直接反應用戶對於影音品質的真實感受。儘管如此，網路品質指標對於傳輸問題的診斷確實有其必要性，因此實務上不論使用何種視訊品質量測方法，大都伴隨網路品質指標的同時量測與記錄。

在第五節之一中所介紹的主觀 QoE 量測方法，主要由真人（where）依據主觀的觀看體驗與感覺（how）給予量化的評估分數（which）。優點是真實地反應用戶的感受，缺點是必須由真人參與執行，不適用於自動化量測應用。再者，嚴謹的主觀測試與用戶反應研究通常需要在實驗室環境下（以便控制環境因素），不方便性與成本昂貴也是其缺點。

在第五節之二中所介紹的客觀 QoE 量測方法，主要是在用戶端應用層與傳輸層（where）依據影音解碼資訊與封包紀錄（how）量測應用性能指標、網路品質指標與其它外部因素（which），再經過預定的客觀 QoE 模型預估 QoE（可完整考量空間品質、時間品質、量化品質與內容型態等部份）。優點是綜合考量整個傳輸架構上對 QoE 造成影響的因素，較能接近完整地反應用戶的真實感受。缺點則是欲量測的度量指標越完整也會造成客觀 QoE 模型複雜度越高，在有限的資源下實作的可行性也相對越低，故必須適當地選擇關鍵的度量指標。

相較之下，以低成本、可自動化、高可靠性等訴求來看，客觀 QoE 量測這一類方法是比較可行的，這也呼應了第三節二中說明的量測方法演進過程。

二、 量測方法之實施建議

在設計（或選擇）量測方法之前，必須先確定目標應用的情境與需求，不同的情境與需求將影響評估指標與量測方法的決定。觀察**目前大多數使用者的收視習慣為使用行動裝置(即小尺寸螢幕)透過行動寬頻或無線網路觀看影片**，建議可以此作為目標應用情境。再者，為了讓一般用戶也能量測服務品質（包含手動與自動執行），建議將低成本與可自動化列入需求。如此，根據第六節一的比較分析，建議**在用戶端執行客觀 QoE 量測並以開發用戶端應用程式作為量測工具**，以利直接蒐集與用戶實際感受最接近的度量指標結果或是事件紀錄，同時也可以一併**記錄用戶端量測到的網路 QoS 參數**（從用戶端主動量測 RTT、RTT 變異、丟包率與吞吐量等參數）。

在度量指標的考量上，可依允許的複雜度選擇下列參數的組合：

（一）視音訊編碼配置（Profile）與等級（Level）

1. 視訊解析度 （Resolution）
2. 視訊影格率 （Frame Rate）
3. 量化精確度 （No. of bits）

（二）螢幕尺寸、觀看距離

（三）初始緩衝時間和卡頓（重新緩衝）事件（時間、分佈與頻率）

（四）首頁/節目選單載入時間

（五）使用情境

依據第五節之一中摘要的主觀測試研究結論，建議至少考量 Mobile U-vMOS Model 中所採用的**視訊解析度、初始緩衝時間與卡頓時間**比作為關鍵度量指標，再依所選擇的關鍵度量指標建立一與用戶感受正相關的客觀 QoE 模型。藉由控制在指定的地點與時間（或週期性地）使用客觀 QoE 模型蒐集預定數量的量測結果，經由統計分析後可以得到代表用戶的體驗品質的結果，例如就特定影片可以分析：

（一）該影片本身提供的最高解析度對應之客觀 QoE

（二）全程無卡頓下的最佳解析度與該解析度對應之客觀 QoE

另一方面，基於 QoS 與 QoE 的量測結果，可以使用**錯誤! 找不到參照來源。**中提出的雷達圖以視覺化檢視 QoE 和網路 QoS 之間的關聯性。以圖 6-1 的範例說明，該雷達圖欲顯示 MOS（分三級，以不同灰階顏色表示）與網路 QoS（頻寬（BW）：1 Mbps～unlimited、丟包率（Loss Rate）：0%～8% 與時間延遲（Delay）：0 ms～100 ms） 之間的關係。三道扇區（sector）AB、BC 與 CA 分別固定其中一項 QoS 參數為最佳值：unlimited BW、0 ms Delay 與 0% Loss Rate。三條軸線 A、B 與 C 則分別表示沿著該軸從圓心向外延伸將從最佳到最差方式改變 Delay、Loss Rate 與 BW 數值，且其效應分別只作用於扇區 dd'、ee' 與 ff'。舉例而言，在扇區 Ad' 中，BW 是 unlimited、Loss Rate 以順時針方向增加（從 0%到 8%）、而 Delay 則從圓心向邊緣方向增加（從 0 ms 到 100 ms），如此即可觀察 MOS 與其中兩項 QoS 參數 Loss Rate、Delay 之間的交互關係。從圖 41 的扇區 AB 主要是深色區塊（對應較低的 MOS）可看出在此範例中 Loss Rate 與 Delay 是影響 QoS 的關鍵因素。除了第三節之五中介紹的驗證方式，透過觀察 QoS 與 QoE 的關聯性也可作為驗證所提 QoE 量測方法的合理性一種方式，例如以採用自適性串流技術為前提下，量測到的 QoE 能適當地反映出當下網路 QoS 的狀態。

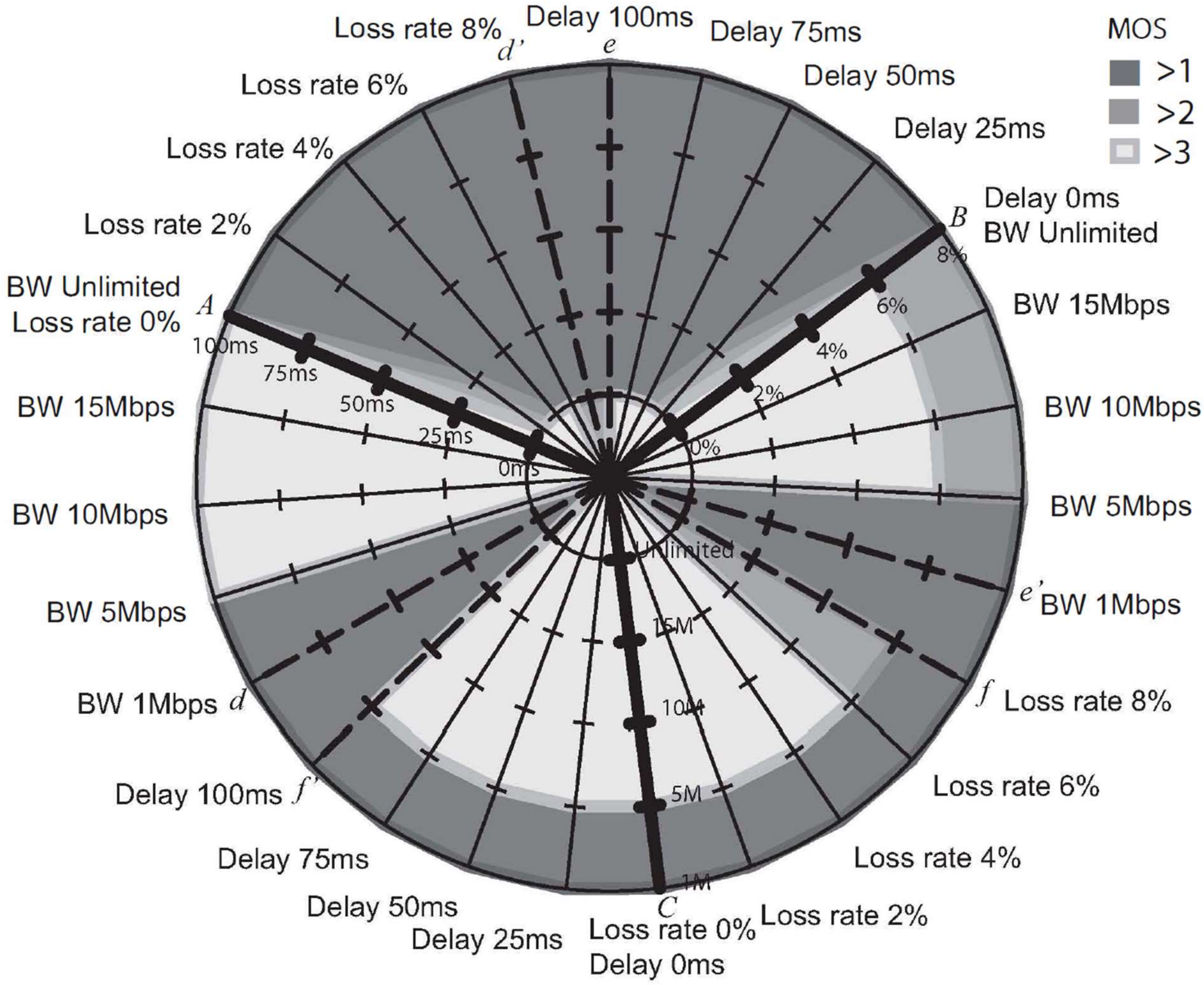


圖 41 網路 QoS 與 QoE 之間對應關係的雷達圖[149]

[149]R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the quality of experience of HTTP video streaming," Proc. IFIP/IEEE International Symposium on Integrated Network Management, Dublin, Ireland, May 23-27, 2011.

第四章 布建量測與結果

第一節 量測工具說明

在設計（或選擇）量測方法之前，必須先確定目標應用的情境與需求，不同的情境與需求將影響評估指標與量測方法的決定。觀察目前大多數使用者的收視習慣為使用行動裝置（即小尺寸螢幕）透過行動寬頻或無線網路觀看影片，為了讓一般用戶也能量測服務品質（包含手動與自動執行），應將低成本與可自動化列入需求。如此，根據第三節之一的比較分析，本研究案選擇在用戶端執行客觀 QoE 量測並以開發用戶端應用程式作為量測工具，以利直接蒐集與用戶實際感受最接近的度量指標結果或是事件紀錄，同時也可以一併記錄用戶端量測到的網路 QoS 參數（從用戶端主動量測 RTT、RTT 變異、丟包率與吞吐量等參數）。

在度量指標的考量上，本研究案採用電信技術中心既有研發之「視訊服務品質綜合指標(Video Mean Opinion Score;vMOS)」：

vMOS = f (視訊解析度, 初始緩衝時間, 卡頓率)

該指標考量了視訊解析度、初始緩衝時間與卡頓率對於 QoE 的綜合影響，其中第一項參數主要反映視訊源本身的空間品質，後二項參數則反映了網路傳輸對時間品質造成的影響。而根據綜多主觀測試研究所得到的結論[150],[151],[152],[153],[154],[155]，視訊品質、卡頓時間與卡頓頻率的確是決定整體 QoE 的關鍵因素，對於本案度量指標的選擇有了良好的支撐。

---

[150]Z. Duanmu et al., “A Quality-of-Experience Index for Streaming Video,” IEEE Journal on Selected Topics in Signal Processing, vo. 11, no. 1, Feb. 2017.
[151]M. Seufert et al., “A survey on quality of experience of HTTP adaptive streaming,” IEEE Comm. Surveys Tut., vol. 17, no. 1, pp. 469–492, Sep. 2014.
[152]M.-N. Garcia et al., “Quality of experience and HTTP adaptive streaming: A review of subjective studies,” Proc. IEEE Int. Conf. Quality Multimedia Exp., Sep. 2014, pp. 141–146.
[153] Akamai, “The Science Behind How Our Bodies React to Video Quality,” April, 2017. [Online]. Available: https://content.akamai.com/gl-en-pg9246-sensum-whitepaper.html
[154]S.S. Krishnan and R.K. Sitaraman, “Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs,” Proc. Internet Measurement Conference (IMC), Boston, MA, USA, Nov. 14-16, 2012.
[155]M.N. Garcia, D. Dytko and A. Raake, “Quality Impact due to Initial Loading, Stalling, and Video Bitrate in Progressive Download Video Services,” Proc. IEEE International Workshop on Quality of Multimedia Experience, Singapore , Sept. 18-20, 2014, pp. 129-134.

在量測工具的實作上，本研究案所需的相關參數含下列測項之實作：

一、網路性能指標（QoS）量測

二、RTT, Jitter, Packet Loss Rate 與 Throughput 等

三、網路問題診斷

四、DNS Lookup 與 Traceroute

五、影音服務體驗品質（QoE）量測

其 vMOS 量測部份乃透過 ExoPlayer API 解析 DASH 串流（以 YouTube 服務為例），監看播放狀態事件來計算所需的參數。此實作方案的優點與限制如下：優點：

一、支援自適性串流（DASH、HLS 與 MSS）

二、不同 Android 版本間的變異較小

三、支持 Android 4.4（API 等級 19）以上的 Widevine 通用加密

四、開發者可以透過客製化來使 ExoPlayer 符合自己的需求

支援數位版權管理（Digital rights management; DRM）的影片限制：ExoPlayer 的標準音頻和視頻組件依 Android 的 MediaCodec API，這是在 Android 4.1（API 16）中發佈的。因此在低於 Android 4.1 版本的裝置上無法使用 ExoPlayer。

第二節 布建量測規劃

一、 量測標的

本研究案以實際量測民眾透過有線廣播電視系統、電信事業固網與行動寬頻等網路媒介使用行動裝置觀看 YouTube 點播影音的服務品質 (vMOS) 作為目標，考量的應用情境包含：

（一）固網環境：行動裝置以 WLAN 連線、定點測試

（二）行網環境：行動裝置以 4G/WLAN 連線、移動/定點測試

藉由控制在指定的地點 (區) 與時間 (或週期性地) 使用所開發之量測工具蒐集預定數量的量測結果 (vMOS 與 QoS 指標)，經由統計分析後可以得到代表用戶的體驗品質的結果。

二、 場測布建規劃

由於本案的場測目的在於確認所開發的量測工具在較具規模 (範圍與時間) 的佈建與自動化量測應用中是可行的，並非針對特定業者進行性能評估或評價，因此關於用戶的取樣，主要考量「能兼具地點、網路型態與業者等多樣性」前提下進行布點測試，以便確認所開發之量測工具在各種應用情境下皆適用。依此前提，本案最終選擇了 222 個用戶進行量測，主要分佈於北/中/南三地區，量測數據量比例 (如圖 46 所示) 約為 2(北)：1(中)：1(南)。這些採樣用戶總共涵蓋了 5 家 4G 業者網路、2 家固網業者網路與 3 家有線電視業者網路，各種網路型態的量測數據量在個別地區所佔比例可參考圖 47 至圖 49。

對於所選擇的每一採樣用戶，在受測期間即安排固定週期的自動測試程序，期間每隔 2 小時即自動執行一次測試流程，該測試流程中依序執行觀看一部業者影片與一部共同影片“big buck bunnie”的 vMOS 評估，評估過程中會針對該測試影片的每一種支援的解析度都進行一次 vMOS 量

測，在進行 vMOS 評估時也同時進行部份較不佔頻寬的 QoS 測項 (RTT、Jitter、DNS Lookup 與 Traceroute) 量測，以便忠實反映串流影音流量當下的網路品質又不至於影響網路品質，待 vMOS 評估完成後再進行佔用頻寬的 QoS 測項 (Throughput) 量測。估計每單次自動測試流程約耗時 7 ~ 15 分鐘 (實際與網路傳輸條件有關)。依此取樣頻率設計，測試預計將可蒐集到超過 20,000 筆總量測數據，之後再依各分析項目取其中適用的子集合數據進行分析。依統計學，一個成功的調查必須至少 1,068 份合格樣本；依本案規劃之樣本蒐集數量至少達 10,000 筆樣本，已超過統計學上成功調查的定義[156]。所有測試結果回傳至敝中心伺服器並建立查詢資料庫，提供之視覺化方式呈現所需的各筆測試結果資料如圖 42。每一筆資料包含了測試的日期與時間 (datetime)、影片解析度 (chosen_video_quality)、影片名稱 (video_name)、vMOS (vmos_value)、ISP 名稱 (isp)、連網介面 (network_type)、終端裝置之廠牌型號 (ue_vendor)、下行吞吐量 (spd_bench)、IMEI (imei) 與地區 (area) 等，之後再依各分析項目取其中適用的子集合數據進行分析，擬進行分析的項目與內容彙整如表 11。

[156]黃文璋, “統計裡的信賴”, 數學傳播 30 卷 4 期, pp. 48-61, https://web.math.sinica.edu.tw/math_media/d304/30406.pdf

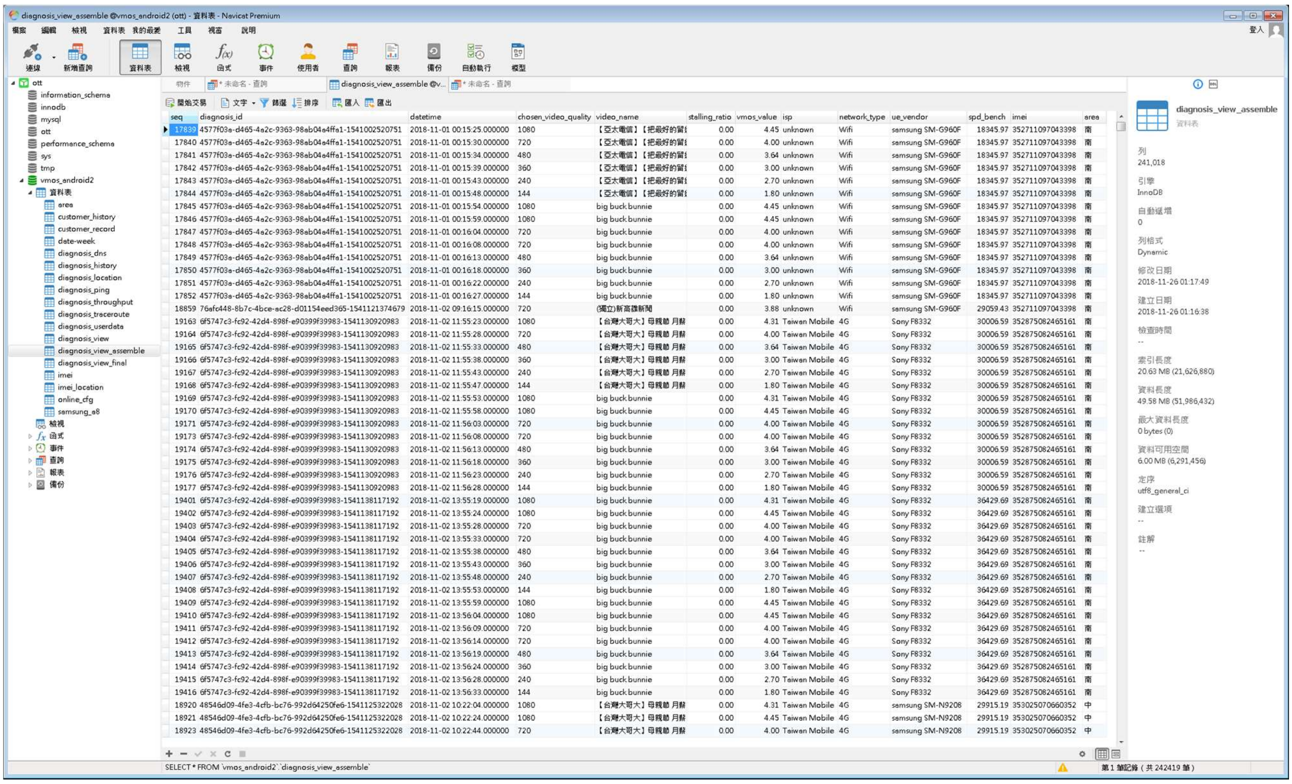


圖 42 測試結果資料(資料來源:本研究整理)

表 11 擬進行分析的項目與內容

內容 項目	固定的條件	控制（因）變數	觀察（應）變數	結果分析與呈現方式
特定影片本身提供的最高解析度對應之 vMOS	• 行動裝置廠牌型號 • 測試所在地點 • 測試發生時間 • 連網介面	• 測試影音媒體	最高解析度對應之 vMOS	對取樣結果取平均值並以柱狀圖呈現各測試影片之結果
特定影片全程無卡頓下的最佳解析度與該解析度對應之 vMOS	• 行動裝置廠牌型號 • 測試所在地點 • 測試發生時間 • 連網介面	• 測試影音媒體	全程無卡頓下的最佳解析度與該解析度對應之 vMOS	對取樣結果取平均值並以柱狀圖呈現各測試影片之結果
用戶體驗品質與上網時間關聯性	• 測試影音媒體 • 行動裝置廠牌型號 • 測試所在地點 • 連網介面	• 測試發生時間	不同時段量測的 vMOS	對取樣結果取平均值並以曲線圖呈現各測試時段之結果
用戶體驗品質與所在位置的關聯性	• 測試影音媒體 • 行動裝置廠牌型號 • 測試發生時間 • 連網介面	• 測試所在地點	不同地點量測的 vMOS	對取樣結果取平均值並以柱狀圖呈現各測試地點之結果
串流影音服務品質的區域性差異	• 測試影音媒體 • 行動裝置廠牌型號 • 測試發生時間 • 連網介面	• 測試所在地區	不同地區量測的 vMOS 平均值	對取樣結果取平均值並以柱狀圖呈現各測試地區之結果

(資料來源:本研究整理)

第三節 量測驗證方法

本研究對於 QoE 量測方法的性能皆需要經過驗證,可靠且有效的量測方法工具才有實際被應用的價值。一般而言,對於測量工具的檢定最常見的便是可靠性 (Reliability) (信度) 與有效性 (Validity) (效度) 的分析。在 QoE 量測中欲度量的單一項目 “使用者感受” 是完全主觀的,基本上無法得到標準答案(即真實值),因此在這領域中實務上通常是僅能透過該項目量測結果 (即觀察值)的某些統計分析來展現信度與效度。本研究為求嚴謹度,先透過統計上量測驗證方法在信度與效度方面的做法於本節說明,並於下節中套用下列的驗證方法加以驗證量測數據,以下分別就信度與效度兩個面向說明在 QoE 量測中普遍用來評估相關性能的方式。

一、 可靠性 (Reliability) (信度) 驗證

信度即可靠性,意指量測結果的一致性或穩定性。根據信度統計模型[157],量測結果 (觀察值) 的平均值 μ_o 與變異量 σ_o^2 皆可以被分解為對應到真實值(系統項; systematic) 與誤差項 (error) 這兩個部份:

$$\text{平均值 } \mu_o = \mu_t + \mu_e$$

$$\text{變異量 } \sigma_o^2 = \sigma_t^2 + \sigma_e^2$$

其中誤差項 σ_e^2 是與量測標的無關的因素造成,該值將影響量測結果的可靠性。故將真實變異量與觀察變異量之比值定義為信度:

$$\text{信度 } \mathrm{r} = \sigma_t^2 / \sigma_o^2$$

[157] J.P. Peter, “Reliability: A review of psychometric basics and recent marketing practices,” *Journal of Marketing Research,* Feb. 1979.

通常 r > 0.7 即表示具有信度。由於無法直接估計真實 (系統項) 變異量 σ_t^2，實務上必須採用某些信度量測方法來決定觀察變異量中來自於系統本身的成份，例如其中一種 “Test-retest” (再測信度) 方法就是利用計算前後兩次量測 (間隔一段時間) 結果間的相關係數作為信度[158]。

就 QoE 量測而言，信度在於表示相同條件下 (影片源、網路傳輸環境與終端設備等) 反覆進行量測 QoE 的結果是否前後一致。量測結果間的彼此誤差越小、相關性越高，則信度越高。因此，可以採用再測信度量測方法，利用計算前後兩次量測間的相關係數作為信度。必須強調的是，在群眾測試 (crowdtesting) 應用中，網路傳輸環境乃隨時間變化，再測信度方法所量測的結果可能會受影響。

二、　有效性 (Validity) (效度) 驗證

效度即正確性，意指量測工具確實能測得所欲測量標的程度。效度越高，表示量測結果越能顯現其所欲測量對象的真正特徵。根據效度統計模型 ，真實值 (系統項) 變異量 σ_t^2 可再進一步被剖析為與測量特質相關的共同變異量 σ_co^2 及與測量特質無關的其他變異量 σ_other^2 兩部份，因此

$$\text{變異量}\ \sigma_o^2 = \sigma_{co}^2 + \sigma_{other}^2 + \sigma_e^2$$

故將 σ_{co}^2 與 σ_o^2 之比值定義為效度：

$$\text{效度}\ \mathrm{v} = {\sigma_{co}^2}/{\sigma_o^2}$$

158 Types of Reliability. *The Research Methods Knowledge Base.* Oct. 2006. [Online]. Available: http://www.socialresearchmethods.net/kb/reltypes.php

通常 v＞0.7 即表示具有中高程度以上的效度。然而，相較於需要估計 σ_t^2 來決定信度，在效度的決定上 σ_{co}^2 又更難估計，因此並非對每一種量測工具都可以客觀地決定其效度。實務上比較容易的作法是採用效標效度 (criterion validity) 量測方法[159]，屬實證效度的一種，經由收集證據來證明量測工具會與目前已廣為大眾肯定具有效度的測量工具 (即效標) 具有高度相關性。

就 QoE 量測而言，效度在於表示量測 QoE 的結果是否真實反映使用者對服務品質的體驗/感受。在此領域中，通常是將第 2.3.1 節中介紹的主觀的 QoE 評估結果作為效標，在一次量測中同時蒐集主觀 QoE 並與量測工具所得量測結果進行比較驗證。若有需要，可以正規化調整量測結果至與主觀 QoE 相同的尺度 (scale) 或範圍 (range)，以利兩種資料間的比對。常見的驗證方式說明如下：

（一）視覺化相關性－二維散點圖(Scatter Plot)

此方式透過類似下圖的二維散點圖（Scatter Plot）以視覺化方式展現量測結果與主觀的 QoE 兩者間的相關性，圖中橫軸代表直接由用戶評分的 QoE，縱軸則代表其它量測方法得到的量測結果，舉例：假設某一筆測試資料中用戶評分為 4 分且客觀 QoE 模型的量測結果為 3.8 分時，便在圖中座標（x, y）= （4, 3.8）處標記一符號（‘o’）。依此方式將所有測試資料繪製於同一張圖上，即可以視覺化方式展示客觀 QoE 模型的量測結果與用戶評分的相關性，當所有符號 （‘o’）越貼近圖中斜率為 1 的直線（紅線）就表示兩者相關性越高，而相關性越高則代表該客觀 QoE 模型越可靠。

159 L.J. Cronbach and P.E. Meehl, “Construct validity in psychological tests,” *Psychological Bulletin*, vol. 52, pp. 281-302, July 1955.

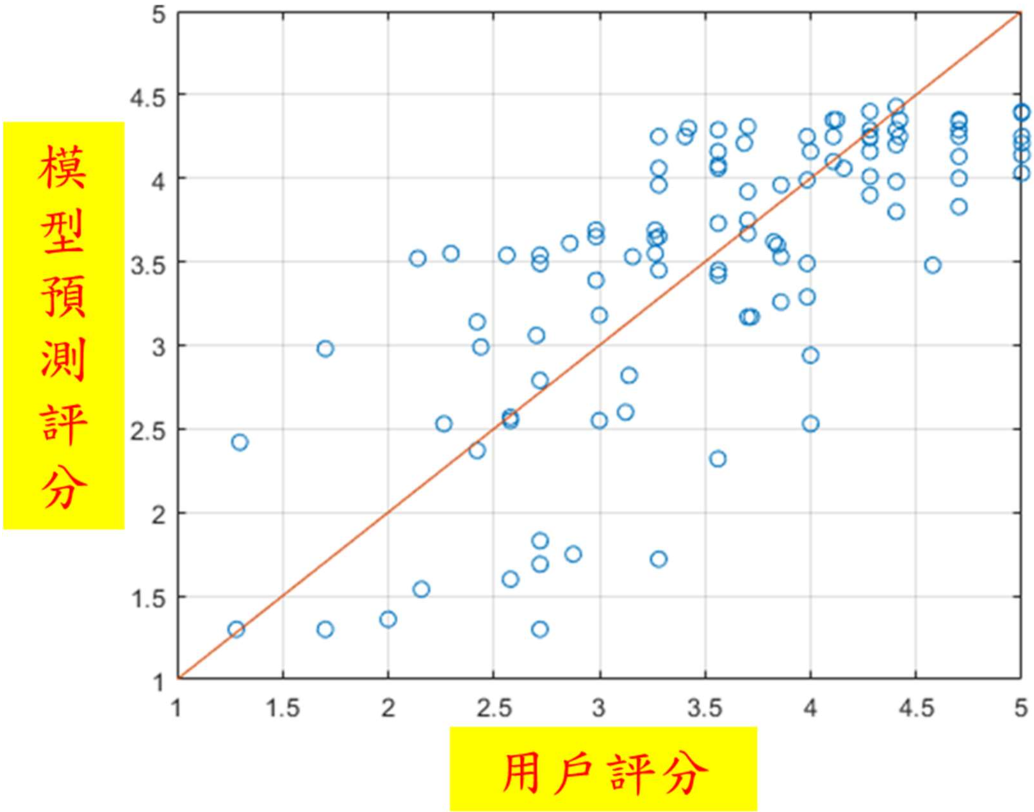


圖 43 二維散點圖（範例）(資料來源:本研究整理)

（二）數值化相關性－皮爾遜相關係數（Pearson correlation coefficient; PCC）

統計上 PCC 定義為兩個變數 X 與 Y 之間的共變異數（covariance）和標準差（standard deviations）的商，即

$$\rho_{X,Y} = \frac{E[(X - E(X))(Y - E(Y))]}{\sqrt{\left[E(X^2) - (E(X))^2\right]\left[E(Y^2) - (E(Y))^2\right]}}$$

該值用於衡量兩個變數之間的相關性（線性相依性），其值介於−1 與 1 之間，值越接近 1 表示兩者正相關性越高。實務上則以樣本平均（sample mean） 取代定義中的總體平均（ensemble average）。

（三）誤差統計

除了前述的相關性，另一種驗證的方式是以誤差統計資料展現兩者間的吻合度，例如以下表的方式呈現完全正確、偏差一個等級、偏差二個等級…佔整體量測結果筆數的比例，也可以一併計算根均方誤差（Root Mean Square Error; RMSE）的值。

表 12 誤差統計作為量測結果驗證 （範例）

	統計結果
實驗數量	120
RMSE	0.615
誤差 1 級	108 （90%）
誤差 2 級	12 （10%）
誤差 3 級	0 （0%）
誤差 4 級	0 （0%）

(資料來源:本研究整理)

第四節 量測數據與驗證

一、 量測數據說明

透過本研究實際布建量測結果目前已超過 20000 筆數據，超過原先預計的 14,400 至 18,000 筆數據。經過 30 天(2081/)的量測，累計了 3144 數據人天數，目前蒐集之數據量已超過原先預計的 100 用戶 15 天量測的 1500 數據人天數。其中使用行網環境行動網路量測的有 2109 數據人天數(已超過原目標的 750 數據人天數)，使用 wifi 環境(包含固網與有線電視之 ISP)量測的有 1035 數據人天數(已超過原目標的 750 數據人天數)，如下表所示。

表 13 依網路介面區分之累計數據人天數

量測網路介面	計畫原目標	量測累計人天數	達成率
行動網路	750	2,109	281.2%
WiFi	750	1,035	138.0%
總合	1500	3,144	209.6%

(資料來源:本研究整理)

而使用 wifi 環境量測之網路服務提供業者又分為固網以及有線電視網路，使用固網累計了 450 數據人天數(已超過原目標的 375 數據人天數)，而有線電視網路累計了 585 數據人天數(已超過原目標的 375 數據人天數)，如下表所示。

表 14 固網與有線電視網路之累計數據人天數

量測網路型態	計畫原目標	量測累計人天數	達成率
固網量測	375	450	120.0%
有線電視網路量測	375	585	156.0%
總合	750	1,035	138.0%

(資料來源:本研究整理)

此次研究共有 222 個用戶進行量測，其中有 104 個用戶量測 15 天以上(已超過原目標 100 個量測用戶 15 天量測目標)，如下圖 44 各用戶量測天數(依天數多寡排序)。

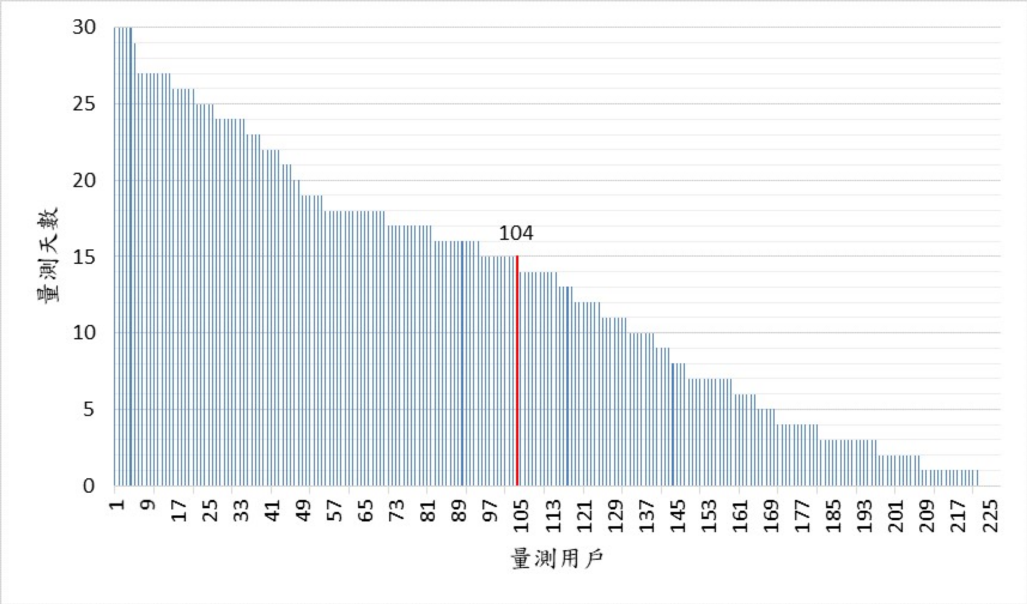


圖 44 各用戶量測天數-依天數多寡排序 (資料來源:本研究整理)

本研究平均 15 天測試人數總共有 210 用戶數(已超過原目標 100)，其中使用行網服務量測平均 15 天測試人數總共有 141 個用戶數(已超過原目標 50)，使用固網量測平均 15 天測試人數總共有 30 個用戶數(已超過原先預計 25)，使用有線電視網路量測平均 15 天測試人數總共有 39 個用戶數(達到原目標 25 個用戶)，如下圖 45 為每日成功量測用戶數的分布圖。

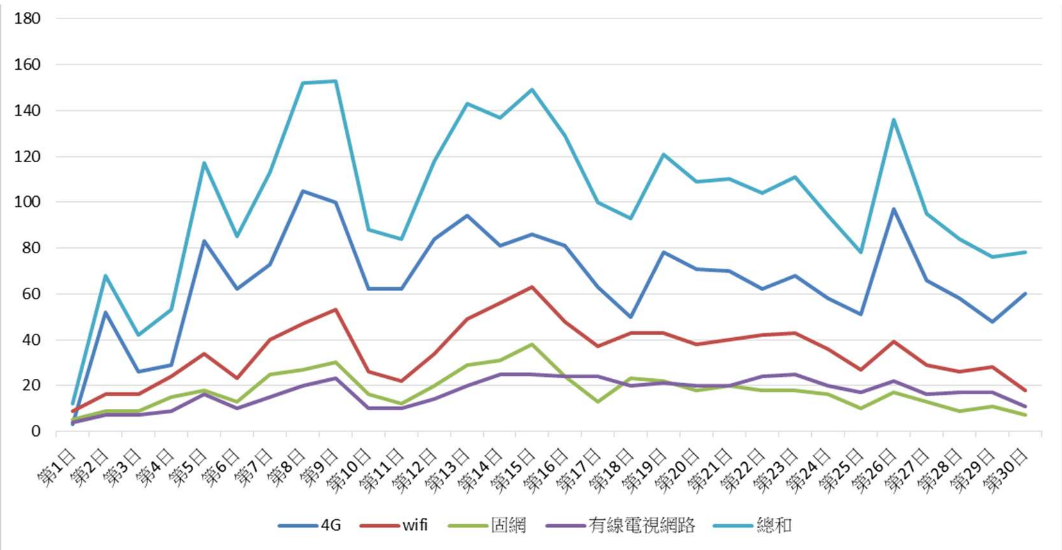


圖 45 每日量測用戶數分布圖(資料來源:本研究整理)

依照區域來劃分，本研究此次量測分布在台灣北、中、南各地進行，共量測 21993 筆量測數據，下圖 46 為依區域劃分量測數據量比例，其中有 11054 筆數據來自北部(占整體的 50%)，有 4865 筆數據來自中部(占整體的 22%)，有 5478 筆數據來自南部(占整體的 25%)，有 596 筆數據來自東部(占整體的 3%)。

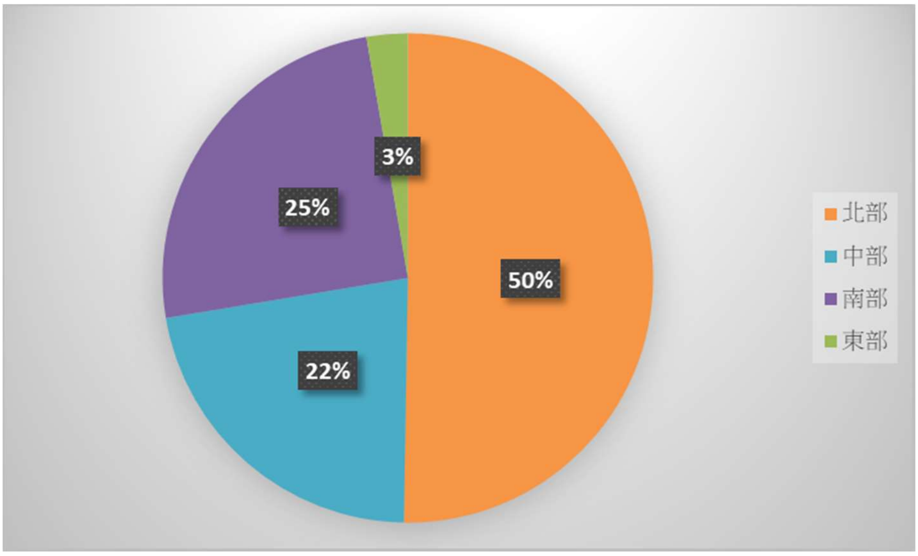


圖 46 依區域劃分量測數據量比例(資料來源:本研究整理)

若單以北部來看量測分布，下圖 47 為北部地區量測數據比例，其中有 7456 筆數據來自行動網路量測(占整體的 68%)，有 2005 筆數據來自固網量測(占整體的 18%)，有 1593 筆數據來自有線電視網路量測(占整體的 14%)。

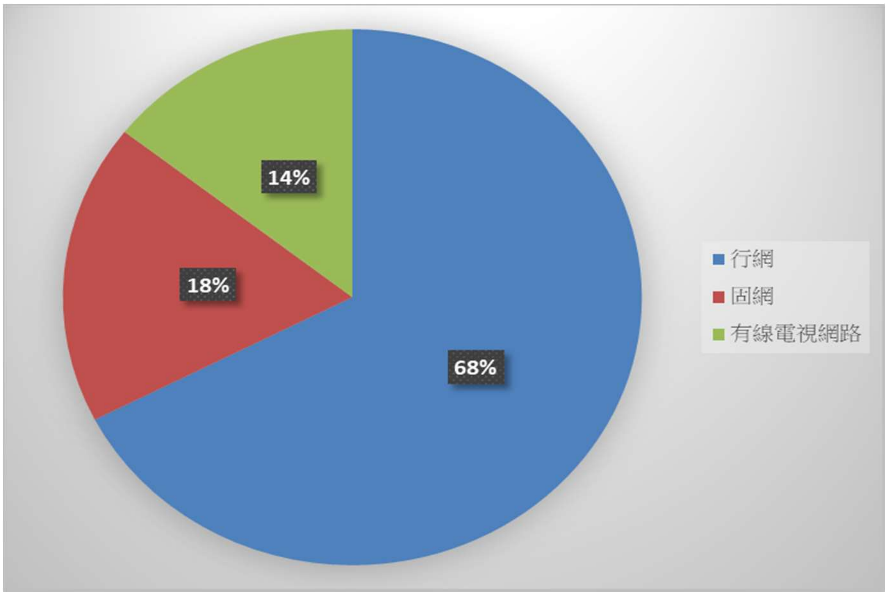


圖 47 量測用戶在北部之各種網路型態百分比(資料來源:本研究整理)

圖 48 為中部地區量測數據比例，其中有 3292 筆數據來自行動網路量測(占整體的 68%)，有 826 筆數據來自固網量測(占整體的 17%)，有 747 筆數據來自有線電視網路量測(占整體的 15%)。

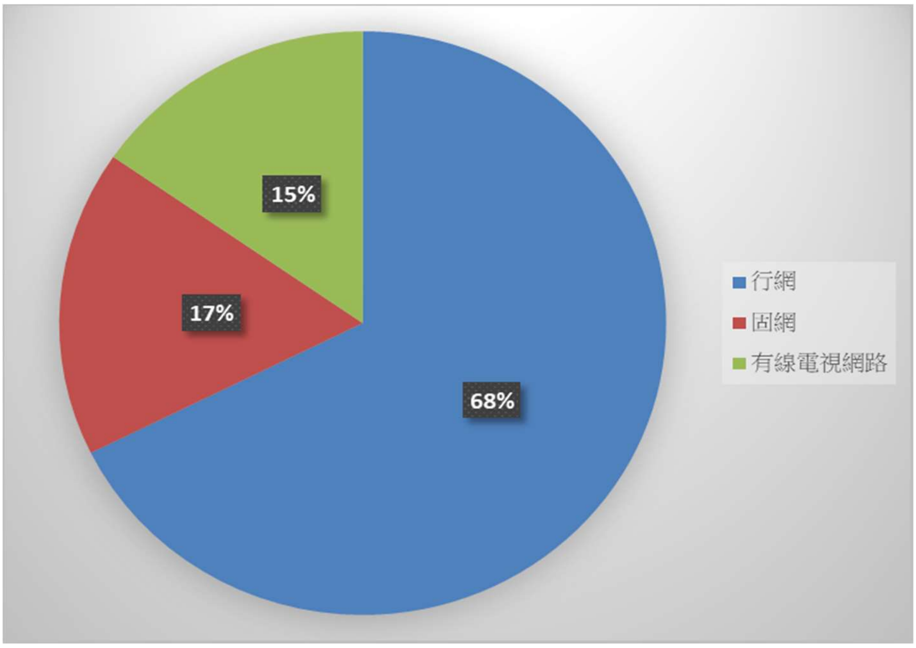


圖 48 中部地區量測數據比例(資料來源:本研究整理)

圖 49 為南部地區量測數據比例，其中有 3397 筆數據來自行動網路量測(占整體的 62%)，有 785 筆數據來自固網量測(占整體的 14%)，有 1296 筆數據來自有線電視網路量測(占整體的 24%)。

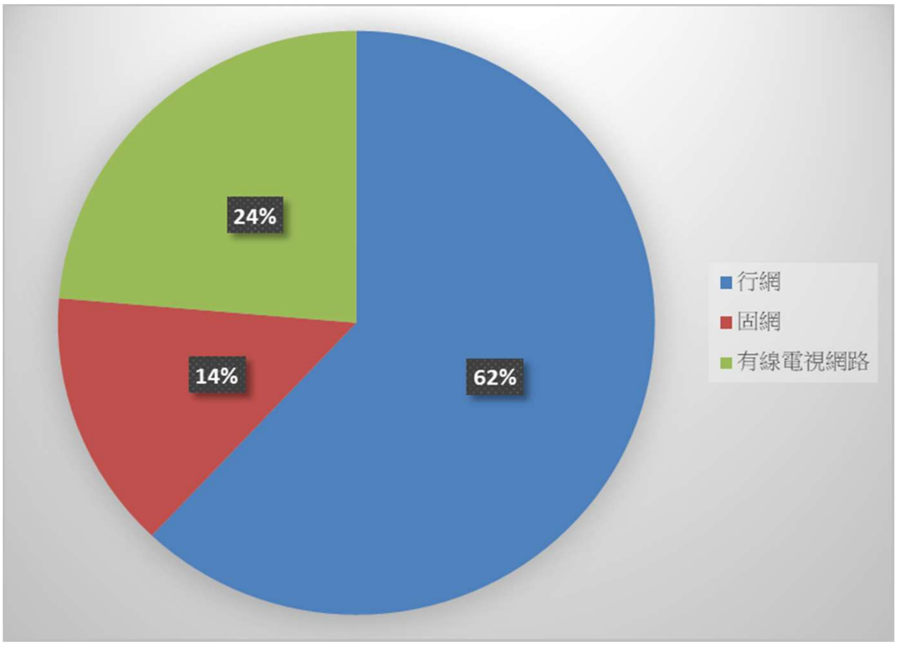


圖 49 量測用戶在南部之各種網路型態百分比(資料來源:本研究整理)

從上述結果分析，本研究目前量測之數據量皆有達到原目標之數據量，並且在量測的數據上也可以發現，目前行動網路的確佔有大部分的使用比例，其中又以北部數據量最多，整體的分布狀況合理，並具有足夠之數據量進行下個階段的信效度分析作業。

二、 信度驗證

本案採用電信技術中心既有研發之「視訊服務品質綜合指標(vMOS)」為依據，並且在後述段落針對該指標進行信度與效度驗證實驗。本研究針對從量測方法指標公式到本研究開發之量測工具可信度驗證方面，為了驗證量測工具的正確性與可信度，我們將影片播放時所產生的事件與相關數據記錄下來，並透過數值計算將該數據帶入指標公式計算出 vMOS 值，再與本量測工具執行測試後所得 vMOS 值進行對照，其數值運算結果與量測工具計算結果兩相符合，證明量測工具的計算方法與我們提出的量測方法指標公式正確性與可信度皆可信賴。

依據 vMOS 指標是基於前述第三章第五節一過往主觀 QoE 研究的結果所產生的，同時該指標也是可以被客觀的量測所取得的重要指標，因此具有高度客觀性。據此本研究依前一節所提出的信度量測驗證方法，我們設計了一項實驗：在連續的 10 天內固定時段、量測工具於採樣地點 (分佈在北/中/南區共 28 處地點)、相同影片源與相同聯網介面進行量測 vMOS。以營造出相似的網路環境條件，將影片解析度作為控制(因)變數，每一次量測皆包含高/中/低三種影片解析度 (分別對應 1080p/480p/240p) 的結果。根據蒐集到的量測結果，我們進行了下列兩項分析：

(一) 再測信度 (Test-retest Reliability) 一週時間差分析

在量測資料中挑選相隔一週的兩天數據組進行再測信度分析，計算前後兩組對應量測結果的相關係數作為信度。以下說明分析方法的細節：將蒐集到的 vMOS 量測結果整理成為 Day 1 - Test 與 Day 2 - Retest 兩組數據，如下表所示僅為其中的部份資料，將 Day 1 - Test 的所有數據串成一列並將之視為一隨機變數 X 的取樣值，將 Day 2 - Retest 的所有數據依對應的順序串成一列並將之視為一隨機變數 Y 的取樣值，再依據第前一節中所介紹的皮爾遜相關係數定義計算 X 與 Y 兩變數間的相關係數 ρ_(X,Y)，該係數即為再測信度 r 值，計算時乃以有限採樣的樣本平均 (sample mean) 取代定義中的總體平均 (ensemble average)。分析結果得到 r = 0.99 的高信度。這意謂著在相同 (似) 的網路環境下，量測工具對於視訊源品質本身的變化總是能具有一致性的反應。

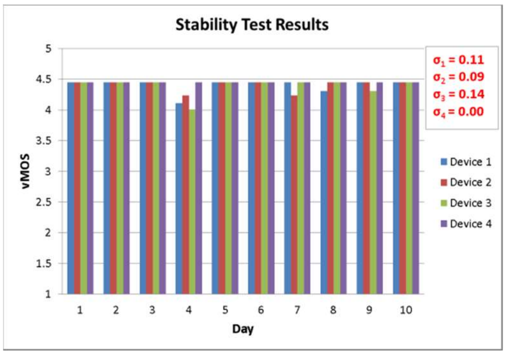
表 15 一週時間差 vMOS 測試結果 (節錄部份數據)

Test Day / Resolution / Device IMEI	Day 1 - Test			Day 7 - Retest		
	1080p	480p	240p	1080p	480p	240p
352711097043216	4.45	3.64	2.70	4.45	3.64	2.70
352711097043414	4.24	3.53	2.60	4.45	3.64	2.70
353759094755968	4.45	3.64	2.50	4.31	3.64	2.70
356772061028589	4.31	3.64	2.70	4.45	3.64	2.63
356810091194896	4.31	3.64	2.63	4.45	3.64	2.70
358491094164711	4.45	3.64	2.70	4.31	3.64	2.70
358491097986888	4.45	3.64	2.70	4.31	3.64	2.70
358491098507691	4.45	3.64	2.70	4.45	3.64	2.70
358491098542813	4.45	3.64	2.70	4.24	3.53	2.63
357008087019081	4.45	3.64	2.70	4.45	3.64	2.70
⋮	⋮	⋮	⋮	⋮	⋮	⋮

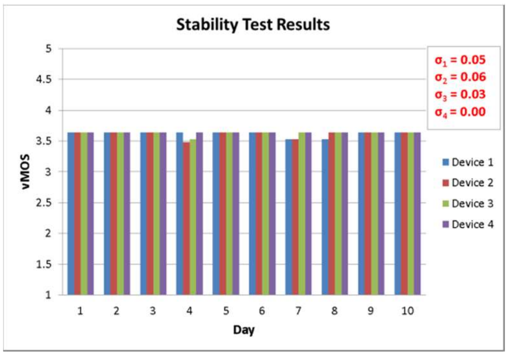
(資料來源:本研究整理)

(二) 連續 10 日同一時段定點測試結果穩定性分析

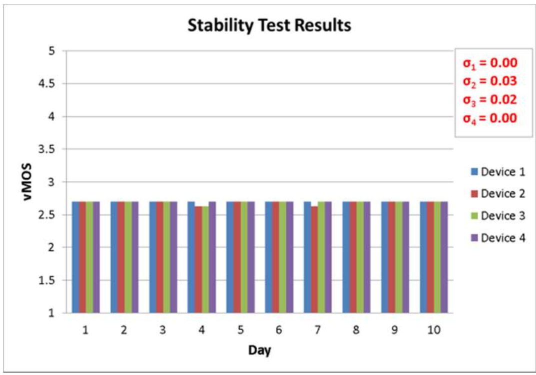
在量測資料中挑選 4 個採樣地點 (涵蓋北/中/南區) 的連續 10 日測試結果，再以柱狀圖呈現個別解析度下個別採樣點連續 10 日量測結果的穩定性，如下圖所示。下圖中 (a)、(b) 與 (c) 分別對應 1080p、480p 與 240p 解析度的結果。從圖中可以觀察出在各種解析度下，個別採樣地點的量測結果都具有明顯的穩定性 (vMOS 之標準差皆低於 0.15)。



(a) 1080p 解析度之結果



(b) 480p 解析度之結果



(c) 240p 解析度之結果

圖 50 量測工具連續 10 日同一時段定點測試。(a) 1080p 解析度之結果、(b) 480p 解析度之結果與 (c) 240p 解析度之結果(資料來源:本研究整理)

三、 有效性驗證

為了驗證量測工具的效度，我們設計了群眾測試 (crowdtesting) 實驗：由近 30 位人員 (年齡分佈為 20 歲至 40 歲，平時有在收看線上影音服務的習慣) 觀看四部影片，其中一部影片最高解析度支援到 720p，其餘三部影片則支援到 1080p，影片長度可選擇為 30 sec 至 90 sec。為了貼近使用者實際觀測的效果，在此測試中我們啟用了自適性串流機制，觀看影片的過程中播放器會依網路頻寬動態地調整要求的影片解析度。觀看每一部影片後即由人員給予 MOS 評分 (1.0～5.0 分，含一位小數) (即 User Score)，作為 Ground Truth 與同時間量測工具量測的 vMOS 一同被記錄。在自適性串流機制下，所開發的量測工具會考量解析度的變化與持續時間在預測整體 QoE 時給予適當的權重。根據蒐集到約 1100 筆回饋與量測結果，我們進行了下列兩項分析：

(一) 效標效度 (Criterion Validity) 分析

採用效標效度量測方法計算 vMOS 與用戶評分 (MOS) 的相關係數作為效度 (v)。分析時，將影片時間長度分為短與長 (分別對應 30 sec 與 90 sec) 兩種級距，個別影片時間長度級距的效度如下表所示。該表顯示了在兩種影片時間長度下，量測工具皆具有高效度。這意謂著量測工具量測到的 QoE 值與真人的主觀 QoE 具有高度正相關性，因此未來可採納量測工具量測取代真人評分，達到自動化與節省人力成本的目的。

表 16 量測工具之效標效度測試結果

影片長度	30 sec	90 sec
效標效度 (v)	0.92	0.93

(資料來源:本研究整理)

(二) 線性回歸分析

利用散點圖搭配線性趨勢線，以視覺化方式呈現 vMOS 與用戶評分 (MOS) 的相關性，如下圖所示，該圖中的紅色實線為理想的完全正相關參考線 (斜率為 1 的對角線)。從圖中可以觀察到散點非常集於對角線上，代表量測工具的量測結果與使用者評分具有高度正相關性。且經由線性回歸分析可以得到兩者間的線性函式如下：

$$vMOS = 0.8475 \times MOS + 0.4066$$

其中，$vMOS$ 為量測工具量測的 QoE，MOS 則為用戶回饋的主觀 QoE，下圖 51 中的藍色虛線即為對應該函式的線性趨勢線。

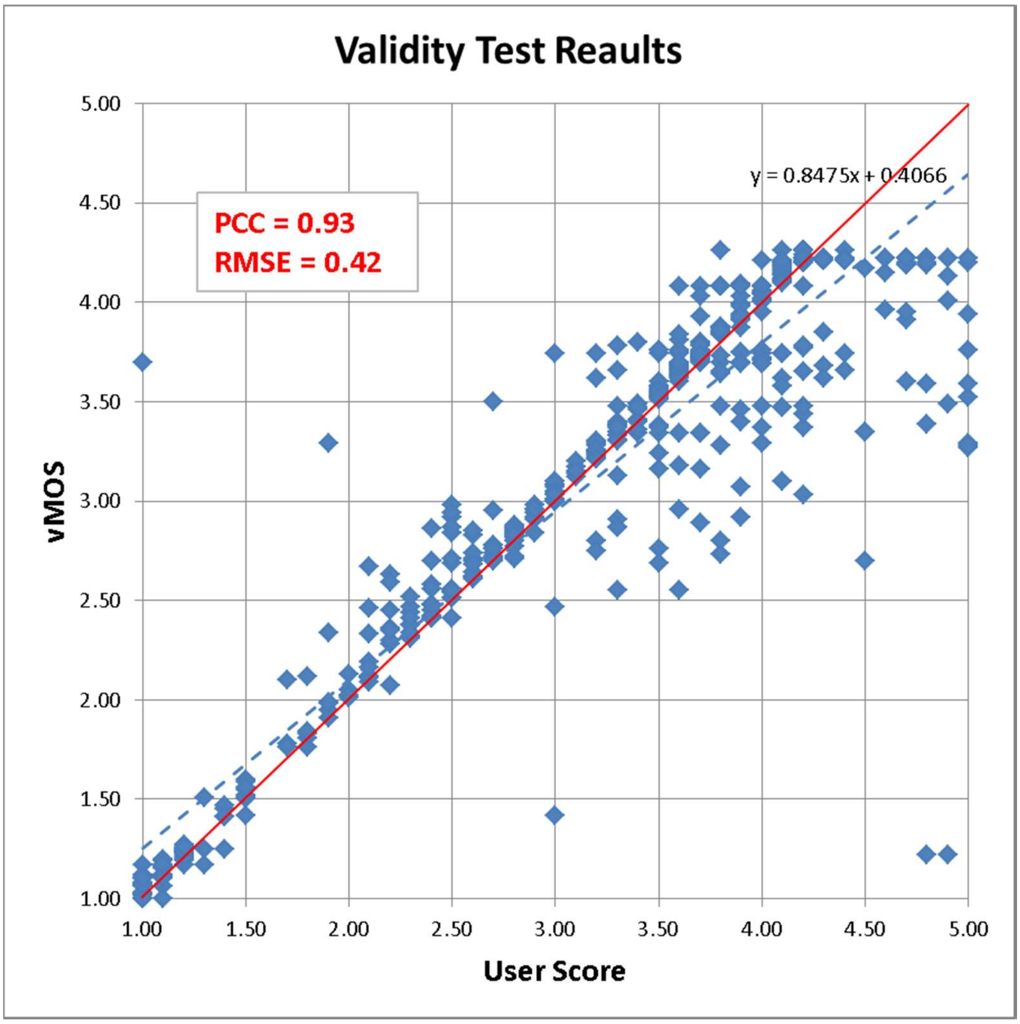


圖 51 量測結果與用戶評分之視覺化相關性(資料來源:本研究整理)

圖 51 中 vMOS 量測結果確實有天花板 (實際為 4.45 分) 現象，其原因是所採用的「視訊服務品質綜合指標 (Video Mean Opinion Score; vMOS)」

$$vMOS = f(\text{視訊解析度}, \text{初始緩衝時間}, \text{卡頓率})$$

針對各種視訊解析度 (4K, 2K, 1080p, 720p, 480p, 360p, …) 設定了個別的空間品質量化分數，以反映視訊解析度對服務體驗的影響。其中 1080p 解析度所對應的空間品質量化分數搭配完美的時間品質 (當網路傳輸條件極佳時) 得到的最佳 vMOS 即為 4.45 分。若解析度達 2K 或 4K 品質時，才會對應更高的最佳 vMOS 值。由於本案場測中所有測試影片的最高解析度都是 1080p(考量在行動裝置上測試，受限於螢幕的尺寸與支援的解析

度，所以選擇 1080p 作為目標最高解析度)，因此導致了模型預測的 vMOS ≤4.45 的現象。

在第四章第四節第三點第(二)款中所做的線性迴歸分析目的並不是為了建構客觀 QoE 模型，而僅是為了呈現 vMOS 預測結果與用戶評分的相關性，方便與圖 51 中理想的完全正相關參考線 (單一斜率為 1 的對角線) 作比較，因此選擇了單一斜率的線性迴歸分析，並在圖 51 中同時繪出該參考線 (紅實線) 與迴歸分析得到的線性趨勢線 (藍虛線)，以利觀察兩者間的差距。

另一方面，也透過誤差統計方式呈現量測結果的 RMSE 與各誤差等級的百分比，如下表所示。由表中可以觀察到大多數情況下，量測結果與真人評分的誤差 97.75%都在 1 分以內。

表 17 量測結果之誤差統計

	統計結果
實驗數量	1110
RMSE	0.34
誤差 1 級	1085 (97.75%)
誤差 2 級	22 (1.98%)
誤差 3 級	1 (0.09%)
誤差 4 級	2 (0.18%)

(資料來源:本研究整理)

通過可靠性驗證，有了可靠的評估指標與量測方法，即可提供串流影音服務營運與管理（或監理）公開服務品質資訊與訂定品質規範的參考，協助相關服務品質之提升，並維護民眾之消費權益。

第五節 圖形化結果與關聯性分析

一、 圖形化網站介紹

資料視覺化是指運用視覺的方式呈現數據，有效的圖表可以將繁雜的數據簡化成為易於吸收的內容，透過圖像化的方式，我們更容易辨別數據的規律、趨勢及關聯。本中心開發的資料視覺化平台，能讓使用者清楚的了解數據，有便於使用者分析資料數據。

以本中心所建置的影音資料視覺化平台為例，如下圖 52 所示，在圓餅圖方面以鮮豔對比的顏色，用使用者清楚的記憶資料數據的各項比例，而資料顯示的重要數據字體也會偏大，讓使用者易讀且易見。

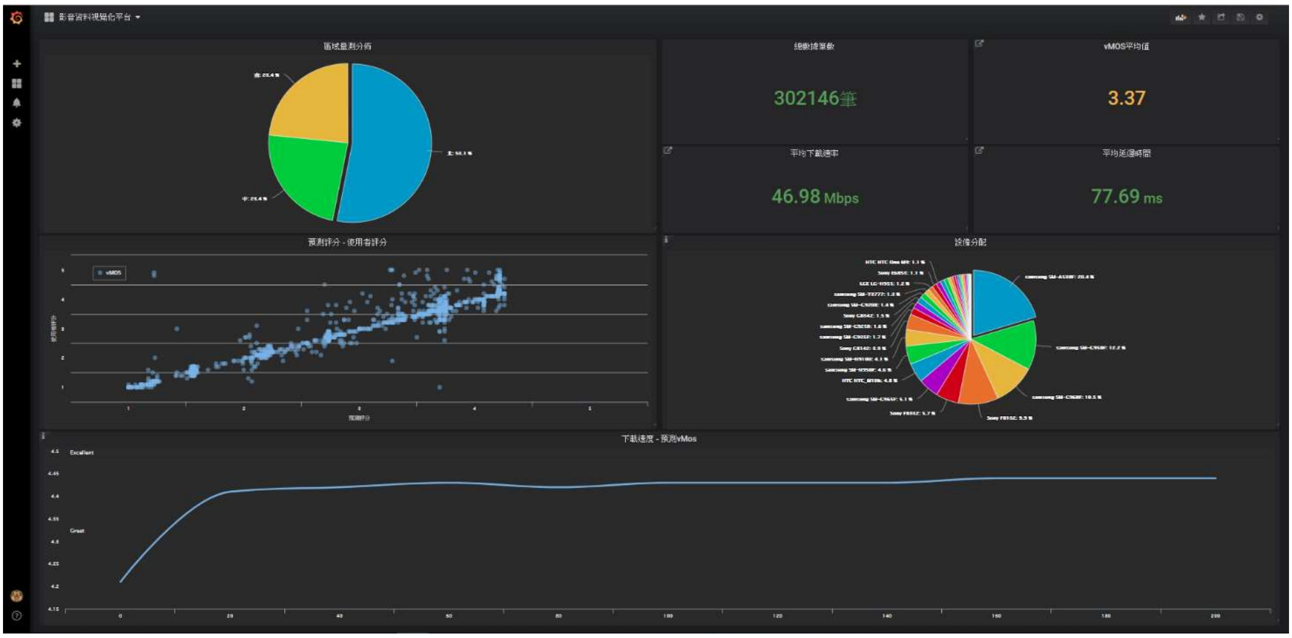


圖 52 影音資料視覺化平台(資料來源:本研究整理)

本研究區域量測的分布，如下圖 53 所示，此圓餅圖可顯示在台灣北、中、南各地方的量測比率，滑鼠移動觸及到之處，不僅會顯示該地區的占比率，也會顯示測試筆數與裝置數量，建置此圖的目的是要讓使用者清楚的知道裝置分布的比例。

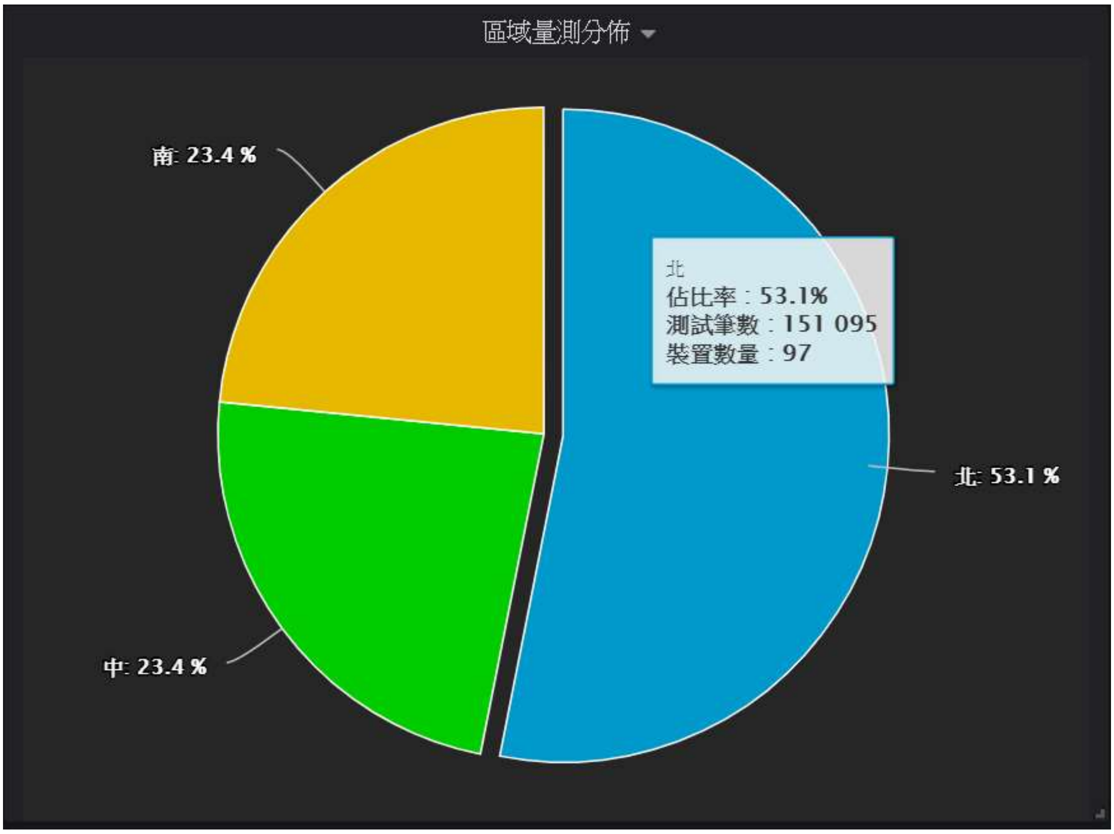


圖 53 區域量測分布(資料來源:本研究整理)

如下圖 54 所示，本平台顯示了「總數據筆數」、「vMOS 平均值」、「平均下載速率」、「平均延遲時間」，以上數據都是本研究所需的重要資訊。

圖 54 本研究之重點數據顯示(資料來源:本研究整理)

點擊 vMOS 平均值後，會顯示三個直方圖，如下圖 55 所示，分別為「每日平均 vMOS 值」、「各網路平均 vMOS 值」、「區域平均 vMOS 值」。首先是每日平均 vMOS 值所表現的是一周星期一到星期日的平均 vMOS 值，如圖可見，現在平均 vMOS 值趨於穩定，同時也代表可以看到網路狀況於不同日期的趨勢表現，可以發現在星期六日因為假日上網量較高狀況下的數值偏低；再者，各網路平均 vMOS 值所表現有分三種型態分別為:cable(有線電視網路之 ISP)、固網、行網(行動網路)；最後，區域平均 vMOS 值所表現的是北中南三個區域。

圖 55 平均下載速率(資料來源:本研究整理)

點擊 vMOS 下載速率後，所各顯示的三個直方圖，如下圖 56、圖 57 所示，都是以「每日平均」、「各網路平均」以及「各區域平均」，來做分析。

圖 56 平均下載速率(資料來源:本研究整理)

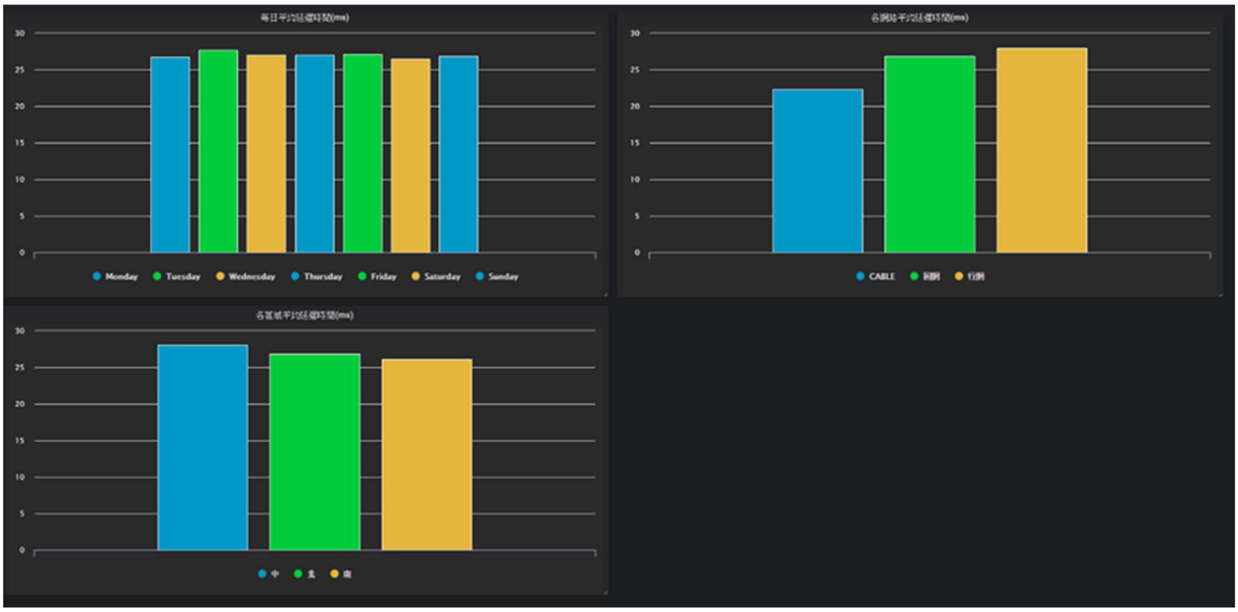


圖 57 平均延遲時間(資料來源:本研究整理)

本平台使用用 XY 散步圖來顯示 vMOS「使用者評分」以及「預測評分」，並用此兩種評分來做分析，如下圖 58 所示，此兩種評分為高度正相關。

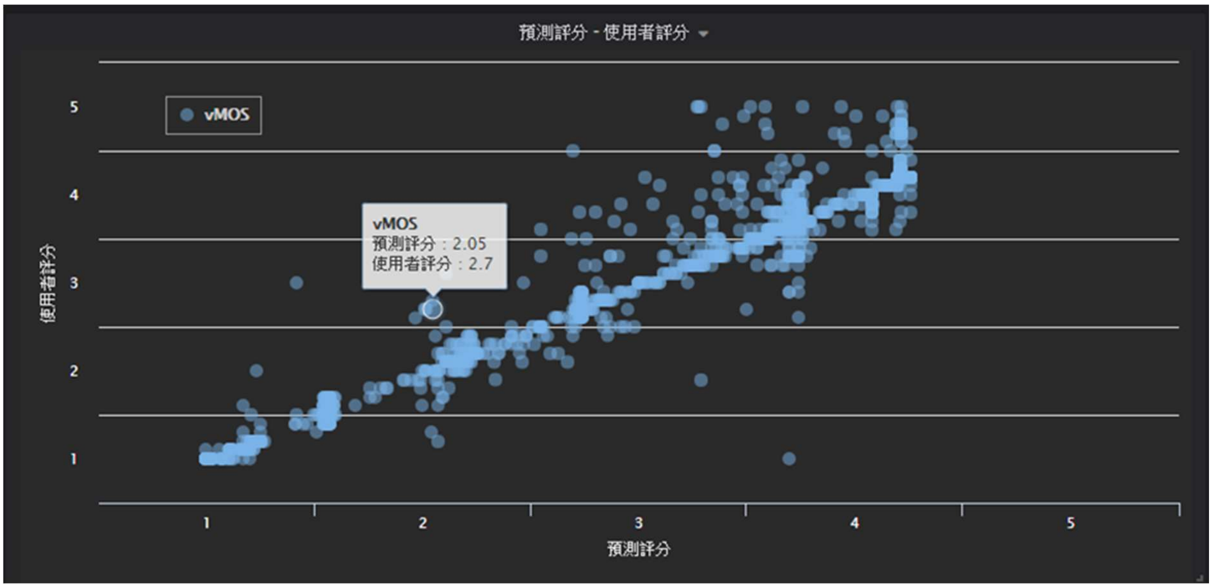


圖 58 使用者評分以及預測評分之 XY 散步圖(資料來源:本研究整理)

在設備分配方面，亦是用圓餅圖表示，本次量測的裝置類型眾多，如下圖 59 所示，其中以「samsung SM-A530F」最多占了 20.4%，其次為「samsung SM-G950F」占了 12.2%，而「samsung SM-G960F」占了 10.5% 排名第三。

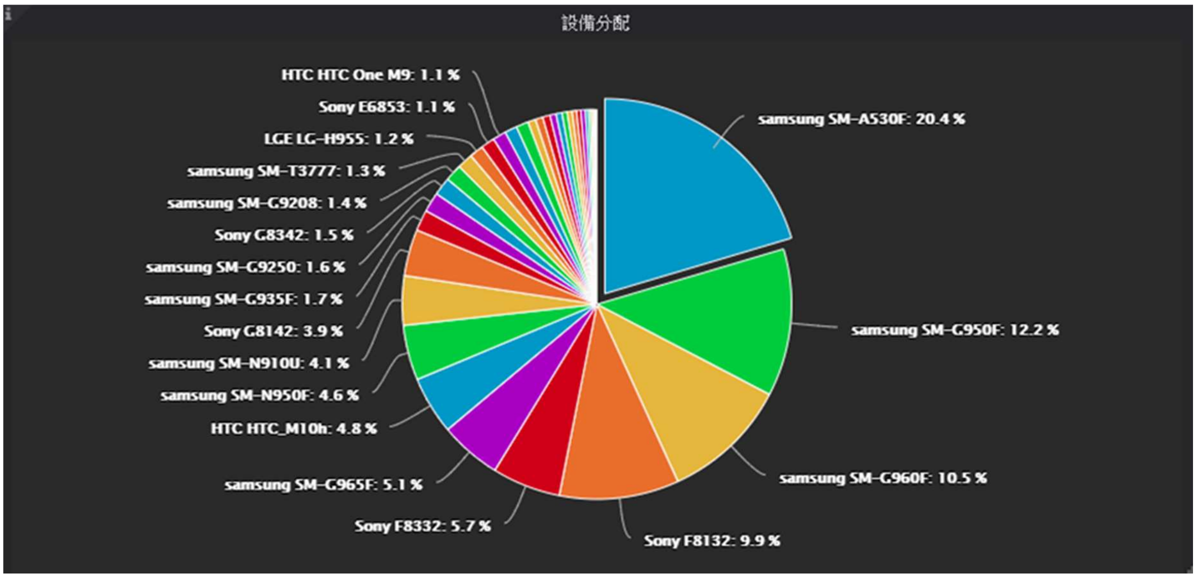


圖 59 設備分配比率圖(資料來源:本研究整理)

本平台亦會用下載速度與預測 vMOS 進行比對分析，如下圖 60 所示，當下載速度超過 20 後，其預測 vMOS 值穩定在 4.4 上下。

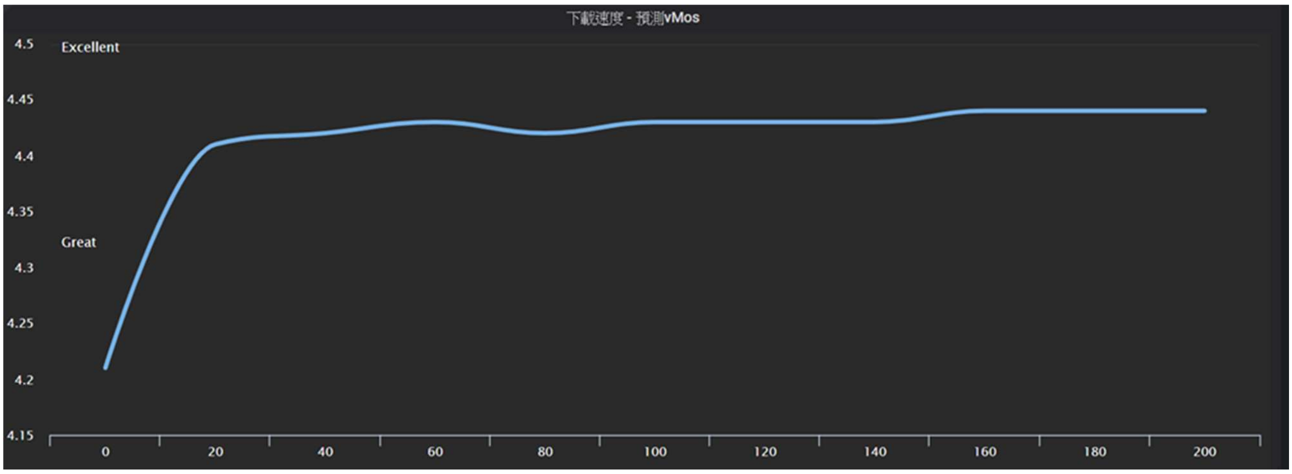


圖 60 下載速率與 vMOS 值關係圖(資料來源:本研究整理)

二、 關聯性分析

(一) vMOS 與影片解析度之關係

針對特定影片 “big buck bunnie” 統計個別解析度下所有量測 vMOS 之最小值、最大值與平均值，結果如圖 4-2 所示，其中粗體字標籤表示平均值。由圖中可以觀察出影片解析度直接反應了平均 vMOS 值的變化，解析度越高則平均 vMOS 也越高。然而，越高解析度的影片串流需要更高品質的網路環境條件，當網路品質不佳時反而會因頻繁地發生卡頓造成嚴重地體驗品質下降，這說明了圖 61 中較高解析度 (720p 與 1080p) 具有較廣的 vMOS 分佈 (1.00～4.45) 與較低的 vMOS 最小值 (1.00) 發生的狀況。

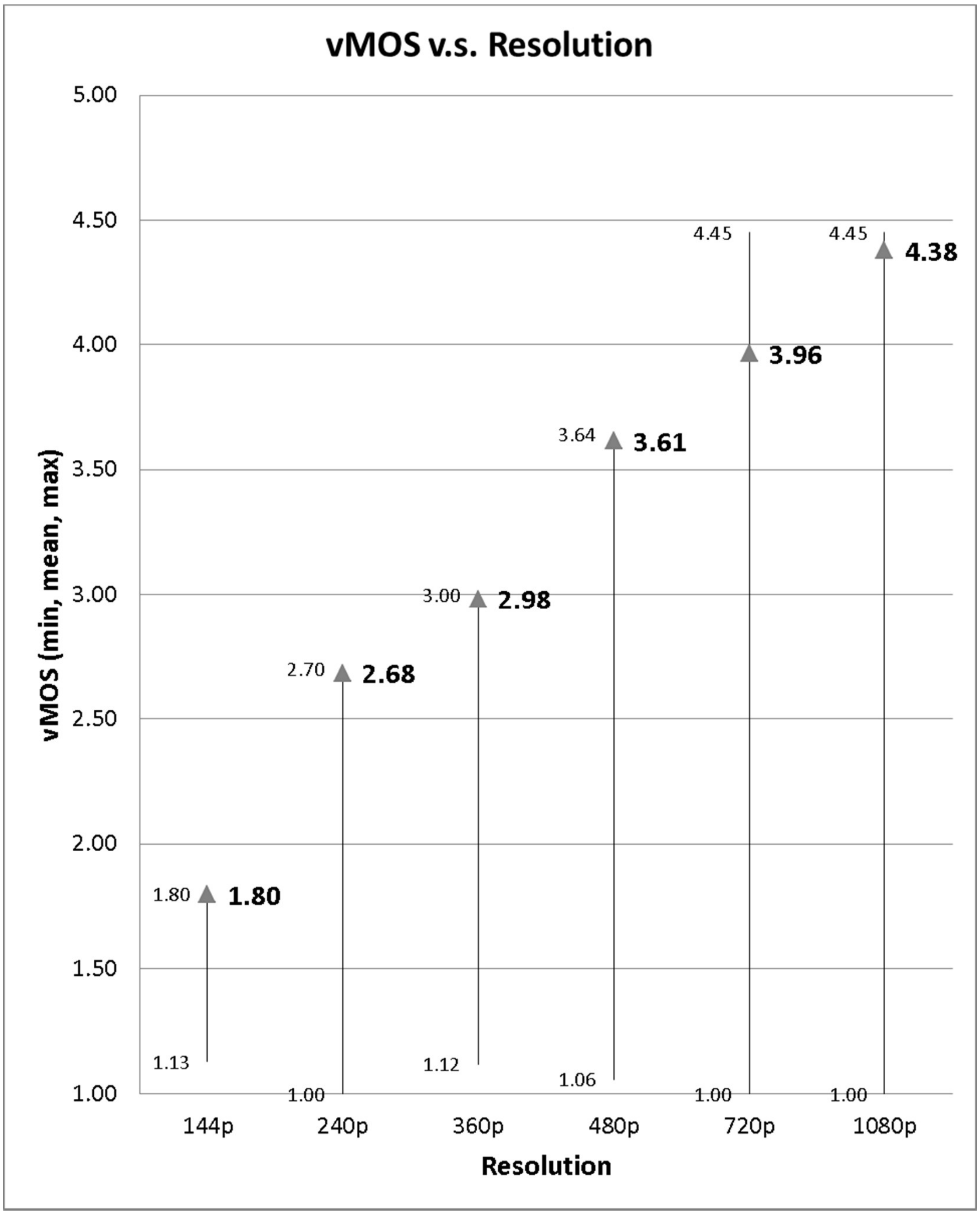


圖 61 vMOS 與影片解析度之關係(資料來源:本研究整理)

(二) vMOS 與影片源之關係

針對不同影片 (四部業者影片與一部共同影片 “big buck bunnie”) 統計個別影片最高解析度下所有量測 vMOS 之最小值、最大值與平均值，結果如圖 62 所示，其中粗體字標籤表示平均值。由於個別影片的最高解析度皆為 1080p，且量測工具不受影片內容影響，因此由圖中可以觀察出個別影片都呈現類似的平均體驗品質 (平均 vMOS 4.35 ~ 4.42)。

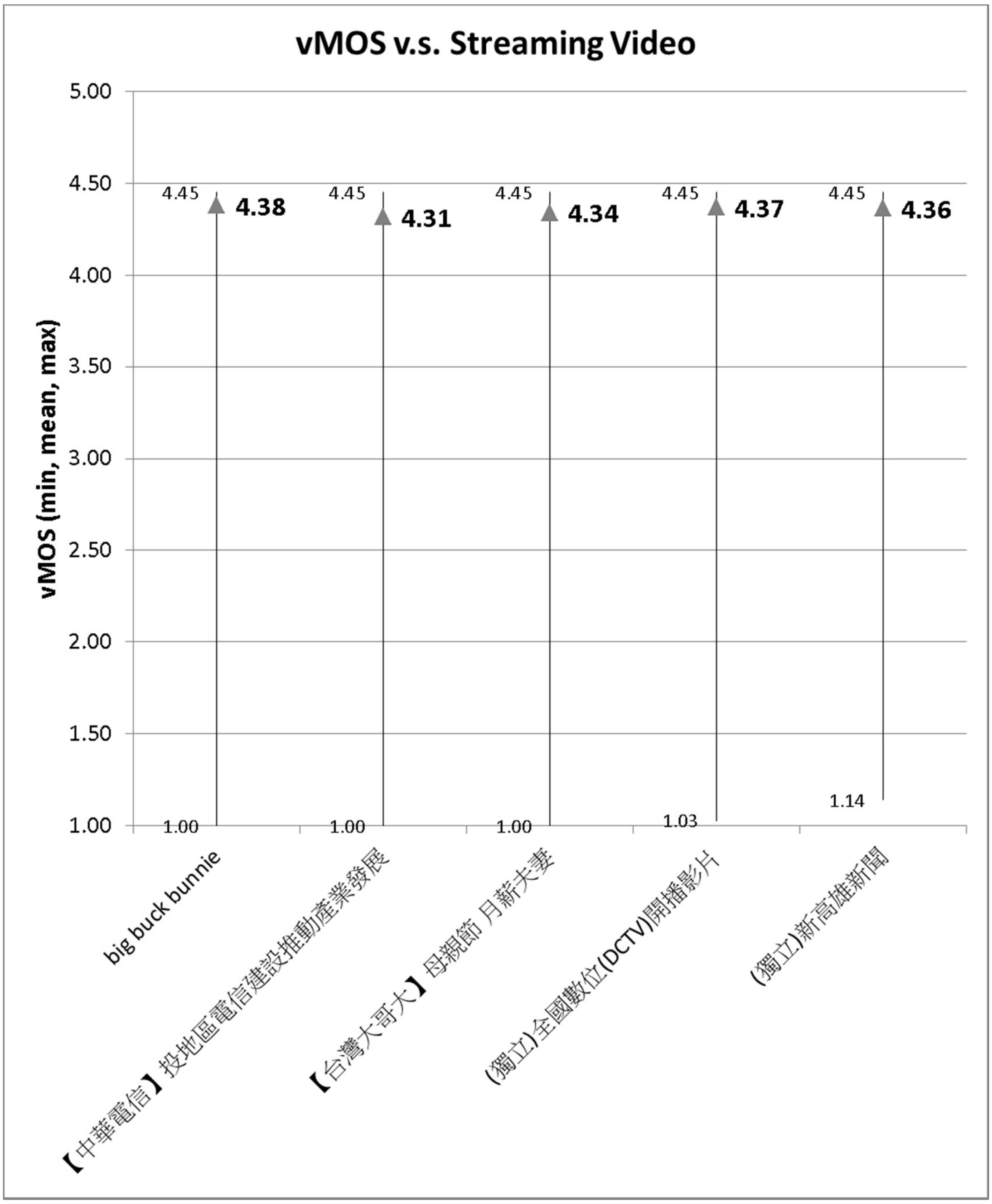


圖 62 vMOS 與影片源之關係(資料來源:本研究整理)

(三) 固網環境下 vMOS 與上網時間之關聯性

針對固網環境與特定影片 “big buck bunnie” (最高解析度) 統計個別時間區間 (每兩小時為一區間) 所有量測 vMOS 之平均值，北/中/南三地區統計結果彙整如圖 63 所示。從圖中可以觀察到固網環境下在固定的地點較易反映出網路環境隨時間的變化，而 vMOS 量測工具恰好可以作為監控的工具，例如在中部地區顯示了 00:00～02:00、12:00～14:00 與 22:00～24:00 這些區間網路環境較其它時段差 (平均 vMOS 低於 4.20)，而這些幾個區間也正好是民眾休息時間，這顯示出中部民眾在這幾個休息的時段上對於固網使用量上有明顯增加，而使用量上升所增加的頻寬流量足以影響到使用者觀看影音之品質體驗，但整體而言中部固網環境仍提供了良好的網路傳輸品質，大多數時段平均 vMOS 都可達 4.40 以上。而南部則是在早上七點與下午五點時段有下滑，但整體而言都低於中部與北部。這表示固網業者在中部與南部服務建置上，還有進步的空間。

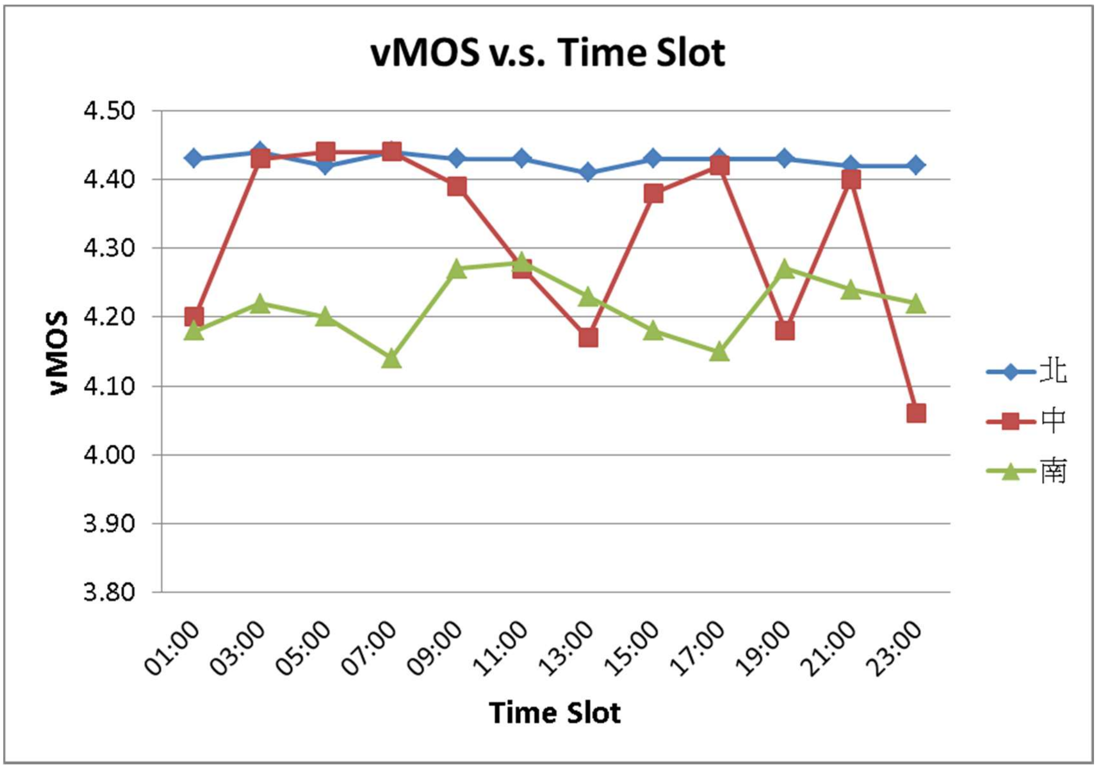


圖 63 固網環境下 vMOS 與上網時間之關聯性(資料來源:本研究整理)

(四) 行網環境下 vMOS 與上網時間之關聯性

針對行網環境與特定影片 “big buck bunnie” (最高解析度) 統計個別時間區間 (每兩小時為一區間) 所有量測 vMOS 之平均值，北/中/南三地區統計結果彙整如圖 64 所示。從圖中可以觀察到行網環境下在非固定的地點不容易反映出網路環境隨時間的變化 (影響被行動特性平均化了)，整體而言由這樣的結果可以看到台灣的 4G 行網環境提供了良好的網路傳輸品質，幾乎所有時段平均 vMOS 都可達 4.30 以上。

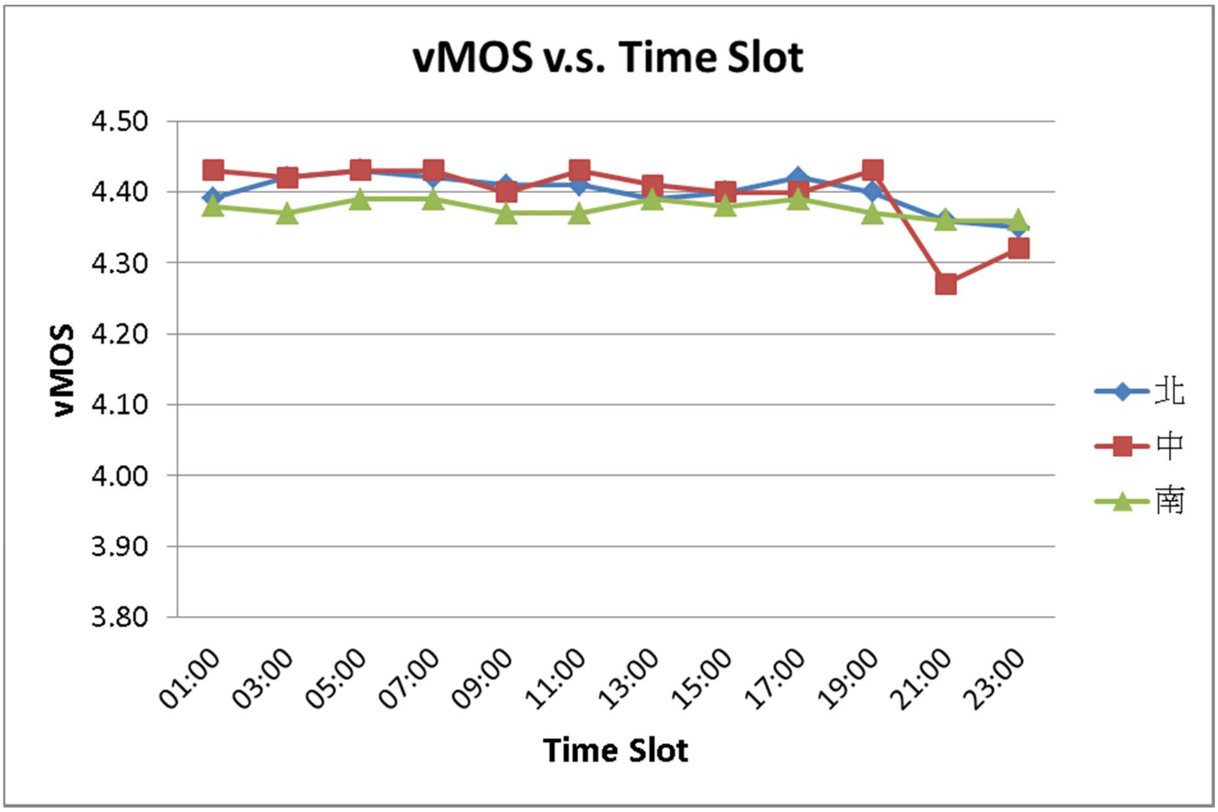


圖 64 行網環境下 vMOS 與上網時間之關聯性(資料來源:本研究整理)

從圖 63 與圖 64 中可以觀察網路型態、地區差異與時段變化對於服務品質的影響，造成這些影響背後的因素實際上可能與網路架構及民眾生活習慣相關。

以圖 63 為例，該圖呈現了不同地區固網環境下服務品質與上網時間之關聯性。整體而言北區的品質 (用戶體驗) 最佳、中區次之、南

區再次之，此類地區差異性通常是因業者網路與路由架構導致。由於固網環境下的定點測試裝置乃分享固網或有線電視網路提供的無線網路進行聯網，因此該固網環境中存在其它分享頻寬的應用 (網頁瀏覽、網路遊戲、網路電視等) 將會對可用頻寬造成影響。在圖 63 中可觀察出中區與南區時段變化對於服務品質較為明顯，在中部地區顯示了 12:00 ~ 14:00 (午餐/午休時段)、18:00 ~ 20:00 (晚餐時段) 與 22:00 ~ 24:00 (就寢前) 這些區間網路環境較其它時段差，而在南部地區則顯示了 06:00 ~ 08:00 (早餐/上(學)班前時段) 與 16:00 ~ 18:00 (放學/下班後時段) 這些區間網路環境較其它時段差，推論這些網路環境較差的時段基本上都是固網用戶較會使用網路的悠閒時間。至於圖 64 的行網環境在非固定的地點不容易反映出網路環境隨時間的變化 (影響被行動特性平均化了)，僅觀察到在 20:00 ~ 24:00 (晚上休閒時間) 區間服務品質有些微變化。

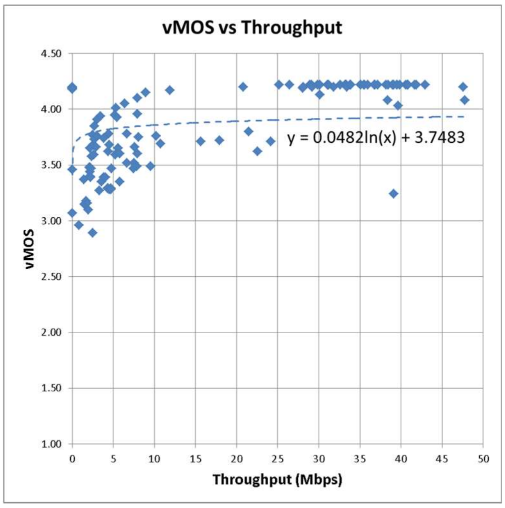
(五) QoE (vMOS) 與 QoS (Throughput) 之關聯性

一般而言，相對於語音服務主要訴求為低傳輸延遲與抖動，串流影音點播服務更需要穩定的傳輸速率才可以讓使用者順暢的播放影片。為了驗證此論述，我們將有啟用自適性串流機制的量測結果用於分析 vMOS 與 Throughput 之關聯性，因為在自適性串流機制下才會根據網路頻寬的變化動態地調整影片解析度需求。分析時，將連網介面分成 4G 與 WiFi，個別以散點圖搭配對數趨勢線呈現該網路介面下 vMOS 與 Throughput 之關聯性，如圖 65 所示。圖 65 (a) 與 (b) 分別對應 4G 與 WiFi 連網介面的結果。從圖中可以看出，vMOS 與 Throughput 呈現近似對數函式的關係：

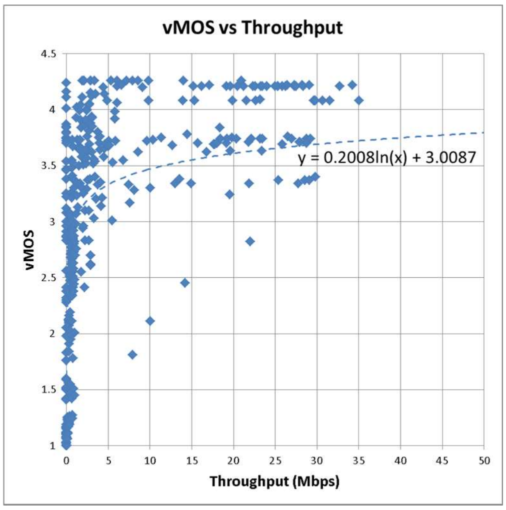
$$\text{4G:} \quad vMOS = 0.0482 \times \ln(MOS) + 3.7483$$

$$\text{WiFi:} \quad vMOS = 0.2008 \times \ln(MOS) + 3.0087$$

由圖中可以觀察出當 Throughput 高達 5 Mbps，通常可以有較高且穩定的 vMOS 評分。至於當 Throughput 介於 0 ~ 5 Mbps 時，可能會因切換至較低解析度或甚至發生卡頓導致較低的 vMOS 評分。



(a) 4G 行網結果



(b) WiFi 固網結果

圖 65 (a) 行網環境與 (b) 固網環境下之平均 vMOS 與上網時間之關聯性(資料來源:本研究整理)

第五章 活動辦理紀錄與成效

第一節 座談會

本研究計畫擬針對通傳事業經營者透過行動寬頻或自建 WiFi 網路提供的影音服務，建立一套可佈建於消費者端的影音服務體驗品質 (QoE) 量測方式並結合國際間主要先進國家傳輸網路服務品質 (QoS) 量測方法分析影音品質不佳的因素，進一步建立具有公信力之影音品質量測機制，作為提升通傳產業之數位匯流影音品質之參考。

此次座談會將針對數位匯流影音平臺服務品質量測之政策法規及量測方法進行說明交流，並期望能進一步得到認同、支持與回饋，促使法規與量測方法能更加完善，以帶動各領域產業發展。座談會相關資訊如下：(詳細座談會紀錄可參閱附件十)

時間	107 年 9 月 4 日（星期二）下午 14:00-16:00
地點	財團法人電信技術中心 高雄本部 1F 國際會議廳
與會單位與人次	與會單位：電信技術中心、凱擘大寬頻、台灣大哥大、國立高雄科技大學、遠傳電信、中華電信等共 6 家，共計 21 人次
內容	一、 數位匯流影音平臺服務品質量測之政策法規： **線上影音平臺服務品質量測的政策意涵** 目前線上影音無論是國內外皆無法律強制的量測標準，主要原因是線上影音服務類型眾多，各家廠商所實施或注重的服務也不盡相同，若需量測，須由廠商發布其 API 或串流源供測試所需，但現階段市場生態如要執行量測勢必遭遇困難。 ● 雖線上影音服務量測準則仍處在模糊地帶，但可針對市售產品列舉較常使用 KPI 作為參考，如： ● 資源上下載速率、資源使用延遲時間、系統穩定度、影音幀速率、影音幀分辨率、影音壓縮率、網路頻寬速率、網路延遲時間、封包遺失率、DNS 解析時間。 ● 為研析針對國內有線電視系統、電信事業固網與行動寬頻等業者了解其影音服務架構與產業現況，並對國際間主要國家之線上影

音服務法規、監理政策，研析國際監理機關對線上影音服務之立場與看法。

美歐網路中立性政策立場

歐盟網路中立性法規執行準則（2016）

- 準則一：應保障之使用者權利 （法規第 3 條之 1）
- 準則二：合理之商業行為 （第 3 條之 2）
- ISP 與使用者間所達成之協議或契約中，唯侵害使用者權利之項目，否則不受本法之限制。例如吃到飽方案、達流量上限之額外加購流量
- 準則三：合理之商業行為-零費率爭議 （第 3 條之 2）
- 準則四：合理之流量管理機制 （第 3 條之 3）
- 符合資訊透明（transparent）、無歧視（non-discriminatory）、比例原則（proportionate）等原則
- 準則五：網路中立性允許提供專業服務（第 3 條之 3）

二、 串流影音平臺服務品質量測方法

服務品質簡介：

串流影音服務：透過開放式網際網路直接對用戶提供各種視音訊內容的服務。視音訊內容以串流方式經由網際網路，再透過用戶端的行網或固網傳送至使用者的電視、電腦、智慧型手機或平板電腦等各種終端設備。用戶端可以邊載邊看（只需等待相對短暫的初始片段下載時間，就可以持續收看完整的影音內容）。

串流影音平臺：串流影音服務的提供者

用戶端需求：支援的瀏覽器或特定的播放器/應用程式（App）

營運模式：廣告、贊助、付費訂閱、授權、週邊商品與大數據運用等

服務品質量化：

定性而言，影音服務的「高品質」意謂著「低延遲、流暢穩定、高畫質與高傳真」，具體的量化數據才有助於資料蒐集、處理、統計與分析。基於網路性能指標的量測，包含傳輸延遲（DELAYORLATENCY）、傳輸延遲變異（IPDV）、

丟包率（PACKETLOSSRATE）、上/下行吞吐量。

客觀 QOE 模型範例：預估QoE得分=（初始緩衝時間得分×$\theta1$）+（PPI得分×$\theta2$）+（卡頓率得分×$\theta3$）

第二節 說明會

本研究計畫擬針對通傳事業經營者透過行動寬頻或自建 WiFi 網路提供的影音服務，建立一套可佈建於消費者端的影音服務體驗品質 (QoE) 量測方式並結合國際間主要先進國家傳輸網路服務品質 (QoS) 量測方法分析影音品質不佳的因素，進一步建立具有公信力之影音品質量測機制，作為提升通傳產業之數位匯流影音品質之參考。

此次說明會將對數位匯流影音平臺服務品質量測內容進行說明，以期望數位匯流影音平臺服務品質之量測可以順利進行。座談會相關資訊如下：(詳細說明會紀錄可參閱附件十一)

場次	台北場
時間	107 年 9 月 7 日（星期五）上午 10:00-12:00
地點	財團法人電信技術中心 台北辦公室 大會議室
與會單位與人次	與會單位：國家通訊傳播委員會、電信技術中心、凱擘大寬頻、中華電信、台灣有線寬頻產業協會、國立高雄科技大學、台灣大哥大、中嘉網路、遠傳電信等共 9 家，共計 27 人次
場次	台中場
時間	107 年 9 月 13 日（星期四）下午 14:00-16:00
地點	國立台中教育大學 求真樓 K401 會議室
與會單位與人次	與會單位：電信技術中心、台灣基礎開發科技、台灣寬頻通訊顧問股份有限公司、群健有線電視、台灣寬頻、台灣大哥大、國立台中教育大學、中華電信、台基科等共 9 家，共計 23 人次。

內容	**串流影音服務簡介** 串流影音服務:透過開放式網際網路直接對用戶提供各種視音訊內容的服務 串流影音平臺：串流影音服務的提供者 用戶端需求：支援的瀏覽器或特定的播放器/應用程式(App) 營運模式：廣告、贊助、付費訂閱、授權、週邊商品與大數據運用等 **服務品質量化：** 定性而言，影音服務的「高品質」意謂著「低延遲、流暢穩定、高畫質與高傳真」，具體的量化數據才有助於資料蒐集、處理、統計與分析。基於網路性能指標的量測，包含傳輸延遲(DELAYORLATENCY)、傳輸延遲變異（IPDV)、丟包率（PACKETLOSSRATE)、上/下行吞吐量。 客觀 QOE 模型範例：預估QoE得分=（初始緩衝時間得分×$\theta 1$）+（PPI得分×$\theta 2$）+（卡頓率得分×$\theta 3$） **量測方法概覽** 基於網路性能指標的量測：傳輸延遲(Delay or Latency)、傳輸延遲變異(IPDV)、往返時間延遲(RTT)、丟包率(Packet Loss Rate)、上/下行吞吐量(DL/UL Throughput) 客觀 QOE 模型範例：預估 QoE 得分=(初始緩衝時間得分 × $\theta 1$) +(PPI得分 ×$\theta 2$)+(卡頓率得分 ×$\theta 3$) **量測工具說明** 提出可行的 QoE 模型：vMOS=f(視訊解析度,初始緩衝時間,卡頓率) 開發量測工具：以 YouTube 服務為量測對象開發 Android App，含網路性能指標（QoS）量測、網路問題診斷、影音服務體驗品質(QoE）量測等功能 **佈建規劃** 進行多個採樣用戶（多家固網與行網業者）的長期測試:共進行三梯測試，每一梯測試 40~50 支行動裝置，每一梯進行 15 天量測，期間每隔 2 小時即自動執行一次測試（一部業者影片與一部共同影片），每單次自動測試約耗時 7~15 分鐘（實際與網路傳輸條件有關）。依此取樣頻率設計，測試預計將可蒐集到 14,400 至 18,000 筆總量測數據，之後再依各分析項目取其中適用的子集合數據進行分析。

第三節 教育訓練

本研究計畫建立一套可佈建於消費者端的影音服務體驗品質 (QoE) 量測方式並結合國際間主要網路服務品質 (QoS) 量測方法分析影音品質不佳的因素，進一步建立具有公信力之影音品質量測機制，作為提升通傳產業之數位匯流影音品質之參考。此次教育訓練將對數位匯流影音平臺服務品質量測方法進行說明，座談會相關資訊如下：(詳細教育訓練紀錄可參閱附件十二)

場次	北部場
時間	107 年 11 月 22 日(星期四) 上午 10 點至 11 點
地點	交通通訊傳播大樓 2003 會議室(台北市中正區仁愛路 1 段 50 號 20 樓)
與會單位與人次	與會單位：國家通訊傳播委員會、電信技術中心，共計 11 人次
場次	中部場
時間	107 年 11 月 27 日(星期二) 上午 10 點至 11 點
地點	國家通訊傳播委員會中區監理處 1 樓研習室
與會單位與人次	與會單位：國家通訊傳播委員會、電信技術中心，共計 24 人次。
場次	南部場
時間	107 年 11 月 30 日(星期五) 上午 10 點至 11 點
地點	國家通訊傳播委員會南區監理處 402 會議室
與會單位與人次	與會單位：國家通訊傳播委員會、電信技術中心，共計 20 人次。

內容	**串流影音服務簡介** 串流影音服務：透過開放式網際網路直接對用戶提供各種視音訊內容的服務 串流影音平臺：串流影音服務的提供者 用戶端需求：支援的瀏覽器或特定的播放器/應用程式(App) 營運模式：廣告、贊助、付費訂閱、授權、週邊商品與大數據運用等 **服務品質量化：** 定性而言，影音服務的「高品質」意謂著「低延遲、流暢穩定、高畫質與高傳真」，具體的量化數據才有助於資料蒐集、處理、統計與分析。基於網路性能指標的量測，包含傳輸延遲（DELAYORLATENCY)、傳輸延遲變異（IPDV)、丟包率（PACKETLOSSRATE)、上/下行吞吐量。 客觀 QOE 模型範例：預估QoE得分=（初始緩衝時間得分×$\theta 1$）+（PPI得分×$\theta 2$）+（卡頓率得分×$\theta 3$） **量測方法概覽** 基於網路性能指標的量測：傳輸延遲(Delay or Latency)、傳輸延遲變異(IPDV)、往返時間延遲(RTT)、丟包率(Packet Loss Rate)、上/下行吞吐量(DL/UL Throughput) 客觀 QOE 模型範例：預估 QoE 得分=(初始緩衝時間得分 × $\theta 1$) + (PPI得分 ×$\theta 2$)+(卡頓率得分 ×$\theta 3$) **量測工具說明** 提出可行的 QoE 模型：vMOS=f(視訊解析度,初始緩衝時間,卡頓率) 開發量測工具：以 YouTube 服務為量測對象開發 Android App，含網路性能指標（QoS）量測、網路問題診斷、影音服務體驗品質(QoE）量測等功能 **佈建規劃** 進行多個採樣用戶（多家固網與行網業者）的長期測試：共進行三梯測試，每一梯測試 40~50 支行動裝置，每一梯進行 15 天量測，期間每隔 2 小時即自動執行一次測試（一部業者影片與一部共同影片），每單次自動測試約耗時 7~15 分鐘（實際與網路傳輸條件有關）。依此取樣頻率設計，測試預計將可蒐集到 14,400 至 18,000 筆總量測數據，之後再依各分析項目取其中適用的子集合數據進行分析。

第四節 會議回饋與成效

本案辦理座談會、說明會、教育訓練活動，不重複與會廠商共 15 家，其中包括國家通訊傳播委員會、電信技術中心、凱擘大寬頻、台灣大哥大、國立高雄科技大學、遠傳電信、中華電信國家通訊傳播委員會、台灣有線寬頻產業協會、中嘉網路、台灣基礎開發科技、台灣寬頻通訊顧問股份有限公司、群健有線電視、台灣寬頻、國立台中教育大學、台基科等產、官、學界人員共 126 人次。透過這場座談會讓與會人員了解影音服務品質的重要性與實際量測的實施辦法。針對各會議的做列點式彙整。

1. 針對網站管理權則區分，目前 NCC 的權責僅能規管國內的影音服務平臺業者，對於國外的影音網站若產生糾紛，則以消費者爭議的民事方式處理。
2. 以網路中立性來說，消費者要求低價吃到飽，又要求網路品質須達到某個程度的要求，對行動網路業者來說所需要進行的基礎建設負擔相當龐大與吃重。低價吃到飽確實會影響台灣電信業者的發展，網路中立與網路費率有連帶關聯，政府管越多，資費就不會往上爬，對網路建設期會有很大的負面影響。另一方面，台灣的電信法規範要求以比其他國家嚴謹與嚴格，因此無須再討論網路中立的問題。
3. 對於影音服務品質的量測標準，QoS 採用國際間通用的量測方法，QoE 參考國際間（只有）學術與業界使用的方法進行測試。
4. 對影音服務品質的量測方式，必須使用使用客觀的量測方式進行。而量測的參考點，則會在不同地區(縣市)進行單一用戶測試 15 天。
5. 業者表示國外影音服務平台的崛起，對國內業者眾多業者帶來強烈的衝擊。透過本案所提出的量測方法，能有助於消費者釐清影音服務平台服務品質不佳的問題點，但期望勿成為替非規管之影音平台背書工具而造成更大衝擊。
6. 有業者擔心未來該測試方法是否會列入主管機關固定評測項目。就現階段而言，並不會納入定期評量的評測項目之中。

第六章 結論

在本研究報告中，在第一章介紹目前的數位匯流影音服務的各種型態，與國內通傳事業之數位影音匯流服務架構解析，了解國內數位匯流影音平臺服務產業現況，並探討匯流影音產業面臨轉型挑戰與侵權困境。台灣數位匯流影音平臺服務產業面臨強勁境外對手，如何提升本身平臺差異化，並積極製作優質內容，是所有業者共同目標；而面對另一個更強大的對手—侵權網站（或侵權機上盒），業者在面臨缺乏智財保護的環境之下，在產業整體發展上顯然較他國有相當的差距，因此對於侵權問題的處置上需要政府在修法與執法行動上的積極協助，始能徹底解決此一問題。

在第二章透過了解數位匯流影音服務品質監理的政策意涵，再研析美國、加拿大、英國、法國、新加坡五個主要國家的數位匯流影音服務品質監理政策，探討網路中立性對數位匯流影音服務品質監理之影響，並說明服務品質與網路中立性的監理目前之挑戰。從管制觀點來看，QoS 管理面臨許多挑戰。對 QoS 進行區別管理（differentiated management）不僅對網路業者、內容/應用業者，甚至消費者或其他終端使用者皆具有潛在利益。在單一市場上讓所有電信服務（從企業界角度看是好的）皆具備網路中立服務，事實上就是數位匯流線上影音，使其成為具備 QoS 保證的資料服務（例如全球資訊網、電子郵件、VoIP、Facebook、BitTorrent 等）。目前全球並無明確獲最佳的方案因應網路中立性，各國根據其內國環境所採取的態度各有不同，約略可分為謹慎觀察、低管制、特定管制三類：第一類是採取謹慎觀察態度，這類國家並未採取任何特定措施以因應網路中立性，認為既有規範即為已足；第二類是採取低度管制態度（light-handed approach），例如資訊揭露與透明性原則、降低轉換障礙、最低服務品質（minimum QoS）而與既有規範之間有所微調，但不至於禁止特定行為；第三類則是採取特定管制措施禁止 ISP 執行特定行為，通常是根據合理網路管理實務作為。

而如何提升通傳產業之數位匯流影音品質，本研究有以下三點建議提供未來相關政策法規之修法方向與建議：

一、具體落實「資訊透明化」原則。

二、承諾服務品質形同消保法中「廣告」之義務。

三、避免以政府監理方式要求線上影音平臺服務遵守服務品質規範。

今日的線上影音平臺與傳統有線電視、IPTV 在訊息傳遞範圍上有相當大的差異，對於在網際網路上提供線上影音平臺所面對的市場已是全球性的，採取由上而下的管制方式並不實際也欠缺規管基礎，建議朝向讓業者自行管理與使用者締約，政府的角色即在單純維護消費者權益，只有這樣才能擺脫過去的規管思維，讓管制機關更聰明地進行規管、讓產業能更聰明地將能力發揮在該發揮的地方。

在第三章介紹了串流影音服務之量化品質量測概觀，包含對問題的描述、對量測方法的演進彙整與介紹此領域常用的可靠性驗證方式。再分別對基於 QoS 量測與基於 QoE 量測的方法做進一步的說明與探討，包含了常見的 QoS 指標量測方法、QoS-to-QoE 對應關係、主觀的與客觀的 QoE 量測指標與方法等。其中，與本研究案研究目的最相關的客觀 QoE 模型視訊品質量測方法部份，介紹了涵蓋研究論文、產業白皮書與標準文獻中的十種客觀 QoE 模型。

基於對現有視訊品質量測方法的掌握，我們對不同類方法進行了客觀地比較，並由本研究案研究目的角度提出了量測方法之實施建議。在行動裝置應用前提下，基於模型複雜度與實作可行性的綜合考量，建議採用類似 MobileU-vMOS 的客觀 QoE 模型作為量測指標。本研究案後續將朝此方向設計與開發量測工具，規劃佈建量測工具並蒐集測試結果，以期在產業應用上能作為串流影音服務的監控與改善方面的參考。藉由即時且自動化量測並公開影音服務品質量測結果，可以讓使用者、服務供應商或監理單位掌握服務品質。監理單位也可依據這樣的量測方法訂定相關的服務品質規範，要求服務供應商進行改善以保障使用者權益。

在第四章說明本研究佈建量測的方式與結果。透過在用戶端執行客觀 QoE 量測以利直接蒐集與同時也可以一併記錄用戶端量測到的網路 QoS 參數。藉由投入人力超過 200 個使用戶、至少量測 15 天、蒐集超過 20,000 筆測試結果進行分析，在信度驗證的實驗中再測信度達 0.99，測試時影片長度為 30 秒的條件下，校標效度為 0.92；影片長度為 90 秒的條件下，校標效度為 0.93，這意謂著量測工具量測到的 QoE 值與真人的主觀 QoE 具有高度正相關性。另一方面，也透過誤差統計方式呈現量測結果的 RMSE 與各誤差等級的百分比，量測結果與真人評分的誤差 97.75%都在 1 分以內，具有高度代表性。因此未來可採納量測工具量測取代真人評分，達到自動化與節省人力成本的目的。

在第五章中我們透過辦理一場座談會、兩場說明會、三場教育訓練，促進產、官、學界的交流。在座談會中我探討目前國內影音市場的規模與趨勢，以及國內外政策法規，並且針對國際間影音服務的 QoS 與 QoE 的量測方法進行介紹，在與蒐集先進們的問題與建議後，本案提出針對影音服務品質的 QoE 量測方法 (vMOS)。並且舉辦說明會與教育訓練，透過舉辦這些活動讓社會大眾更清楚了解本案所執行的內容與成效。

總結上述，現在各國對於影音服務產業的發展並未有嚴格的限制與管理，且線上影音服務的結構並非由單一業者提供完整的服務，因此消費者使用影音服務的得到的品質若有問題，無法釐清問題點規責何處在。本案透過研析國際間網路的 QoS 與 QoE 的量測方法，採用電信技術中心既有研發之「視訊服務品質綜合指標(Video Mean Opinion Score;vMOS)」來進行大規模布建量測，藉由此方式驗證此指標的可信度與大規模布建的可行性是有高度正向性的。期在產業應用上除了能協助進行判斷 QoE 不佳的可能因素外，亦能作為串流影音服務的監控與改善方面的參考。藉由自動化量測可以讓使用者、服務供應商或監理單位掌握服務品質。監理單位也可將這樣的量測方法訂定相關的服務品質規範，要求服務供應商定期公告服務品質測試結果，並且進行改善以保障使用者權益。

本案目前已驗證 vMoS 之 QoE 量測方法的確可以透過設備端 APP 量測進行大規模布建量測，並此方法收集之指標數據具有信效度之基礎，未來研究方面可以建議提供給有經營線上影音業務之有線電視、固網或行網業者，導入到業者的機上盒系統或平台系統，並透過業者提供給消費者之機上盒或線上影音系統平台收集使用者 QoE 數據。通傳會可以將相關的數據收集並分析，以了解消費者感受狀況，當消費者有所疑慮時，通傳會可以有完整數據支持與說明，一方面可以讓本國規管業者提供 vMoS 品質保證之影音服務，提升產業價值與競爭力。同時進而消彌消費爭端與維護消費者權益。

參考文獻

[1] DIGITIMES（2011/10/31）。Smart TV 應用環境與技術發展。檢自 https://www.digitimes.com.tw/iot/article.asp?cat=130&id=0000256722_kg46jrky3vlzbd1o4yhe5

[2] 江明晏（2012/11/29）。OTT 潮流 台灣大攻行動影音，大紀元電子報。檢自 http://www.epochtimes.com/b5/12/11/29/n3741240.htm

[3] 葉志良（2015）。我國線上影音內容管制的再塑造：從 OTT 的發展談起。*資訊社會研究*，29 期，頁 47-92。

[4] 曾俐穎、陳人傑（2015）。眼球經濟新藍海：影音 OTT 平臺產業發展模式之研究。2015 中華傳播學會年會暨第 12 屆傳播與媒體生態學術研討會，高雄義守大學。

[5] 葉志良（2015）。從歐盟網路中立性法制發展談網路創新與管制意涵。*東海大學法學研究*，46 期，頁 151-227。

[6] 許琦雪（2015/4/30）。OTT 競爭下有線電視產業的危機與轉機—跳脫傳統電視框架。*NCC News*，104 年 4 月號。

[7] 江明晏（2015/8/5）。台灣大攜手凱擘 OTT 服務明年問世，Yahoo 奇摩新聞。檢自 https://tw.news.yahoo.com/台灣大攜手凱擘-ott 服務明年問世-083338994--finance.html

[8] 葉志良、何明軒（2016）。OTT 產業政策白皮書。元智大學大數據與數位匯流創新中心政策法規研究團隊。

[9] 黃晶琳（2016/1/30）。遠傳 friDay 影音 多螢幕吸睛，聯合新聞網。檢自 http://udn.com/news/story/7240/1476029

[10] 何英煒（2016/7/27）。台灣大：Q3 獲利可望優於 Q2，中時電子報。檢自 http://www.chinatimes.com/newspapers/20160727000208-260206

[11] 何佩珊（2016/11/22）。不靠打賞、不捧網紅，為什麼他們也要做直播，數位時代。檢自 http://www.bnext.com.tw/article/41958/the-reason-why-they-build-live-platform

[12] 林淑惠（2016/12/19）。新聞分析－攻 OTT 5 大電信揪眾打群架，中時電子報。檢自 http://www.chinatimes.com/newspapers/20161219000041-260202

[13] 張家華（2017）。電信業者進入 OTT 市場之競爭策略研究 —以台灣大哥大 myVideo 為例。國立中正大學電訊傳播研究所碩士論文。

[14] 張恩齊（2017）。OTT 平臺重度使用者經驗之研究。世新大學資訊傳播研究所碩士論文。

[15] OVO 展雋創意股份有限公司（2017）。OTT 網路電視數據報告。檢自 https://www.ovotv.com/blog/zh/2017/09/15/ovodata2017q3/

[16] 陈楚雄, 柯江毅與覃道满, “视频业务体验评估和优化提升探讨（Discussion on Video Service Experience Evaluation and Optimization）,” 邮电设计技术, no. 2, pp. 17-23, 2017.

[17] TechNews, “OVO 月觀看突破 700 萬次，公布台灣首份 OTT 網路電視數據報告,” Sept. 15, 2017。檢自 http://technews.tw/2017/09/15/ovo-taiwan-internet-tv-report/

[18] 柯思瑪, “你付費看影集了嗎？串流影音平臺火力全開！,” 數位時代, Nov. 24, 2017。檢自 https://www.bnext.com.tw/article/46954/video-stream-netflix

[19] 黃晶琳（2017/12/17）。台灣 4G 傳輸量 稱冠全球，經濟日報。檢自 https://money.udn.com/money/story/8888/2879232

[20] 資誠聯合會計師事務所（2018）。2018 全球與臺灣娛樂暨媒體業展望報告。檢自 https://www.pwc.tw/zh/publications/topic-report/outlook.html

[21] 中華電信（2018/3/28）。中華電信數位匯流事業處成立揭牌 強攻數位匯流龍頭。檢自 https://www.cht.com.tw/zh-tw/home/cht/messages/2018/msg-180328-162206

[22] 台灣大哥大新聞中心（2018/5/7）。慶 APP 下載數破 350 萬 myVideo 把整個城市變成電影院。檢自 https://corp.taiwanmobile.com/press-release/news/press_20180508_767429.html

[23] 雲端暨聯網電視論壇（2018/7/3）。多螢媒體與匯流政策的對話—政策建言白皮書。

[24] Almes, S. Kalidindi and M. Zekauskas. (1999). “A One-way Delay Metric for IPPM,” *IETF RFC 2679*.

[25] Almes, S. Kalidindi and M. Zekauskas. (1999). “A Round-trip Delay Metric for *IPPM,” IETF RFC 2681*.

[26] Almes, S. Kalidindi and M. Zekauskas. (1999). “A One-way Packet Loss Metric for IPPM,” *IETF RFC 2680*.

[27] Apple Inc. (2013). HTTP live streaming technical overview 2013. Retrieved from https://developer.apple.com/library/ios/documentation/networkinginternet/conceptual/streamingmediaguide/Introduction/Introduction.html

[28] Adobe Systems Inc. (2013). HTTP dynamic streaming 2013. Retrieved from http://www.adobe.com/products/hds-dynamic-streaming.html

[29] Akamai. (2017). The Science Behind How Our Bodies React to Video Quality. Retrieved from https://content.akamai.com/gl-en-pg9246-sensum-whitepaper.html

[30] B. Constantine et al. (1998). “Framework for TCP Throughput Testing,” *IETF RFC 6349, Aug. 2011.Kostas*, T.J., Borella, M.S. Sidhu, I. Schuster, G.M., Grabiec, J. & Mahler, J. Real-time voice over packet-switched networks. IEEE Network, 12（1）, 18-27.

[31] Body of European Regulators for Electronic Communications. (2011). A framework for Quality of Service in the scope of Net Neutrality. BoR （11） 53.

[32] B. Constantine et al. (2011). “Framework for TCP Throughput Testing,” *IETF RFC 6349*.

[33] Balachandran. (2013). “Developing a Predictive Model of Quality of Experience for Internet Video,” *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 339-350.

[34] Belias, V. (2013). A Study on QoE for Multimedia Systems. Bachelor Thesis on Informatics. Alexander Technological Educational Institute of Thessalniki.

[35] Baah-Acheamfuor, K. (2014). Quality of Service and Quality of Experience in Fixed-Line and Mobile Multimedia Services. The 13th Regulatel--BEREC Summit Meeting, Buenos Aires.

[36] Argentina, retrieved from http://www.regulatel.org/wordpress/wp-content/uploads/2014/09/martes/kwame_baah-acheamfuo.pdf

[37] Bennett, R. (2015). Assessing Minimum Quality of Service. Retrieved from http://hightechforum.org/assessing-minimum-quality-of-service/

[38] BEREC. (2017). “Net Neutrality Regulatory Assessment Methodology,” BoR （17） 178.

[39] Crawley et al. (1998). “A Framework for QoS-Based Routing in the Internet,” IETF RFC 2386.

[40] C. Demichelis and P. Chimento. (2002). “IP Packet Delay Variation Metric for IP Performance Metrics （IPPM）,” *IETF RFC 3393*.

[41] Cisco. (2017). “Cisco Visual Networking Index: Forecast and Methodology, 2016 – 2021,” Retrieved from https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html

[42] Chhabra, S. (2018). CRTC establishes network quality framework for basic internet service rollout. Retrieved from https://mobilesyrup.com/2018/07/13/

[43] D. Paxson et al. (1998). "Framework for IP Performance Metrics," *IETF RFC 2330*.

[44] DASH Industry Forum. (2013). For promotion of MPEG-DASH 2013. Retrieved from http://dashif.org

[45] ETSI, "Network Aspects （NA）; General Aspects of Quality of Service （QoS） and Network Performance （NP）.(1994)." Tech. rep. ETR003, 2nd ed..

[46] Eden. (2007). "No-Reference Estimation of the Coding PSNR for H.264-coded Sequence," *IEEE Trans. on Consumer Electronics*, vol. 53, no. 2, pp. 667-674.

[47] Electronic Communications and Postal Regulatory Authority.(2017). The State of Internet in France. France.

[48] Electronic Communications and Postal Regulatory Authority (2018). The State of Internet in France. France.

[49] Federal Communications Commission . (2016) FCC Technological Advisory Council. Retrieved from https://transition.fcc.gov/bureaus/oet/tac/tacdocs/

[50] Gómez et al. (2014). "YouTube QoE evaluation tool for Android wireless terminals," *EURASIP Journal on Wireless Communications and Networking*.

[51] H.J. Kim et al. (2008). "The QoE Evaluation Method through the QoS-QoE Correlation Model," *Proc. IEEE International Conference on Networked Computing and Advanced Information Management*, Gyeongju, South Korea, Sept. 2-4, pp. 719-725.

[52] H.J. Kim and S.G. Choi. (2010). "A Study on a QoS/QoE Correlation Model for QoE Evaluation on IPTV Service," *Proc. IEEE International Conference*

on Advanced Communication Technology, Phoenix Park, South Korea, Feb. 7-10, pp. 1377-1382.

[53] Huawei. (2015). Mobile Video Service Performance Study （White Paper）.

[54] Huawei. (2016) Video Experience-based Bearer Network （White Paper）.

[55] Huawei. (2016). Mobile Bearer Network Requirements for Mobile Video Services （White Paper）.

[56] Holznagel, B. & Hartmann, S. (2016). The EU ‘open Internet access’ regulation and its impact on the digital press. Convergence: The International Journal of Research into New Media Technologies. 22（5）, 488-493.

[57] Hotel Avala Resort &Villas Budva, Montenegro. (2016). International Regulatory Conference for Europe Regulating Electronic Communication Market, Milan JANKOVIC. Retrieved from https://pdfs.semanticscholar.org/presentation/afb2/f1802870e134caf697947ae2c5ee22a8fda9.pdf

[58] ITU-T Rec. E.800. (1993) “Terms and Definitions Related to Quality of Service and Network Performance Including Dependability”.

[59] ITU-R Rec. BT.500-10. (2000). Methodology for Subjective Assessments of the Quality of Television Picture.

[60] ITU-T Rec. P.910. (2008). Subjective video quality assessment methods for multimedia applications.

[61] ITU-T Rec. P.913. (2014). Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment.

[62] ITU-T Rec. J.343. (2014). Hybrid perceptual bitstream models for objective video quality measurements.

[63] ITU-T Rec. J.343.1. (2014). Hybrid-NRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of encrypted bitstream data.

[64] ITU-T Rec. J.343.2. (2014). Hybrid-NR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of non-encrypted bitstream data.

[65] ITU-T Rec. J.343.3. (2014). Hybrid-RRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and encrypted bitstream data.

[66] ITU-T Rec. J.343.4. (2014). Hybrid-RR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and non-encrypted bitstream data.

[67] ITU-T Rec. J.343.5. (2014). Hybrid-FRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and encrypted bitstream data.

[68] ITU-T Rec. J.343.6. (2014). Hybrid-FR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and non-encrypted bitstream data.

[69] International Telecommunication Union (2017). Quality of Service Regulation Manual. Switzerland: ITU.

[70] J. Gozdecki, A. Jajszczyk, and R. Stankiewicz. (2003). "Quality of Service Terminology in IP networks," IEEE Comm. Magazine.

[71] Jiang et al. (2016). "A Practical Prediction System for Video QoE Optimization," *Proc. USENIX Symposium on Networked Systems Design and Implementation*, Santa Clara, CA, March 16-18, pp. 137–150.

[72] Jankovic, M. (2016). Regulatory challenges related to the Quality of Service and Experience. International Regulatory Conference for Europe Regulating

Electronic Communication Market, Retrieved from
https://www.itu.int/en/ITU-D/Regional-Presence/Europe/Documents/

[73] Janevski, T. & Jankovic, M. (2017). Manual on Quality of Service Regulation. Telecommunication Development Sector, International Telecommunication Union.

[74] Ketykó et al. (2010). "QoE measurement of mobile YouTube video streaming," *Proc. Workshop on Mobile video delivery*, Firenze, Italy, Oct. 25, pp. 27-32.

[75] Klein, J., Freeman, J., Morland, R. & Revell, S. (2011). Traffic Management and Quality of Experience. Project commissioned by Ofcom, conducted by Technologia. Retrieved from
https://www.ofcom.org.uk/__data/assets/pdf_file/

[76] Khan, L. Sun and E. Ifeachor. (2012). "QoE Prediction Model and its Application in Video Quality Adaptation Over UMTS Networks," *IEEE Transactions on Multimedia*, vol. 14, no. 2, pp. 431–442.

[77] Li, M., Yeh, C. & Lu, S. (2018). Real-Time QoE Monitoring System for Video Streaming Services with Adaptive Media Playout. Hindawi International Journal of Digital Multimedia Broadcasting,
https://doi.org/10.1155/2018/2619438

[78] Morton, "Round-Trip Packet Loss Metrics. (2012)." *IETF RFC 6673*.

[79] M. Seufert et al. (2014). "A survey on quality of experience of HTTP adaptive streaming," *IEEE Comm. Surveys Tut.*, vol. 17, no. 1, pp. 469–492.

[80] M.-N. Garcia et al. (2014). "Quality of experience and HTTP adaptive streaming: A review of subjective studies," *Proc. IEEE Int. Conf. Quality Multimedia Exp.*, pp. 141–146.

[81] M.N. Garcia, D. Dytko and A. Raake. (2014). "Quality Impact due to Initial Loading, Stalling, and Video Bitrate in Progressive Download Video

Services," *Proc. IEEE International Workshop on Quality of Multimedia Experience*, Singapore, pp. 129-134.

[82] M. Seufert et al. (2014). "A survey on quality of experience of HTTP adaptive streaming," *IEEE Comm. Surveys Tut.*, vol. 17, no. 1, pp. 469–492.

[83] Mux. (2017). 2017 VIDEO STREAMING PERCEPTIONS REPORT. Retrieved from https://static.mux.com/downloads/2017-Video-Streaming-Perceptions-Report.pdf

[84] Poretsky et al., "Terminology for Benchmarking Network-layer Traffic Control Mechanisms," IETF RFC 4689, Oct. 2006.Geddes, M. (2015). How should regulators measure broadband quality? Retrieved from http://www.martingeddes.com/how-should-regulators-measure-broadband-quality/

[85] Pal and V. Vanijja. (2017). "Effect of Network QoS on User QoE for a Mobile Video Streaming Service Using H.265/VP9 Codec," *Procedia Computer Science*, vol. 111, pp. 214-222.

[86] Qadir, Q.M., Kist, A.A. & Zhang, Z. (2015). Mechanisms for QoE Optimization of Video Traffic: A Review Paper. Australian Journal of Information, Communication Technology and Applications, 1（1）, 1-18.

[87] Q. Wang et al. (2018). "Data Analysis on Video Streaming QoE over Mobile Network," *EURASIP Journal on Wireless Communications and Networking*.

[88] R. Serral-Gracià et al. (2010). "An Overview of Quality of Experience Measurement Challenges for Video Applications in IP Networks," *Proc. International Conference on Wired/Wireless Internet Communications*, Luleå, Sweden, pp. 252-263.

[89] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang. (2011). "Measuring the quality of experience of HTTP video streaming," *Proc. IFIP/IEEE*

International Symposium on Integrated Network Management, Dublin, Ireland.

[90] R. K. P. Mok et al. (2011). "Inferring the QoE of HTTP video streaming from user-viewing activities," *Proc. ACM SIGCOMM workshop on Measurements up the stack*, Toronto, Ontario, Canada, pp. 31-36.

[91] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang. (2011). "Measuring the quality of experience of HTTP video streaming," *Proc. IFIP/IEEE International Symposium on Integrated Network Management*, Dublin, Ireland.

[92] Robitza, W., Ahmad, A., Kara, P., Atzori, L., Martini, M., Raake, A. & Sun, L. (2017). Challenges of Future Multimedia QoE Monitoring for Internet Service Providers. Multimedia Tools and Applications, 76（21）, 22243-22266.

[93] Rodríguez, M. & Muñoz, E. (2017). Review of Quality of Service （QoS） mechanisms over IP Multimedia Subsystem （IMS）. Ingeniería y Desarrollo, 35（1）, 262-281.

[94] Schulzrinne et al. (1996). "RTP: A transport protocol for real-time applications," *IETF RFC 1889*.

[95] S. Bradner and J. McQuaid. (1999). "Benchmarking Methodology for Network Interconnect Devices," *IETF RFC 2544*.

[96] Szigeti, T. & Hattingh, C. (2004). Quality of Service Design Overview. End-to-End QoS Network Design. Cisco Press.

[97] S. Poretsky et al. (2006). "Terminology for Benchmarking Network-layer Traffic Control Mechanisms," *IETF RFC 4689*.

[98] S.S. Krishnan and R.K. Sitaraman. (2012). "Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs," *Proc. Internet Measurement Conference （IMC）*, Boston, MA, USA.

[99] S.S. Krishnan and R.K. Sitaraman. (2012). "Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs," *Proc. Internet Measurement Conference （IMC）*, Boston, MA, USA, Nov. 14-16.

[100] Shen,Y., Liu, Y., Qiao, N., Sang, L. &Yang, D. (2012). QoE-based Evaluation Model on VideoStreaming Service Quality, in: IEEE Globecom Workshops （GC Wkshps）.Retrieved from https://10.1109/GLOCOMW.2012.6477772.

[101] SamKows. (2015) "SamKnows Test Methology （Document Reference: SQ301-005-EN）".

[102] SamKows. (2015). "SamKnows Smartphone-based Testing: SamKnows App for Android （Document Reference: SQ312-003-EN）

[103] Sandvine. (2017). Reveal network quality with Scoreboard, White Paper. Retrieved from https://www.sandvine.com/hubfs/downloads/solutions/analytics-

[104] Takahashi, A. (2008). Standardization Activities in the ITU for a QoE Assessment of IPTV. IEEE Communications Magazine. Feb. 2008, 78-84.

[105] T. Wang, A. Pervez and H. Zou, "VQM-based QoS/QoE Mapping for Streaming Video. (2010)." *Proc. IEEE International Conference on Broadband Network and Multimedia Technology*, Beijing, China, pp. 807-812.

[106] Takahashi, A. (2015). Standardization Activities for QoS/QoE-related Technologies in ITU-T. 13（6） NTT Technical Review. 1-3.

[107] V. Jacobson, K. Nichols and K. Poduri. (1999). "An Expedited Forwarding PHB," *IETF RFC 2598*.

[108] Vlessing, E. (2014). Netflix offers peek at biz as standoff continues with CRTC. Retrieved from http://mediaincanada.com/2014/09/23/as-crtc-netflix-stand-off-continues-u-s-streamer-offers-sneak-peek-at-business/.

[109] Video Quality Experts Group. (2014). Hybrid Perceptual/Bitstream Validation Test Final Report.

[110] W. C. Hardy. (2001). QoS Measurement and Evaluation of Telecommunications Quality of Service, Wiley.

[111] Webach, K. (2009). New Rules for a New Age: Creating an "Economic Stimulus Agency" out of the FCC Knowledge. Retrieved from http://knowledge.wharton.upenn.edu/article/new-rules-for-a-new-age-creating-an-economic-stimulus-agency-out-of-the-fcc/

[112] White Paper on Network Traffic Management and the Evolving Internet. (2010). Committee on Communications Policy. Retrieved from https://ieeeusa.org/wp-content/uploads/2017/07/IEEEUSAWP-NTM2010.pdf

[113] Y. Chang et al. (2010). "Radar chart: Scanning for high QoE in QoS dimensions," *Proc. IEEE International Workshop Technical Committee on Communications Quality and Reliability*, Vancouver, Canada.

[114] Y. Shen et al. (2012). "QoE-based Evaluation Model on Video Streaming Service Quality," *Proc. IEEE Globecom Workshop*, Anaheim, CA, USA.

[115] Y. Chen, K. Wu, and Q. Zhang. (2015). "From QoS to QoE: A Tutorial on Video Quality Assessment," IEEE Comm. Surveys Tut, vol. 17, no. 2, pp. 1126-1165.

[116] Y. Sun et al. (2016). "Improving video bitrate selection and adaptation with data-driven throughput prediction," *Proc. ACM SIGCOMM, Florianopolis*, Brazil, Aug. 22-26, pp. 272–285.

[117] Zambelli. (2009). Smooth streaming technical overview. Retrieved from http://www.iis.net/learn/media/on-demand-smooth-streaming/smoothstreamingtechnical-overview

[118] Z. Duanmu et al. (2017). "A Quality-of-Experience Index for Streaming Video," *IEEE Journal on Selected Topics in Signal Processing*, vo. 11, no. 1.

[119] Z. Duanmu et al. (2017). "A Quality-of-Experience Index for Streaming Video," *IEEE Journal on Selected Topics in Signal Processing*, vo. 11, no. 1.

[120] Zboralska, E. & Davis, C. (2017). Transnational over-the-top video distribution as a business and policy disruptor: The case of Netflix in Canada. The Journal of Media Innovation, 4（1）, 4-25.

附錄

附錄一 FCC Guidance on Open Internet Transparency Rule Requirements

附錄二 Telecom Decision CRTC 2018-241: Non-consensus report on quality of service metrics to define high-quality fixed broadband Internet access service

附錄三 Technologia: Traffic management and quality of experience

附錄四 Ofcom: Measuring mobile voice and data quality of experience

附錄五 PNS: A study of traffic management detection method & tools

附錄六 Actual Experience: Investigation of Internet Quality of Experience

附錄七 Arcep State of Internet in France 2017: Quality of service, from the perspective of UFC-Que CHOISIR

附錄八 Arcep State of Internet in France 2018: A high quality Internet to support innovation

附錄九 IMDA Code of Practice for television broadcast standards

附錄十 座談會會議記錄

附錄十一 說明會會議記錄

附錄十二 教育訓練會議記錄

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄一

FCC Guidance on Open Internet Transparency Rule Requirements

PUBLIC NOTICE

Federal Communications Commission
445 12th St., S.W.
Washington, D.C. 20554

News Media Information 202 / 418-0500
Internet: https://www.fcc.gov
TTY: 1-888-835-5322

DA 16-569
May 19, 2016

GUIDANCE ON OPEN INTERNET TRANSPARENCY RULE REQUIREMENTS

GN Docket No. 14-28

In this Public Notice, the Chief Technologist, Office of General Counsel, and Enforcement Bureau (collectively the Bureaus) offer guidance regarding acceptable methodologies for disclosure of network performance to satisfy the enhanced transparency requirements in the *2015 Open Internet Order.*[1] Further, the Bureaus offer guidance on compliance with the point of sale disclosure requirement under the *2015 Open Internet Order*. This guidance is intended to clarify what disclosure practices will satisfy the Transparency Rule.[2]

In the *2010 Open Internet Order*, the Commission concluded that effective disclosure of network management practices, performance characteristics, and commercial terms helps to ensure that "broadband providers will abide by open Internet principles," enhances the general public's and the Commission's ability to identify and address open Internet violations, and correspondingly increases "the chances that harmful practices will not occur."[3] The Commission further found that these disclosures empower consumers and promote competition and investment, further reducing broadband Internet access service (BIAS) providers' incentives and ability to engage in harmful conduct.[4]

To achieve these objectives, the Commission adopted the Transparency Rule:

---

[1] *Protecting and Promoting the Open Internet*, GN Docket No. 14-28, Report and Order on Remand, 30 FCC Rcd 5601, 5673-75, para. 166 (2015) (*2015 Open Internet Order*), *pets. for review pending sub nom USTA v. FCC No. 15-1063* (D.C. Cir. filed May 22, 2015). When it adopted the *2015 Open Internet Order*, the Commission cited approvingly previous staff guidance "in interpreting and applying the general requirements of the transparency rule" and anticipated that "further such guidance may be appropriate concerning the transparency rule." *Id.* at 5682, para. 185. Parties have raised concerns about disclosures regarding network performance. *See*, *e.g.*, AT&T PRA Comments at 4-5, 12, 17-18, 19-20, 22-27 (July 20, 2015); CTIA PRA Comments at 13-14, 24-25 (July 20, 2015); Mobile Future PRA Comments at 6-7 (July 20, 2015); USTelecom PRA Comments at 9-10, 14-15 (July 20, 2015). Anticipating the need to address these kinds of concerns, the Order delegated to the Chief Technologist authority to provider further guidance on the subject. *2015 Open Internet Order*, 30 FCC Rcd at 5673-75, para. 166.

[2] The Transparency Rule under the *2010 Open Internet Order* remains in effect. The Commission will publish a notice in Federal Register announcing an effective date for the transparency rule enhancements upon approval of the modified information collection by the Office of Management and Budget.

[3] *Preserving the Open Internet*, GN Docket No. 09-191, WC Docket No. 07-52, Report and Order, 25 FCC Rcd 17905, 17937, 17941, paras. 56, 59 (2010) (*2010 Open Internet Order*), *aff'd in relevant part Verizon v. FCC*, 740 F.3d 623 (D.C. Cir. 2014).

[4] *Id.* at 17940, para. 59.

A person engaged in the provision of broadband Internet access service shall publicly disclose accurate information regarding the network management practices, performance, and commercial terms of its broadband Internet access services sufficient for consumers to make informed choices regarding use of such services and for content, application, service, and device providers to develop, market, and maintain Internet offerings.[5]

In 2011 and in 2014, Commission staff released advisory guidance to assist BIAS providers in complying with the Transparency Rule.[6] The *2011 Advisory Guidance* gave fixed and mobile BIAS providers guidance regarding acceptable methodologies for disclosure of network performance metrics to satisfy the Transparency Rule.[7] For example, fixed BIAS providers could use Measuring Broadband America (MBA) data, or "disclose actual performance based on internal testing; consumer speed test data; or other data regarding network performance, including reliable, relevant data from third-party sources."[8] Mobile providers who had access to reliable data regarding their network performance could "disclose the results of their own or third-party testing."[9] For those providers without access to reliable data, the *2011 Advisory Guidance* allowed disclosure of "a Typical Speed Range (TSR) representing the range of speeds and latency that can be expected by most of their customers, for each technology/service tier offered, along with a statement that such information is the best approximation available to the broadband provider."[10] The *2011 Advisory Guidance* further stated that the *2010 Open Internet Order* does not require distribution of disclosure materials in hard copy or extensive training of sales employees to provide the disclosures themselves.[11] It also stated that BIAS providers can comply with the point of sale requirement by "directing prospective customers at the point of sale, orally and/or prominently in writing, to a web address at which the required disclosures are clearly posted and appropriately updated."[12]

The text of the codified Transparency Rule, which was upheld by the D.C. Circuit Court in *Verizon v. FCC*,[13] was not changed in the *2015 Open Internet Order*. Instead, the *2015 Open Internet Order* enhanced the Transparency Rule by clarifying certain aspects of the rule including disclosure of specific commercial terms, performance characteristics, and network management practices.[14] Among other things, BIAS providers are specifically required to disclose expected and actual download and upload speeds, latency, and packet loss, but are no longer required to disclose the typical frequency of

---

[5] 47 CFR § 8.3.

[6] *See FCC Enforcement Bureau and Office of General Counsel Issue Advisory Guidance for Compliance with Open Internet Transparency Rule*, Public Notice, 26 FCC Rcd 9411, 9414-15 (2011) (*2011 Advisory Guidance*); *FCC Enforcement Advisory, Open Internet Transparency Rule: Broadband Providers Must Disclose Accurate Information to Protect Consumers*, Public Notice, 29 FCC Rcd 8606, 8607 (2014) (*2014 Advisory Guidance*). To the extent they are not superseded by the *2015 Open Internet Order* or this guidance, the *2011 Advisory Guidance* and *2014 Advisory Guidance* continue to apply to enhancements made to the Transparency Rule. *See* Letter from Thomas Cohen, Counsel for the American Cable Association, to Marlene H. Dortch, Secretary, FCC, GN Docket No. 14-28, at (filed Oct. 2, 2015) (requesting clarification regarding the effectiveness of past guidance).

[7] *2011 Advisory Guidance*, 26 FCC Rcd at 9411-18.

[8] *Id.* at 9414-15.

[9] *Id.* at 9415.

[10] *Id.*

[11] *Id.* at 9413.

[12] *Id*. at 9414.

[13] *See Verizon v. FCC*, 740 F.3d 623.

[14] *See 2015 Open Internet Order*, 30 FCC Rcd at 5672-79, paras. 162-175.

congestion.[15] In addition, the *2015 Open Internet Order* reconfirmed—but did not modify—the Transparency Rule requirements around disclosure at the point of sale, requiring "at a minimum, the prominent display of disclosures on a publicly available website and disclosure of relevant information."[16]

In May 2015, the Commission issued the *PRA Notice*[17] seeking comment on the information collection requirements for the enhancements to the Transparency Rule made in the *2015 Open Internet Order*. The comments filed in response reflected concern regarding some of the requirements.[18] Some areas of particular concern were the geographic and peak hour disclosure requirements and the packet loss disclosure.[19] They also noted concerns about the lack of a mobile Measuring Broadband America safe harbor and uncertainty about whether the point of sale disclosure requirement had changed.[20]

To provide further clarity regarding the Transparency Rule, this Public Notice offers guidance for compliance with certain aspects of the Transparency Rule. We emphasize that, with the exception of the requirements for participation in the MBA safe harbor, any examples provided here are not exhaustive; BIAS providers may implement alternative approaches that disclose information sufficient to adequately inform consumers and relevant third parties.

I. NETWORK PERFORMANCE METRICS

A. Guidance on Disclosure of Network Performance Metrics

Service Tiers. The *2015 Open Internet Order* clarified that, "for mobile broadband providers, the obligation in the Transparency Rule to disclose performance characteristics for 'each broadband service' refers to separate disclosures for services with each technology (e.g., 3G and 4G)".[21] We similarly clarify that, for fixed BIAS providers, the obligation in the Transparency Rule to disclose performance characteristics for "each broadband service" refers to separate disclosures for services with each *technology* (e.g., digital subscriber line (DSL), cable, fiber, or satellite) and *service tier* (e.g., 50 Mbps download / 10 Mbps upload), as it is the combination of technology and service tier that BIAS providers use to market the service.[22]

Disclosure of Actual Network Performance Metrics. The *2015 Open Internet Order* requires all BIAS providers to disclose both *expected* and *actual* download and upload speeds, latency, and packet loss for each service.[23] Here, we give guidance on disclosure of *actual* network performance metrics.

---

[15] *Id.*

[16] *Id.* at 5677, para. 171.

[17] Federal Communications Commission, Information Collection Being Reviewed by the Federal Communications Commission, 80 Fed. Reg. 29000 (May 20, 2015) (PRA Notice).

[18] *See, e.g.*, AT&T PRA Comments at 4-6, 17-18, 20-27, 29-30, 32-33, 37-41; CTIA PRA Comments at 13-14, 17-25; Mobile Future PRA Comments at 6-7, 9; USTelecom PRA Comments at 9-10, 14-15.

[19] *See* AT&T PRA Comments at 4-6, 17-18, 20-27, 30; CTIA PRA Comments at 13-14, 22, 24-25; Mobile Future PRA Comments at 6, 9; USTelecom PRA Comments at 9-10, 14-15.

[20] *See* AT&T PRA Comments at 5, 20, 29; CTIA PRA Comments at 13; Mobile Future PRA Comments at 6-7.

[21] *2015 Open Internet Order*, 30 FCC Rcd at 5674, para. 166.

[22] *See 2010 Open Internet Order*, 25 FCC Rcd at 17939, para. 56 ("*Service Description*: A general description of the service, including the service technology, expected and actual access speed and latency, and the suitability of the service for real-time applications.").

[23] *See 2015 Open Internet Order*, 30 FCC Rcd at 5673-74, paras. 165-166. The MBA program measures speed by the throughput over a five second time window, latency by the round trip time between an end user and an off-net measurement server, and packet loss by the percentage of packets transmitted from an end user to a measurement

The Measuring Broadband America (MBA) program, which may be used by fixed BIAS providers as a safe harbor in meeting the requirement to disclose *actual* network performance,[24] presents both median speeds and percentiles of speed.[25] We clarify that providers may comply with the requirement to disclose *actual* speeds—both download and upload—by disclosing either the median speed or a range of actual speeds that includes the median speed (e.g., 25th to 75th percentile). However, we note that speed ranges may be more appropriate when there is substantial variation in speed, e.g. for fixed BIAS using DSL technology[26] and for mobile BIAS.[27] Similarly, latency may be disclosed using either the median latency or a range of actual latencies that includes the median latency (e.g., 25th to 75th percentile). If speed or latency ranges are used, the percentiles used to determine the endpoints of the ranges must also be disclosed.[28] Packet loss may be disclosed as the average packet loss.[29] In order to ensure that the *actual* and *expected* network performance metrics can be compared, it is best to provide *actual* and *expected* performance in comparable formats. For example, if *actual* download speed is provided as a range, the *expected* download speed should use a range with the same percentile endpoints.

Geographic Granularity for Actual Network Performance Metrics. The *2015 Open Internet Order* requires that disclosures of *actual* speed, latency and packet loss "be reasonably related to the performance the consumer would likely experience in the geographic area in which the consumer is purchasing service."[30] BIAS performance may vary by location.[31] For fixed BIAS, however, except for fine-grained variations in performance based on the distance between a consumer and network equipment, commenters agree that there are few variations in actual BIAS performance across a BIAS provider's service area for a particular combination of technology and service tier unless the BIAS provider is using different network management practices in different geographical areas.[32] Therefore, we clarify that fixed BIAS providers may meet this requirement by disclosing *actual* performance metrics for "each

---

server for which no acknowledgement was received. The detailed methodology is available in the Technical Appendix of the 2015 Measuring Broadband America Report. FCC, 2015 Technical Appendix; Measuring Broadband America Fixed Broadband: A Report on Consumer Fixed Broadband Performance in the US (2015), http://data.fcc.gov/download/measuring-broadband-america/2015/Technical-Appendix-fixed-2015.pdf.

[24] *2015 Open Internet Order*, 30 FCC Rcd at 5674-75, n.411.

[25] *See* FCC, 2015 Measuring Broadband America Fixed Broadband Report: A Report on Consumer Fixed Broadband Performance in the US at 30-36 (2015), http://data.fcc.gov/download/measuring-broadband-america/2015/2015-Fixed-Measuring-Broadband-America-Report.pdf (2015 Fixed MBA Report).

[26] *See id.* at 8.

[27] *See 2015 Open Internet Order*, 30 FCC Rcd at 5674, n.409 ("Given that the performance of mobile broadband networks is subject to a greater array of factors than fixed networks, we note that disclosure of a range of speeds may be more appropriate for mobile broadband consumers."), n.410 ("Per the 2011 Advisory Guidance, those mobile broadband providers that 'lack reasonable access' to reliable information on their network performance metrics may disclose a 'Typical Speed Range (TSR)' to meet the requirement to disclose actual performance.").

[28] The percentiles used should be sufficient for consumers to make informed choices regarding the use of such services and for content, application, service, and device providers to develop, market, and maintain Internet offerings.

[29] *See* 2015 Fixed MBA Report at 18.

[30] *2015 Open Internet Order*, 30 FCC Rcd at 5674, para. 166.

[31] *See* 2015 Fixed MBA Report at 14-17.

[32] *See* AT&T PRA Comments at 20-21; Letter from Fred Baker, Fellow, and Russ Gyurek, Director, Cisco Systems Inc., to Marlene H. Dortch, Secretary, FCC, GN Docket No. 14-28 (Nov. 2, 2015) (Cisco *Ex Parte*) at 2; Letter from Steven F. Morris, Vice President and General Counsel, NCTA, to Marlene H. Dortch, Secretary, FCC, GN Docket No. 14-28 (Nov. 25, 2015) (NCTA *Ex Parte*) at 2.

broadband service"[33] in each geographic area in which the service has a distinctive set of network performance metrics (operational area). We expect that operational areas will be determined by the technology used and by network management practices, and that many fixed BIAS providers will have a single operational area for each broadband service offered.

For mobile BIAS, the *2015 Open Internet Order* stated that, "with the exception of small providers, mobile broadband providers can be expected to have access to reliable actual data on performance of their networks representative of the geographic area in which the consumer is purchasing service – through their own or third-party testing – that would be the source of the disclosure" of actual network performance.[34] Mobile BIAS performance may vary based on a BIAS provider's access to spectrum in various geographic areas.[35] We therefore clarify that mobile BIAS providers with access to reliable actual data on network performance may meet this requirement by disclosing *actual* performance metrics for each Cellular Market Area (CMA) in which the service is offered, as further described below.[36]

Disclosure of Expected Network Performance Metrics. BIAS providers are also required to disclose *expected* download and upload speeds, latency, and packet loss.[37] We clarify that there is no corresponding requirement to disclose different *expected* network performance metrics in different geographic areas. However, to ensure that information regarding performance is accurate and sufficient for consumers to make informed choices, *expected* network performance disclosed for a geographic area should not exceed *actual* network performance in that geographic area.

Peak Usage Periods. The *2015 Open Internet Order* stated the Commission's expectation that network performance would be measured "during times of peak usage".[38] Some commenters asked whether times of peak usage would need to vary by geographic area, e.g. according to whether the area is dominantly commercial or residential.[39] We clarify that peak usage periods may be based solely on the local time zone.[40] We further clarify that BIAS providers retain flexibility to determine the appropriate peak usage periods for their network performance metrics but must disclose the peak usage periods chosen for such disclosures.

B. Guidance for BIAS Providers using the Measuring Broadband America Safe Harbor.

Fixed BIAS Network Performance. Participation in the Measuring Broadband America (MBA) program remains a safe harbor for fixed BIAS providers in meeting the requirement to disclose actual network performance.[41] As a result, fixed BIAS providers may disclose their results from the MBA program, for each service for which the program provides network performance metrics, as a sufficient

---

[33] *See supra* Part I.A (discussing service tiers).

[34] *2015 Open Internet Order*, 30 FCC Rcd at 5674, para. 166.

[35] *See, e.g.*, Andrea Goldsmith, Wireless Communications 505-527 (2005).

[36] *See infra* Parts I.B, I.C.

[37] *2015 Open Internet Order*, 30 FCC Rcd at 5673-74, paras. 165-166.

[38] *Id.* at 5674, para. 166.

[39] *See* ATT PRA Comments at 25-27; CTIA PRA Comments at 24-25.

[40] *See* 2015 Fixed MBA Report at 32 (defining the peak usage period for the 2015 Fixed MBA Report as between 7:00 p.m. and 11:00 p.m. local time).

[41] *2015 Open Internet Order*, 30 FCC Rcd at 5674-75, n.411.

representation of actual download and upload speeds, actual latency, and actual packet loss of those services.[42]

Mobile BIAS Network Performance. The *2015 Open Internet Order* stated that the MBA program could at the appropriate time be declared a safe harbor for *mobile* BIAS providers in meeting the requirement to disclose actual network performance.[43] The MBA program has been measuring mobile BIAS performance since November 2013.[44] We anticipate that the MBA program will publish its first Mobile Broadband Report in 2016 for those services for which it has a sufficient national sample size. The program will provide, at a minimum, network performance metrics for each such service for each CMA in which the program has a sufficient CMA sample size, and additional sets of these network performance metrics aggregated among sets of other CMAs.[45] Today, we establish that mobile BIAS providers may disclose their results from the mobile MBA program as a sufficient disclosure of actual download and upload speeds, actual latency, and actual packet loss of a service[46] if the results satisfy the above sample size criteria and if the MBA program has provided CMA-specific network performance metrics of the service in CMAs with an aggregate population of at least one half of the aggregate population of the CMAs in which the service is offered.

C. Guidance for BIAS Providers not using the Measuring Broadband America Safe Harbor.

Fixed BIAS Network Performance. As initially articulated in the *2011 Advisory Guidance*, fixed BIAS providers not using the MBA safe harbor may disclose actual network performance metrics based on the MBA methodology, "internal testing; consumer speed test data; or other data regarding network performance, including reliable, relevant data from third-party sources."[47] In the *2015 Open Internet Order*, the Commission explained that disclosed actual network performance metrics should be based on measurements during peak usage periods.[48] It also explained that any of these methodologies may be used

---

[42] *See 2011 Advisory Guidance*, 26 FCC Rcd at 9414-9415. Fixed BIAS providers should inform the MBA program for each broadband service offered if there is more than one operational area so that the MBA program may provide separate sets of network performance metrics as needed. The safe harbor is available only for those services for which the program provides network performance metrics, which is subject to the program's policies regarding the minimum number of subscribers to a service and the minimum number of panelists in a service.

[43] *2015 Open Internet Order*, 30 FCC Rcd at 5674, para. 166.

[44] *See* Press Release, FCC, Fact Sheet: FCC Unveils New, Free Speed Test App to Empower Consumers with U.S. Mobile Broadband Performance Information (Nov. 14, 2013), https://www.fcc.gov/document/fcc-unveils-mobile-broadband-speed-test-app-empower-consumers.

[45] *See 2015 Open Internet Order*, 30 FCC Rcd at 5674, n.410 ("In any event, we expect that mobile broadband providers' disclosure of actual performance data will be based on accepted industry practices and principles of statistical validity."). The Chief Technologist, with consultation of the Wireless Telecommunications Bureau and the Office of Engineering and Technology, will determine policies regarding sufficient national and CMA sample sizes.

[46] There is no requirement to disclose different *expected* network performance metrics in different geographical areas. For instance, a mobile BIAS provider may choose to advertise a single set of network performance metrics nationally, providing the advertised *expected* performance does not exceed the *actual* network performance in each CMA in which the MBA program has provided CMA-specific metrics, and does not exceed the aggregate *actual* network performance in each set of other CMAs. Alternatively, it may choose to advertise one set of network performance metrics in some CMAs and another set in other CMAs, provided that each such advertisement similarly does not exceed the actual network performance in the corresponding geographical areas. *See supra* Part 1.A.

[47] *2011 Advisory Guidance* 26 FCC Rcd at 9414-15.

[48] *2015 Open Internet Order*, 30 FCC Rcd at 5674, para. 166.

as the basis for measurement of actual download and upload speeds, actual latency, and actual packet loss, provided that the methodology be disclosed and be grounded in commonly accepted principles of scientific research, good engineering practices, and transparency.[49]

Mobile BIAS Network Performance. Commenters note that there is a tradeoff between the likely per capita cost incurred to obtain performance metrics in a geographic area and the population density in that geographic area.[50] We interpret the requirement for disclosures of actual speed, latency, and packet loss to "be reasonably related to the performance the consumer would likely experience in the geographic area in which the consumer is purchasing service"[51] to be satisfied by sufficient disclosures of aggregate actual network performance in low population density areas. Specifically, mobile BIAS providers that, instead of taking advantage of the MBA safe harbor, measure network performance by their own or third-party testing may disclose performance metrics for each CMA in which the service is offered, except that actual network performance may be aggregated among CMAs with a population density below 250 people per square mile.[52]

Routes. The Transparency Rule requires a BIAS provider to disclose "accurate information regarding the network management practices, performance, and commercial terms of its broadband Internet access services sufficient for consumers to make informed choices regarding use of such services and for content, application, service, and device providers to develop, market, and maintain Internet offerings."[53] The routes over which network performance metrics are measured should thus be chosen to accurately represent the actual network performance experienced by consumers within the designated geographic area. The *2010 Open Internet Order* noted that "our rules apply only as far as the limits of a broadband provider's control over the transmission of data to or from its broadband customers".[54] The *2015 Open Internet Order* noted that "congestion may originate beyond the broadband provider's network and the limitations of a broadband provider's knowledge of some of these performance [metrics]".[55] Therefore, a sufficient representation of actual network performance of the service may be obtained from measurements of speed, latency, and packet loss on a representative sampling of routes between end users and the points of interconnection with edge providers or other networks. Fixed BIAS providers may, for example, measure speed, latency, and packet loss between measurement clients in broadband modems and measurement servers that are located in close proximity to the links on which traffic is exchanged with edge providers or other networks. As an alternative to placing measurement clients in broadband modems, a fixed BIAS provider may place measurement clients in access networks, provided that this internal testing accurately measures performance metrics in a manner that represents the performance experienced by consumers of the service. Mobile BIAS providers may, for example, achieve a representative sampling of end users by running measurement clients on end-user devices (e.g.,

---

[49] *Id.* at 5675, n.412.

[50] *See* AT&T PRA Comments at 16-18.

[51] *2015 Open Internet Order*, 30 FCC Rcd at 5674, para. 166.

[52] There is no requirement to disclose different *expected* network performance metrics in different geographical areas. For instance, a mobile BIAS provider may choose to advertise a single set of network performance metrics nationally, providing the advertised *expected* performance does not exceed the *actual* network performance in each CMA with a population density above 250 people per square mile, and does not exceed the aggregate *actual* network performance in all other CMAs. Alternatively, it may choose to advertise one set of network performance metrics in some CMAs and another set in other CMAs, provided that each such advertisement similarly does not exceed the actual network performance in the corresponding geographical areas. *See supra* Part I.A.

[53] 47 CFR § 8.3.

[54] *2010 Open Internet Order*, 25 FCC Rcd at 17933, n.150.

[55] *2015 Open Internet Order*, 30 FCC Rcd at 5675-76, para. 168.

using consumer speed test data), by placing measurement clients in locations near a representative set of mobile broadband consumers (e.g., by combining drive-test data with an estimate of the reduction in speed from drive-test locations to user locations), or by measuring performance in the network and estimating the relationship between this measured performance and that experienced by end users.

II. POINT OF SALE REQUIREMENT

The Transparency Rule, as adopted in the *2010 Open Internet Order*,[56] requires BIAS providers to disclose network management practices, performance characteristics, and commercial terms "at the point of sale."[57] As discussed in more detail below, this requirement was not modified in the *2015 Open Internet Order*. As the Commission has explained previously, BIAS providers are not required to provide hard copies of the disclosures required under the Transparency Rule at the point of sale and instead may direct consumers to links to online disclosures.[58] However, for disclosures made through such links to be sufficient under the Transparency Rule, BIAS providers must ensure that consumers actually receive the information necessary to make informed decisions prior to making a final purchasing decision at all potential points of sale, including in a store, over the phone, and online.

Background. During the rulemaking proceeding leading up to the adoption of the Transparency Rule in the *2010 Open Internet Order*, some commenters raised concerns that a point of sale requirement could be interpreted to compel distribution of physical materials at retail outlets and extensive training of sales employees at retail outlets as well as at telephone and Internet sales centers that may be operated by BIAS providers or other third parties. The Commission addressed those concerns in the *2010 Open Internet Order* by stating that "broadband providers must, at a minimum, prominently display or provide links to disclosures on a publicly available, easily accessible website that is available to current and prospective end users and edge providers."[59] The Commission further explained that "broadband providers may be able to satisfy the transparency rule through a single disclosure."[60]

In the *2011 Advisory Guidance*, the Commission's Office of General Counsel and Enforcement Bureau issued joint guidance on the Transparency Rule, including how the point of sale requirement should be interpreted.[61] The *2011 Advisory Guidance* states that the *2010 Open Internet Order* does not require distribution of disclosure materials in hard copy, or extensive training of sales employees to provide the disclosures themselves.[62] It also states that BIAS providers can comply with the point of sale requirement by "directing prospective customers at the point of sale, orally and/or prominently in writing, to a web address at which the required disclosures are clearly posted and appropriately updated."[63]

The text of the Transparency Rule, upheld by the D.C. Circuit Court in *Verizon v. FCC*,[64] was not changed in the *2015 Open Internet Order*. The *2015 Open Internet Order*, however, reconfirms that the Transparency Rule requires "at a minimum, the prominent display of disclosures on a publicly available

[56] *2010 Open Internet Order,* 25 FCC Rcd at 17936-41, paras. 53-61.

[57] *2010 Open Internet Order,* 25 FCC Rcd at 17940, para. 57.

[58] *2011 Advisory Guidance,* 26 FCC Rcd at 9413-14.

[59] *2010 Open Internet Order,* 25 FCC Rcd at 17939-40, paras. 57-58 & n.186 ("[W]e expect that broadband providers will make disclosures in a manner accessible by people with disabilities.").

[60] *Id.* at 17940, para. 58.

[61] *2011 Advisory Guidance*, 26 FCC Rcd at 9411.

[62] *Id.* at 9413.

[63] *Id*. at 9414.

[64] *See Verizon v. FCC*, 740 F.3d 623.

website and disclosure of relevant information."[65] While the Commission made some enhancements to the disclosures required under the rule in the *2015 Open Internet Order*, the requirements relating to point of sale disclosures were not modified.[66]

In response to the *PRA Notice*,[67] several commenters expressed concern about perceived changes to the point of sale disclosure requirements.[68] Specifically, these commenters point to language in a footnote in the *2015 Open Internet Order*, which states that:

> [b]roadband providers must actually disclose information required for consumers to make an "informed choice" regarding the purchase or use of broadband services at the point of sale. *It is not sufficient for broadband providers simply to provide a link to their disclosures*.[69]

This language apparently raised concerns that the Commission intended to eliminate the option for BIAS providers to provide point of sale disclosures by directing prospective customers to a web address and that providers would be required under the enhanced Transparency Rule to print and distribute paper materials at every point of sale location.[70] According to these commenters, such a disclosure requirement would be unduly burdensome, particularly for prepaid wireless services because there is little space available on packaging containing prepaid devices and services.[71] These commenters contend that wireless BIAS providers cannot reasonably be expected to present all or even a portion of the detailed disclosures required by the Commission under the Transparency Rule on packaging.[72] CTIA notes that mobile BIAS providers often rely on multiple, distinct sales channels (e.g., online outlets, carrier stores, big box stores, small retail shops, independent dealers, and call centers)[73] to sell their products and services, and that each of these channels likely would have different systems and processes that would need to be modified in order to comply with a requirement to distribute hard copies of all disclosures.[74]

Discussion. We clarify that in the *2015 Open Internet Order*, the Commission did not change the requirement that disclosures required under the Transparency Rule be prominently displayed "on a publicly available website" and made at the point of sale, nor did the Commission eliminate the permissibility of making disclosures at the point of sale by directing consumers to a website link.[75]

The Commission's statement that "it is not sufficient for broadband providers simply to provide a

[65] *Id.* at 5677, para. 171.

[66] *Id.* at 5672-79, paras. 162-175.

[67] PRA Notice, 80 Fed. Reg. 29000.

[68] *See*, *e.g.*, CTIA PRA Comments at 6, AT&T PRA Comments at 17.

[69] *2015 Open Internet Order*, 30 FCC Rcd at 5677, n.424 (emphasis added).

[70] AT&T PRA Comments at 35, CTIA PRA Comments at 10, USTelecom PRA Comments at 11.

[71] *Id.*

[72] CTIA PRA Comments at 20.

[73] *Id.* USTelecom also states that "typical providers have multiple point of sale channels including store fronts (provider-owned and other retail outlets such as Best Buy and Wal-Mart), call centers, and website. To bar notice by a website link to customers who seek service via website or telephone, in particular, would add considerable layers and cost to the disclosure process." USTelecom PRA Comments at 15.

[74] USTelecom PRA Comments at 15.

[75] *2015 Open Internet Order,* 30 FCC Rcd at 5677, para. 171; *2011 Advisory Guidance*, 26 FCC Rcd at 9413.

link to their disclosures" was intended to explain that, while disclosures may be made via a link to a website, for those disclosures to be meaningful, BIAS providers must ensure that consumers actually receive any Open Internet-related information that is relevant to their purchasing decision at all potential points of sale, including in a store, over the phone, and online. BIAS providers should ensure that their point of sale disclosure methods actually lead potential customers to the relevant disclosure information so that informed purchasing decisions can be made by those customers.

Safe Harbor for Form of Disclosure to Consumers. In the *2015 Open Internet Order*, the Commission determined that there should be a voluntary safe harbor for the format and nature of the disclosures to consumers required under the Transparency Rule. To take advantage of the safe harbor, a BIAS provider must provide a consumer-focused, standalone disclosure.[76] The Commission tasked the Consumer Advisory Committee with developing and recommending a disclosure format.[77] The safe harbor disclosure format was approved by the relevant Bureaus and publicly released on April 4, 2016.[78] Accordingly, a BIAS provider that wishes to qualify for the safe harbor must use the safe harbor format for disclosures made at the point of sale, but also may choose not to take advantage of the safe harbor. Thus, a provider may choose to make the disclosures in the safe harbor format at retail outlets and ensure that consumers have access to disclosures in that format at the other various points of sale.

-FCC-

[76] *2015 Open Internet Order,* 30 FCC Rcd at 5680, para. 179.

[77] *Id.* at 5680-81, paras. 179-180.

[78] *Consumer and Governmental Affairs, Wireline Competition, and Wireless Telecommunications Bureaus Approve Open Internet Broadband Consumer Labels*, GN Docket No. 14-28, Public Notice, DA 16-357 (CGB/WCB/WTB 2015) (*Consumer Broadband Label Notice*).

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄二

Telecom Decision CRTC 2018-241: Non-consensus report on quality of

Canadian Radio-television and Telecommunications Commission
Conseil de la radiodiffusion et des télécommunications canadiennes

Telecom Decision CRTC 2018-241

PDF version

Ottawa, 13 July 2018

Public record: 8663-C12-201503186 and 8621-C12-01/08

CISC Network Working Group – Non-consensus report on quality of service metrics to define high-quality fixed broadband Internet access service

With this decision, the Commission further defines the universal service objective by establishing the broadband quality of service (QoS) that should be provided to all Canadians. The Commission determines that to meet the broadband portion of the universal service objective, fixed broadband Internet access service is defined as a high-quality service if it provides the subscriber with a smooth experience when using real-time QoS-critical applications, as described in this decision. Specifically, the Commission establishes a round-trip latency threshold of 50 milliseconds, and a packet loss threshold of 0.25%, both based on measurement during peak times. The Commission is launching, today, a separate proceeding to establish an appropriate QoS metric for jitter.

Introduction

1. In Telecom Regulatory Policy 2016-496, the Commission determined that the availability of fixed broadband Internet access service offerings that meet certain levels of speeds, data allowance, and quality of service (QoS) will help ensure that Canadians are receiving services that meet their needs and enable them to participate in the digital economy. Accordingly, the Commission established a universal service objective: Canadians, in urban areas as well as in rural and remote areas, have access to voice services and broadband Internet access services, on both fixed and mobile wireless networks. To measure the successful achievement of this objective, the Commission established several criteria, including,

   - Canadian residential and business fixed broadband Internet access service subscribers should be able to access speeds of at least 50 megabits per second (Mbps) download and 10 Mbps upload, and to subscribe to a service offering with an unlimited data allowance; and

   - the latest generally deployed mobile wireless technology should be available not only in Canadian homes and businesses, but also on as many major transportation roads as possible in Canada.

2. The Commission also determined that the QoS levels for latency,[1] jitter,[2] and packet loss[3] need to be established to define high-quality fixed broadband Internet access service and measure the successful achievement of the broadband portion of the universal service objective, in addition to the above-mentioned criteria. The Commission considered that the CRTC Interconnection Steering Committee (CISC) would offer an opportunity for many different parties with technical expertise to provide input on appropriate QoS metrics and measurement methodology.

3. Accordingly, the Commission requested that CISC review and make recommendations on appropriate metrics for latency, jitter, and packet loss to define high-quality fixed broadband Internet access service. These recommendations were to include (i) technical specifications, (ii) the identification of points of interconnection in the Internet service providers' (ISPs) networks where these metrics would apply, and (iii) the methods by which data on the service metrics could be collected and reported by ISPs in a consistent manner. The Commission expected that the QoS metrics would reflect the objective that broadband Internet access services in rural and remote areas be of similar high quality as those in urban areas.

Report

4. The CISC Network Working Group (NTWG) submitted the following non-consensus report, dated 29 November 2017, for the Commission's consideration:

   - *Develop recommendations as to the appropriate metrics and reporting to define high-quality fixed broadband Internet access service* (NTRE061) [the NTWG Report]

5. The NTWG Report can be found under the "Reports" section of the NTWG page, which is available under the CISC section of the Commission's website at www.crtc.gc.ca.

Issues

6. The Commission has identified the following issues to be addressed in this decision:

   - What constitutes an ISP's broadband Internet access network?

---

[1] Latency refers to the time it takes for data packets to travel from a source to a destination. Latency is usually measured in terms of the round trip, i.e. from a source to a destination and back to the source.

[2] Jitter refers to the variation in latency that causes data packets that were sent at regular intervals from a source to arrive at a destination at irregular intervals.

[3] Packet loss refers to the number of data packets that are sent from a source that fail to reach their intended destination.

- What measurement methodology should be used?
- What are appropriate metrics to define high-quality fixed broadband Internet access service?

What constitutes an ISP's broadband Internet access network?

Positions of parties

7. NTWG participants (hereafter, "parties") noted that typically, the ISP supplies a modem or gateway at the customer premises, or customers purchase their own modem.[4] They indicated that the modem is the starting point of a customer's home wireless (Wi-Fi) or wired local area network. The customer's home network directly connects to the customer's computers, laptops, smartphones, tablets, video game consoles, and potentially many other devices. The parties noted that the devices that customers use could affect QoS, but since these devices are not supplied by the ISPs, they are not within the ISPs' control. As a result, the parties submitted that the customer's home network is not part of an ISP's fixed broadband Internet access network.

8. SSi Micro Ltd. and the Eeyou Communications Network submitted that the transport networks[5] should be included in QoS measurement. This is because ISPs providing broadband Internet access service far from Canadian Tier 1 cities[6] have to purchase Internet Protocol (IP) transit services, typically from larger ISPs, to connect to the appropriate Internet exchange point (IXP).[7] These parties also submitted that their choice of IP transit service provider is largely based on balancing the quality and cost of the service; hence, the transit network could have a significant impact on QoS performance and the end-customer experience.

9. Bell Canada noted that not all ISPs are connected to IXPs in Canadian Tier 1 cities and that some Canadian ISPs are connected to United States-based IXPs. Parties indicated that some ISPs exchange traffic with each other at private interconnection points in Canadian Tier 1 cities.

10. Parties agreed that including the QoS of the global Internet beyond Canadian Tier 1 cities would not be appropriate, since this does not constitute Canadian ISPs' fixed broadband Internet access network. Parties noted that it would be impossible for

---

[4] The term "modem" in this decision refers to either a stand-alone modem or a device that is a combination of a modem and router.

[5] Transport networks are also referred to as Internet Protocol transit networks.

[6] The current Tier 1 cities, based on the consensus recommendation in the NTWG Report, are Moncton, Halifax, Toronto, Ottawa, Montréal, Winnipeg, Saskatoon, Edmonton, and Vancouver.

[7] The IXP is where multiple ISPs connect to exchange Internet traffic with other ISPs in Canada and with the global Internet.

Canadian ISPs to measure QoS beyond Canadian Tier 1 cities into the global Internet. Consequently, parties agreed that broadband QoS should be measured in Canadian Tier 1 cities.

11. The Canadian Internet Registration Authority (CIRA), along with Fenwick McKelvey (Concordia University), the Cree Nation / Eeyou Communications Network, and Herb Charles (independent consultant) [collectively, CIRA et al.], Rogers Communications Canada Inc. (RCCI), Clearcable Networks, and SamKnows Ltd. (SamKnows)[8] recommended that the IXPs in Canadian Tier 1 cities should be the end-point of ISPs' fixed broadband Internet access network where QoS should be measured.

Commission's analysis and determinations

12. The primary purpose of an ISP's broadband Internet access network is to connect broadband service subscribers to the Internet. Once connected, subscribers can access various Internet-based services and applications, hosted in Canada and globally.

13. A typical ISP network starts from the customer premises to an IXP or a private interconnection point in Canada.

14. The Commission considers that a customer's home network and devices beyond the modem are not part of ISPs' fixed broadband Internet access network, since they are typically not supplied by or within the control of the ISPs but could affect QoS measurements. As such, it would not be appropriate for broadband QoS measurement to include the performance of customer devices and the customer's home network. Accordingly, the Commission determines that QoS measurement starting from the customer premises should take place at the modem.

15. ISPs that use an IP transit service typically do so to carry their customers' Internet traffic to or from an IXP. The ISP is responsible for providing or choosing the IP transit service provider or routes; therefore, the IP transit service is part of ISPs' fixed broadband Internet access network.

16. It was not the Commission's goal to determine the QoS of the global Internet or United States-based IXPs. The Commission's goal is to measure the QoS of Canadian ISPs' fixed broadband Internet access service, and since the IXPs in Canadian Tier 1 cities are well-established points of interconnection where ISPs typically interconnect to exchange Internet traffic within Canada and with the rest of the global Internet, these IXPs are appropriate end-points for QoS measurement. The

[8] SamKnows is a broadband performance measurement company based in the United Kingdom that has built a global Internet measurement platform. SamKnows also conducts broadband measurement in Canada as part of the Commission's Broadband Measurement Project.

record indicates that ISPs that do not connect to the IXPs in Canadian Tier 1 cities can establish these connections specifically for the purpose of QoS measurement.

17. In light of all the above, the Commission determines that the parts of an ISP's fixed broadband Internet access network to which QoS measurement should apply include all network elements from the modem at the customer premises to a point of interconnection at an IXP in a Canadian Tier 1 city.

What measurement methodology should be used?

Background

18. The measurement methodology identifies (i) the points of interconnection in the ISPs' networks where QoS is measured, and (ii) the method by which QoS performance data should be collected and reported by ISPs across Canada, in a consistent manner.

19. The Commission launched the Broadband Measurement Project in 2015 to objectively measure broadband Internet performance in Canadian homes. This Project is a collaboration between the Commission and major Canadian ISPs. SamKnows conducted the broadband QoS measurement study in Canada on behalf of the Commission, and submitted two broadband measurement reports (one in December 2016 and one in April 2016) [hereafter, "the broadband measurement reports"].

20. While the primary focus of the Broadband Measurement Project is to measure actual Internet connection speeds, ISPs' performance data (latency, packet loss, and jitter) are also measured. In its CRTC Three-Year Plan 2017-2020, the Commission indicated that it would continue to collect performance data from participants and expand the Project to include more ISPs and performance measurement parameters.

Positions of parties

21. Parties proposed the three QoS measurement options below.

Option 1: From the customer premises to an IXP

22. The Canadian Network Operators Consortium Inc. (CNOC), CIRA et al., Clearcable Networks, RCCI, and SamKnows submitted that broadband QoS measurement from the customer premises, either at the modem or the customer's computer/device, to a measurement server[9] located off-net[10] at the IXP is the only

[9] The measurement server refers to the equipment located in the ISP's network or at the IXP to which the measurement probes connect (see footnote 11 for a definition of measurement probes). The measurement server collects and stores the measurement results, among other things.

[10] This refers to a location at an IXP that marks the end of an ISP's network. Since this location is outside the ISP's network, it is referred to as being "off-net."

way to ensure that the full extent of the ISP's network is measured and is the best practice in broadband QoS measurement.

23. CIRA submitted that to support broadband QoS measurement, it has deployed an Internet Performance Test in the form of off-net measurement servers at the IXPs in all the Canadian Tier 1 cities, and that these servers can be used for free by any ISP in Canada in the manner proposed in Option 1.

24. CIRA et al. recommended Option 1 since it aligns with third-party broadband QoS measurement initiatives, such as the measurement study conducted by SamKnows, and CIRA's Internet Performance Test. These initiatives rely on established measurement standards that are already used in regulatory contexts globally and have clear explanations of their methodology. CIRA et al. also submitted that the Commission's Broadband Measurement Project should be part of an ongoing QoS monitoring system by the Commission.

25. SamKnows noted that it uses Option 1 exclusively in almost all the work it carries out for telecommunications regulators globally, including for the Commission's Broadband Measurement Project. SamKnows supported Option 1 since it ensures that the same measurement software is used consistently across all measurement probes.[11]

26. Bell Canada, Cogeco Communications Inc., Quebecor Media Inc., and TELUS Communications Inc. (collectively, Bell Canada et al.) and CNOC expressed the concerns that deploying third-party probes in customers' homes in every community throughout Canada is not cost effective, and that it would be difficult to recruit volunteers who would permit such devices to be installed in their homes. Bell Canada et al. also noted that such measurement would not be completed within reasonable time frames.

27. The parties that supported Option 1 argued that they did not propose deploying third-party probes in the manner that Bell Canada et al. and CNOC were concerned about, since Option 1 involves a sample-based approach.

[11] The measurement probe refers to the measurement equipment located on the customer side of the network. The measurement probe could be a dedicated piece of equipment or software running on a broadband service subscriber's computer.

Option 2: From the access aggregation point[12] to the on-net border router[13]

28. Bell Canada et al. and the Independent Telecommunications Providers Association (ITPA) proposed that broadband QoS measurement take place at or near the access aggregation point to the on-net border router. However, they indicated that if broadband QoS measurement of the access network falls within 20 milliseconds (ms) of the latency threshold to be established by the Commission, ISPs should continue to conduct measurements at the modem in the customer premises to confirm whether or not the established latency threshold is met. If the result is more than 20 ms below the latency threshold to be established by the Commission, the households served by that central office or cable head-end should be assumed to fall below the established threshold.

29. Bell Canada et al. acknowledged that this measurement methodology was not appropriate for some ISPs, and proposed that the Commission give ISPs the choice of using Option 1 or 2, based on the resources available to them. CNOC supported giving ISPs this choice.

30. Bell Canada et al. and the ITPA proposed partial network measurements, based on the cost and effort associated with measurement and reporting. The ITPA noted that the Policy Direction[14] requires that any new regulatory measure imposed by the Commission be efficient and proportionate to its purpose and interfere with the operation of competitive market forces to the minimum extent necessary.

31. SamKnows noted that on-net measurement servers are typically not used in public reporting of ISPs' performance measurement.

32. CIRA et al. did not support Option 2. They submitted that this option does not reflect or accurately capture an ISP's broadband QoS performance, ignores the performance of critical parts of an ISP's network, or factors those parts with unsupported estimates. CIRA et al. added that Option 2 lacks transparency and does not enable the Commission to use an independent third party for broadband QoS measurement.

33. CIRA et al. noted that the parties that supported Option 2 did not address basic design components, such as the measurement protocols to be used, the sampling approach and scheduling, and the calculation of average, maximum, and minimum

---

[12] This refers to a location in an ISP's network where access network transmission lines in an area connect to aggregate traffic, such as a central office or a cable head-end for a traditional phone or cable company, respectively. It could also be a fixed wireless access base station that serves a particular area. In all cases, broadband QoS measurements are made on an interface that is used to serve multiple subscribers (tens to several hundreds).

[13] This refers to a location where an ISP (for example, a small ISP serving a community) hands off traffic to a transit IP service provider. A location or piece of equipment within the ISP's network is referred to as being "on-net."

[14] *Order Issuing a Direction to the CRTC on Implementing the Canadian Telecommunications Policy Objectives*, P.C. 2006-1534, 14 December 2006

performance. They added that Option 2 would require significant elaboration before being put into practice.

Option 3: From the access aggregation point to an off-net server at an IXP

34. Shaw Communications Inc. (Shaw) proposed that if the purpose of broadband QoS measurement is to determine ISPs' QoS performance nationally, the company supports measurement from the access aggregation point to an off-net server at an IXP. Shaw submitted that this option provides a transparent and fair measurement platform to objectively compare end-to-end services between ISPs.

35. Shaw also submitted that if the measurement results at the access aggregation point are close to the high-quality latency threshold to be determined by the Commission, a sufficient number of customer premises should be measured to confirm that the broadband Internet access service meets the threshold. As well, Shaw argued that ISPs should have the flexibility to conduct broadband QoS measurement based on the resources available to them, which may include remotely initiated measurement from the ISP-provided router gateway in a customer premises, if technically feasible, or testing by a technician on-site.

Commission's analysis and determinations

36. The Commission considers that the chosen methodology for broadband QoS measurement should accurately capture ISPs' actual broadband network access QoS performance, as well as subscribers' actual real-world experience. It should also take into account factors such as consistency (i.e. the broadband QoS results from different ISPs must be comparable or equivalent so that they can be aggregated to give a national assessment of QoS performance), accessibility (i.e. the measurement points or equipment should be accessible to all ISPs that use a given network and to third-party measurement organizations, and not require them to install their own equipment), fairness (i.e. the measurement methodology should be neutral and prevent more favourable QoS results from one ISP over another), and the burden on ISPs.

37. The parties that proposed partial network measurements in Options 2 and 3 did so primarily to reduce the cost and effort of conducting measurements. These factors should be considered in the assessment of the potential burden of broadband QoS measurement on ISPs. However, the Commission considers that for the broadband QoS measurement methodology to be accurate and fair, ISPs would have to measure their entire broadband Internet network. Option 1 is the only option through which (i) the entire network is measured, and (ii) all ISP performance is measured from the same network points. Under Options 2 and 3, some ISPs would measure only parts of their network, while other ISPs would measure their entire network, creating unfairness and inconsistency in the measurements. In addition, Options 2 and 3 are based on making assumptions regarding the parts of the network that were not being measured.

38. As such, Options 2 and 3 would lead to unfairness among ISPs in demonstrating their broadband QoS performance, since different ISPs' access networks have significantly different broadband QoS performance depending on technology and design. The use of one estimate for all ISPs would be inaccurate in measuring broadband QoS performance. Furthermore, the appropriateness and feasibility of using estimates to determine ISPs' packet loss and jitter QoS performance have not been demonstrated.

39. As well, Options 2 and 3 would not enable competitive ISPs, or those that use IP transit service, to measure their broadband QoS, since the measurement points would be in network locations to which these ISPs do not have access. Further, if these ISPs were given access to those network locations, they may incur costs to purchase, install, and maintain their measurement equipment, rather than using existing and shared equipment provided by third parties, as would be the case with Option 1.

40. Since broadband QoS measurement is a long-term activity, it is important that ISPs' performance be measured repeatedly over time. The parties that supported Options 2 and 3 noted that these methodologies can be conducted on a one-time basis as required, but not regularly, due to the burden on ISPs and completion time.

41. Furthermore, Option 1 is the only option that supports the use of an independent third party to measure broadband QoS. The Commission considers that the use of a third party ensures fairness and consistency in broadband QoS measurement, since the same measurement probes, servers, software, protocols, and algorithms would be used by all ISPs. The use of an independent third party also ensures that QoS measurement is implemented in a symmetrical and competitively neutral manner, in line with subparagraph 1(b)(iii) of the Policy Direction. The parties that supported Options 2 and 3 did not indicate which probes, servers, software, protocols, or algorithms would be used. In addition, Options 2 and 3 would not ensure consistency, symmetry, or the implementation of the QoS measurements in a competitively neutral manner.

42. In addition to determining where in an ISP's network broadband QoS should be measured, it is important to determine when these measurements should be taken. When congestion occurs, subscribers may not receive high-quality broadband Internet access service. Therefore, the period that best indicates overall QoS performance of broadband Internet access service is during peak usage times. As such, the Commission determines that all QoS measurements should be based on performance at peak times, i.e. from 7 p.m. to 11 p.m. local time on weekdays.

43. The Commission notes that CIRA et al. recommended the continuation of the Commission's Broadband Measurement Project to measure broadband QoS performance. The Project uses a measurement methodology that aligns with Option 1 and measurement during peak times. The methodology is well established and has proven to be suitable for accurate, efficient, fair, and continual broadband QoS measurement. This is evident from its use by regulators in countries that measure

broadband QoS performance, including those in the United States and the United Kingdom.

44. In light of all the above, the Commission determines that the measurement methodology used in the Commission's Broadband Measurement Project is the appropriate measurement methodology to determine ISPs' broadband QoS performance. Specifically, broadband QoS is to be measured using a sample-based approach, during peak times (i.e. from 7 p.m. to 11 p.m. local time on weekdays), and using a measurement probe at the modem in the customer premises to an off-net measurement server connected to an IXP in a Canadian Tier 1 city.

45. In addition, in its CRTC Three-Year Plan 2017-2020, the Commission indicated that it would, in every year up to 2020, continue to collect broadband performance data and publish it as part of its Broadband Measurement Project. As such, the continued use of the Project is the most efficient and least burdensome option for ISPs to collect and report broadband performance data, since it is already being used for participating ISPs (which represent over 80% of Internet subscribers). The Commission's goal is to increase ISPs' participation in the Project.

46. Consequently, the Commission determines that its Broadband Measurement Project is an appropriate means to collect and report on ISPs' broadband QoS measurements for latency, packet loss, and jitter to measure the successful achievement of the broadband portion of the universal service objective.

47. The Commission considers that, in line with subparagraph 1(a)(ii) of the Policy Direction, use of the Broadband Measurement Project for continued broadband QoS measurement is efficient and proportionate to its purpose and interferes with the operation of competitive market forces to the minimum extent necessary.

What are appropriate metrics to define high-quality fixed broadband Internet access service?

Background

48. In Telecom Regulatory Policy 2016-496, the Commission noted that real-time applications – particularly those with audiovisual functionalities – are sensitive to any degradation of the connection and require low levels of latency, jitter, and packet loss to provide a smooth experience to the Canadians who use them. High latency could result in an unsatisfactory user experience for real-time communications services, such as telephone calls or video conferencing. Similarly, high packet loss or jitter causes visible effects, such as video pixilation, sound distortion, or delays in loading Web pages.

Definition of high-quality fixed broadband Internet access service

Positions of parties

49. The NTWG noted that "high-quality" is a subjective term. It defined broadband QoS as "the collective effect of service performance which determines the degree of satisfaction of a user of the service." It defined quality of experience as "the overall acceptability of an application or service, as perceived subjectively by the end-user."

50. The NTWG stated that the best approach to developing QoS metrics objectively was to determine a "basket" of online applications that Canadians commonly use, categorize these applications based on their sensitivity to broadband QoS, and attempt to determine the broadband QoS metrics that would generally lead to a good quality of experience for end-users.

51. Based on this approach, the NTWG set out the categories of applications according to their sensitivity to QoS metrics (i.e. QoS critical, sensitive, or tolerant). Examples of QoS-critical applications are multi-player interactive games and cloud-based applications; examples of QoS-sensitive applications are conversational voice applications, conversational video applications, and Web browsers; and examples of QoS-tolerant applications are file transfers, downloads, high-quality audio streaming, and one-way video streaming.

52. The NTWG reviewed numerous standards, reports, and studies that referred to broadband QoS requirements for various applications. These materials did not identify a common threshold for a "high-quality" broadband Internet connection for specific applications. However, they did indicate a threshold below which the QoS was unacceptable.

53. While the NTWG agreed that low latency, jitter, and packet loss are desirable to provide a high-quality broadband QoS, they were unable to reach consensus on appropriate thresholds.

Commission's analysis and determinations

54. The Commission considers that broadband QoS thresholds should reflect high-quality fixed broadband Internet access service, similar to the 50 Mbps download and 10 Mbps upload speed thresholds it established in Telecom Regulatory Policy 2016-496. In that decision, the Commission recognized that meeting the universal service objective will take time and significant investment to achieve. The above-mentioned speeds do not reflect the minimum speeds that are achievable today in all of Canada.

55. By requesting industry stakeholders to develop QoS metrics, the Commission aimed to establish thresholds that would represent a high benchmark, similar to the 50/10 Mbps speeds, such that the thresholds for latency, jitter, and packet loss, in conjunction with the 50/10 Mbps speeds, would define high-quality fixed broadband Internet access service.

56. It would be contrary to the Commission's determinations in Telecom Regulatory Policy 2016-496 for broadband QoS metrics to be based on the minimum acceptable or adequate QoS required to support the various online applications used by Canadians.

57. For a fixed broadband Internet access service to be considered high quality, it must provide the subscriber with a smooth experience without any degradation of the connection when using a wide variety of real-time applications with audiovisual functionalities, which are commonly used today and will continue to be used in the future.

58. The NTWG reached consensus on the categorization of applications as QoS critical, QoS sensitive, or QoS tolerant. It further agreed on the representative types of applications that would fall within each category. The Commission accepts both this categorization and the examples of representative applications identified in the respective categories.

59. The Commission considers that high-quality fixed broadband Internet access service should be able to support QoS-critical applications. These applications are important in today's digital economy, in which most online and even some offline services are being offered using an online cloud-based model. In addition, fixed broadband Internet access service that supports QoS-critical applications can support important services, such as e-health, remote surgery, online education, teleconferencing, and teleworking through virtual private network access.

60. The QoS metrics required to support QoS-critical applications should therefore serve as the minimum thresholds to define high-quality fixed broadband Internet access service. Accordingly, the Commission determines that fixed broadband Internet access service is defined as a high-quality service if it provides the subscriber with a smooth experience when using real-time QoS-critical applications.

Latency

Positions of parties

61. Various parties proposed different latency metrics to define high-quality fixed broadband Internet access service.

62. CIRA et al. submitted that poor latency can degrade the quality of experience of delay-sensitive applications, even with a relatively high bandwidth connection. CIRA et al. recommended 50 ms as the latency threshold that would provide a high quality of experience for QoS-critical applications, based on measurement by SamKnows, as well as a report commissioned by the United Kingdom's communications regulator, Ofcom (the Ofcom Report).[15]

[15] Assessing Network Quality of Experience – Final Report, Sagentia Media Research, 25 November 2009

63. Bell Canada provided results from broadband QoS measurements in Southern Ontario on its fixed wireless network showing average latencies below 50 ms and none greater than 100 ms. Bell Canada also cited research showing that a good-quality multi-player game experience requires a latency of less than 70 ms, while 200 ms provides an adequate or acceptable experience. Bell Canada stated that remote surgery can be safely conducted with latency levels of less than 200 ms.

64. Bell Canada et al. proposed that instead of establishing QoS metrics for "high-quality" broadband Internet access service, the Commission should establish such metrics for the broadband Internet access service that is presently attainable in all or most parts of Canada. Bell Canada et al. also noted that a latency of up to 750 ms could form part of the definition of a high-quality fixed broadband Internet access service, since it provides adequate-quality voice service.

65. RCCI proposed a latency threshold of 150 ms, stating that this threshold takes into account the broad geographic and technological challenges the telecommunications industry faces in delivering consistent fixed broadband Internet access service throughout Canada.

66. CIRA et al. and Distributel Communications Limited disagreed with Bell Canada et al.'s proposed approach since it would support only an adequate-quality voice service and not QoS-critical applications.

67. CIRA provided latency measurements and maps showing the latency between certain locations in Canada (e.g. from Fort Smith, Northwest Territories, to Calgary, Alberta, [53 ms], to Toronto, Ontario [93 ms], and to Montréal, Quebec [102 ms]), which confirmed that CIRA's proposed latency metrics can be met. Clearcable Networks provided actual latencies measured between various locations in Canada[16] to Toronto, Ontario, and concluded that all of these locations have a latency of less than 100 ms to Toronto. Bell Canada also provided data on the latency between Inuvik, Northwest Territories, and Montréal, Quebec, in the range of 100 to 200 ms.

Commission's analysis and determinations

68. In Telecom Regulatory Policy 2016-496, the Commission indicated that it expected the 50/10 Mbps target speeds for fixed broadband Internet access service to be reached in an incremental manner within 10 to 15 years. Similarly, if the QoS metrics to define high-quality fixed broadband Internet access service were based on the present attainability of those metrics in all or most parts of Canada, the result would be that the lowest QoS attainable would define high-quality services. Therefore, these QoS metrics should be based on the quality of experience that subscribers receive or expect when using high-quality fixed broadband Internet access service.

---

[16] These locations are Cape Breton, Nova Scotia; Southern Yukon; as well as various locations in Alberta, British Columbia, Ontario, and Quebec.

69. The proposed 200 ms to 750 ms latency thresholds are based on evidence that these thresholds are for only an acceptable- or adequate-quality fixed broadband Internet access service. RCCI's proposed latency threshold of 150 ms takes into account the broad geographic and technological challenges the telecommunications industry faces in delivering consistent fixed broadband Internet access service throughout Canada. As a result, this threshold also represents only a medium- or adequate-quality service. Accordingly, setting the latency threshold at these levels would be contrary to the Commission's objective of defining high-quality fixed broadband Internet access service.

70. The Commission considers that CIRA et al.'s recommended threshold of 50 ms most closely aligns with the Commission's intentions based on evidence showing that this threshold is reasonable and achievable, and that it can support QoS-critical applications. As well, the broadband measurement reports indicate that the highest average peak period latency measured from subscribers of the major ISPs in Canada was below 22 ms for digital subscriber line (DSL), cable, and fibre-to-the-home (FTTH) technologies.

71. In light of all the above, the Commission establishes a round-trip latency threshold of 50 ms to define high-quality fixed broadband Internet access service and to measure the successful achievement of the broadband portion of the universal service objective. As mentioned above, this threshold is based on measurement during peak times (i.e. from 7 p.m. to 11 p.m. local time on weekdays), and from the modem in the customer premises to an IXP in a Canadian Tier 1 city.

Packet loss

Positions of parties

72. CIRA et al. noted that latency and packet loss are as important metrics to define high-quality fixed broadband Internet access service as speed. In addition, CIRA submitted evidence that for some QoS-critical applications, the subscriber's quality of experience is affected more by packet loss than by latency. CIRA et al. submitted that the Ofcom Report is the most useful resource for establishing a packet loss threshold, and recommended a threshold of 0.25% for both QoS-sensitive and QoS-critical applications, which corresponds with the latency threshold they recommended.

73. The ITPA recommended a packet loss threshold of less than 1%, but did not provide rationale. Shaw recommended a packet loss threshold of less than 0.5% based on its internal best practice for voice over Internet Protocol (VoIP) services.

74. Bell Canada et al. and RCCI did not propose a specific packet loss threshold, and instead recommended that such a threshold not be established at this time. These companies indicated that only the latency threshold should define high-quality fixed broadband Internet access service.

Commission's analysis and determinations

75. The Commission considers packet loss to be an important metric for defining high-quality fixed broadband Internet access service, since high packet loss prevents many applications from performing at a satisfactory level. The factors that can influence packet loss, such as network design and choice of technology, are under an ISP's direct control.

76. The broadband measurement reports indicate that in 2016, FTTH services yielded the lowest levels of packet loss, averaging 0.04%, while cable services averaged 0.13%, and DSL services averaged 0.17%. These packet loss levels were noted as being extremely small and imperceptible to any common Internet application.

77. CIRA et al.'s recommended packet loss threshold of 0.25% supports QoS-critical and QoS-sensitive applications. Furthermore, the evidence on the record indicates that ISPs' networks should perform at a much lower packet loss level than 0.25%, and that a threshold of 0.5% or 1% for packet loss would fall within the range of a medium-quality broadband Internet access service.

78. In light of all the above, the Commission establishes a packet loss threshold of 0.25% to define high-quality fixed broadband Internet access service and measure the successful achievement of the broadband portion of the universal service objective. As mentioned above, this threshold is based on measurement during peak times (i.e. from 7 p.m. to 11 p.m. local time on weekdays), and from the modem in the customer premises to an IXP in a Canadian Tier 1 city.

Jitter

Positions of parties

79. The ITPA was the only party that proposed a threshold for jitter (i.e. less than 5 ms), and did not provide any supporting evidence for this proposal.

80. CIRA et al. did not provide a threshold for jitter since they submitted that jitter could be compensated through buffering and better latency. Bell Canada et al., CNOC, and RCCI recommended that a specific threshold for jitter not be established at this time.

81. Valve, a major multi-player online interactive game provider, indicated that excessive jitter results in packets being out of order, which negatively impacts end-users' experience when using multi-player online interactive games.

Commission's analysis and determinations

82. The Commission considers that, consistent with its determination in Telecom Regulatory Policy 2016-496, it is necessary and important to establish a QoS threshold for jitter, in addition to the latency and packet loss thresholds.

83. Even with low latency, high jitter can lead to a poor experience for subscribers with real-time applications, such as videos, audio calls, e-health, and multi-player

interactive online games. While many of the effects of jitter can be managed by applications that buffer the data packets, buffering may itself negatively affect the subscriber's experience. Therefore, the use of buffering does not eliminate the need to establish a threshold for jitter, since low jitter reduces or eliminates the need for buffering.

84. There is insufficient data on the record for the Commission to make a determination on what threshold for jitter is appropriate. As well, the broadband measurement reports did not include any statistics regarding jitter that demonstrate Canadian ISPs' performance in this respect.

85. Accordingly, the Commission is launching a separate proceeding to establish an appropriate QoS threshold for jitter to define high-quality fixed broadband Internet access service, through Telecom Notice of Consultation 2018-242, also being issued today. As well, to ensure consistency with the established latency and packet loss QoS thresholds, the jitter threshold to define high-quality fixed broadband Internet access service must be based on the ability to support QoS-critical applications, and jitter performance during peak times (i.e. from 7 p.m. to 11 p.m. local time on weekdays), and from the modem in the customer premises to an IXP in a Canadian Tier 1 city.

Secretary General

Related documents

- *Establishment of an appropriate quality of service metric for jitter to define high-quality fixed broadband Internet access service*, Telecom Notice of Consultation CRTC 2018-242, 13 July 2018

- *Modern telecommunications services – The path forward for Canada's digital economy*, Telecom Regulatory Policy CRTC 2016-496, 21 December 2016

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄三

Technologia: Traffic management and quality of experience

Traffic management and quality of experience

MC069

This document has been prepared for Ofcom

Jeremy Klein, Jonathan Freeman, Rob Morland and Stuart Revell
7 April 2011

Version 1

Contents

Abbreviations used

Abbreviation	Explanation
3G	3rd Generation Mobile
4G	4th Generation Mobile
ADSL	Asymmetric Digital Subscriber Line technology
ATM	Asynchronous Transfer Mode
BGP	Border Gateway Protocol
B-RAS	Broadband Remote Access Server
CAPEX	Capital expenditure
CATV	Cable Television
CDN	Content Distribution Network
CMTS	Cable Modem Termination System
CoS	Class of Service
DCKTN	Digital Communications Knowledge Transfer Network
DNS	Directory Name Server
DOCSIS	Data Over Cable Service Interface Specification
DPI	Deep Packet Inspection
DSL	Digital Subscriber Line
DSLAM	Digital Subscriber Line Access Multiplexer
DWDM	Dense Wavelength Division Multiplexing
FTP	File Transfer Protocol
GGSN	Gateway GPRS Support Node
GPRS	General packet radio service
GW	Gateway
HD	High Definition (Video)
HE	Head End
HLR	Home Location Register
HSPA	High Speed Packet Access
ICT	Information and Communication technologies
IGP	Internal Gateway Protocol
IP	Internet Protocol
IPTV	Internet Protocol television
ISO	International Standards Organisation
ISP	Internet service provider
IuB	UMTS interface between RNC with the Node B
LC	Local Centre
LCON	Local Centre Optical Node
LTE	Long Term Evolution
IuCs	UMTS interface between RNC and Circuit Switched network
IuPs	UMTS interface between RNC and Packet Switched network

IuR	UMTS interface between RNCs
MAC	Media Access Control
MMOG	Massively Multiplayer Online Game
MSC	Mobile Switching Centre
NGA	Next Generation Access
Node B	Base station
NTU	Network Termination Unit
ON	Optical Node
OPEX	Operating Expenditure
P2P	Peer to peer
PC	Personal Computer
PSTN	Public Switched Telephone Network
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RBS	Radio Base Station
RC	Regional Centre
RCON	Regional Centre Optical Node
RF	Radio Frequency
RNC	Radio Network Controller
SD	Standard Definition (Video)
SDH	Synchronous Digital Hierarchy
SGSN	GPRS Support Node
SNR	Signal to Noise Ratio
STB	Set Top Box
TCP	Transmission Control Protocol
TM	Traffic Management
UDP	User Datagram Protocol
Uu	UMTS interface between User Equipment and RBS / Node B
VLAN	Virtual LAN
VLR	Visitor Location Register
VoIP	Voice over IP
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network technology

Executive Summary

This study was commissioned by Ofcom to provide detail on the technical aspects of traffic management, to explore the effects of traffic management on consumers' quality of experience, and to examine ways of measuring and characterising traffic management and connection performance. The study was initiated independently of the Broadband Stakeholder Group's deliberations on a voluntary code of practice on traffic management transparency[1].

The role of traffic management

Traffic management is used by all UK ISPs. However, in the main, traffic management currently follows a 'fair use' paradigm which is intended only to limit 'excessive' or 'unfair' use by the heaviest users. Until recently, the users considered to generate the most traffic have been running peer to peer (P2P) applications, and ISPs have therefore mainly targeted P2P in their policies. However, currently the greatest traffic growth is in video streaming. Video is widely expected to grow further in the context of internet connected TVs and the launch of hybrid TV services. Without a response of some sort, there will be congestion and a reduction in many users' QoE. In the future there will certainly be other applications that will put pressure on internet capacity.

Traffic management is one possible response to these pressures but expanding capacity is an alternative. In practice, traffic management will probably be pursued alongside capacity expansion. Not all network operators face the same cost structure in expanding capacity so we can foresee that traffic management will develop unevenly between network types. Some ISPs will mainly increase capacity while others will respond through more use of traffic management.

We did not find a strong intention among ISPs to utilise either more traffic management, or more complex traffic management. ISPs told us that they recognised traffic management brings with it technical complexity, cost and market communication consequences. Accordingly, current traffic management approaches are generally the minimum necessary to prevent excessive users from degrading the experience of the majority.

However, for the reasons cited above, it is reasonable to assume that moves in the direction of both *more* traffic management and more complex traffic management will be inevitable overall. This will not be a dramatic increase; instead, traffic management is expected to evolve. We have suggested five possible scenarios for the future of traffic management reflecting different ISP strategies. These are:

1. Fair use
2. Traffic management evolves to facilitate congestion-sensitive traffic
3. Traffic management evolves in response to growth in video streaming
4. Traffic management evolves as a business/marketing tool
5. Managed services become the norm.

Traffic management has often been opposed on net neutrality grounds as being injurious to consumers' interests. An alternative view of traffic management is that it is a way to make the consumer experience more controlled and less subject to the vagaries of congestion. By treating different types of data differently, traffic management allows the performance of applications to be

[1] http://www.broadbanduk.org/content/view/479/7/

managed individually so that the most QoS sensitive applications receive the better QoS from the network. Whereas in an unmanaged situation, consumers would tend not to be able to understand and predict the factors that affect their experience, in a traffic managed situation there is potentially more certainty and more transparency, and a better overall quality of experience for the majority of customers.

The technologies of traffic management

ISPs differ in their traffic management implementations but the basic techniques of traffic management are straightforward. In all cases there is a traffic management *decision* which is then enacted in the network as an *intervention*. The decision can take account of the type of traffic, the user's profile and the cumulative usage relative to any caps or limits that are in place. The traffic management intervention can either be to modify the traffic priority or to change the bandwidth allocated (a guaranteed minimum or to impose a maximum speed cap). These two types of intervention affect different traffic protocols differently.

Traffic management technology is reasonably mature and there is no indication that disruptive technology or breakthroughs might occur. Current technology is adequate to identify traffic types. There is no sense that 'internet abusers' are winning the battle against traffic managers.

The challenge of transparency

The principle of transparency has been broadly accepted by the ISPs that implement traffic management[2], but achieving it is not necessarily straightforward. The challenges include those listed below.

- Traffic management is often non-deterministic. The amount of traffic management and its effects on users can differ according to the level of congestion on the network. Both the amount of traffic management and its impact depend on the level of traffic at the time. For example, on one day it could be that reducing the priority of a particular class of data packet would result in increased latency and jitter. On another day, with greater congestion, there could be more traffic management applied and the impact could be that packets are lost altogether. Because traffic management *policies* alone are insufficient to fully describe the effects of traffic management, full transparency would involve providing data that describe the effects of policies over time and therefore the resulting quality of experience for users.

- Lack of standard metrics. There are no standard industry-wide metrics for the measurement and characterisation of traffic management.

- Traceable measurement is not straightforward. Making traceable measurements can be costly and impose overheads on devices and communication channels.

- Apparent complexity of impact. The way in which traffic management works and its effects on the user experience can appear complex, and not easily communicated and understood.

However there are some aspects of traffic management that are helpful to transparency. These are set out below.

- Underlying simplicity. Despite the technical complexity of implementing traffic management, there are only two interventions available to ISPs, namely to change the priorities of data packets or to change the bandwidth (data rate) allocated to a class of

[2] Including the Broadband Stakeholders Group

traffic. This suggests that a common template for describing traffic management is in principle achievable.

- In-network measurement is possible. Data networks embody the potential for in-network measurements at nodes and interfaces, and some equipment is already able to produce certain traffic statistics. However parameters such as speed, latency and jitter only have full meaning at a consumer's connection and may not be able to be measured meaningfully from within a network.

- Commonality of user behaviour. The majority of consumers use a small number of applications (e.g. email, browsing, streaming video, VoIP) so that a 'key facts' summary of how a particular package would perform should meet most people's needs.

Approaches to transparency

Taking account of the traffic management scenarios, we suggest that transparency involves three factors which cannot always be simultaneously satisfied. These are:

- accuracy
- meaningfulness
- comparability.

Implicit traffic management

As well as explicit traffic management through packet prioritisation and bandwidth management there are forms of network design which affect traffic differentially and can also be regarded as a form of traffic management. The dimensioning of networks, the partitioning of access pipes and the use of CDNs all affect QoS, and can do so in ways that discriminate between traffic types. Some of these issues can be addressed through attempts to improve transparency, some will be resolved through market competition, but others may need regulation.

Recommendations

Our recommendations on ways of measuring and characterising traffic management and connection performance - and the relationships between them - are illustrated in Figure 1 overleaf. The boxes are numbered for ease of reference and discussed below.

1. Tariff package design

Current packages include some degree of traffic management. At present we do not see a need to intervene in the design of packages per se because the combination of transparency and market competition appears sufficient. The possibility of imposing minimum connection standards has been raised by Ofcom but traffic management appears unlikely to affect applications such as email and web browsing which would, presumably, be the core of a set of minimum standards.

An argument might be made for packages being designed to be more comparable. For example, some ISPs calculate cumulative volume over a month whereas others calculate volumes over periods of hours. While comparability is indeed a transparency objective, we do not consider that this should override allowing diversity in the packages offered in the market.

2. QoS Policy Form

Despite the technical complexity of traffic management we have found that basic techniques of traffic management are sufficiently bounded for a common template to be used to describe traffic management policies. The general approach is to identify each type of traffic that is treated separately in a particular policy, and then to describe (i) the time and extent to which data packets of that type are prioritised/de-prioritised – which can range from 'guaranteed' to 'blocked', and (ii) the bandwidth specifically allocated to that traffic type – which can range from a minimum 'guaranteed' rate to a maximum 'restricted' or 'capped' rate.

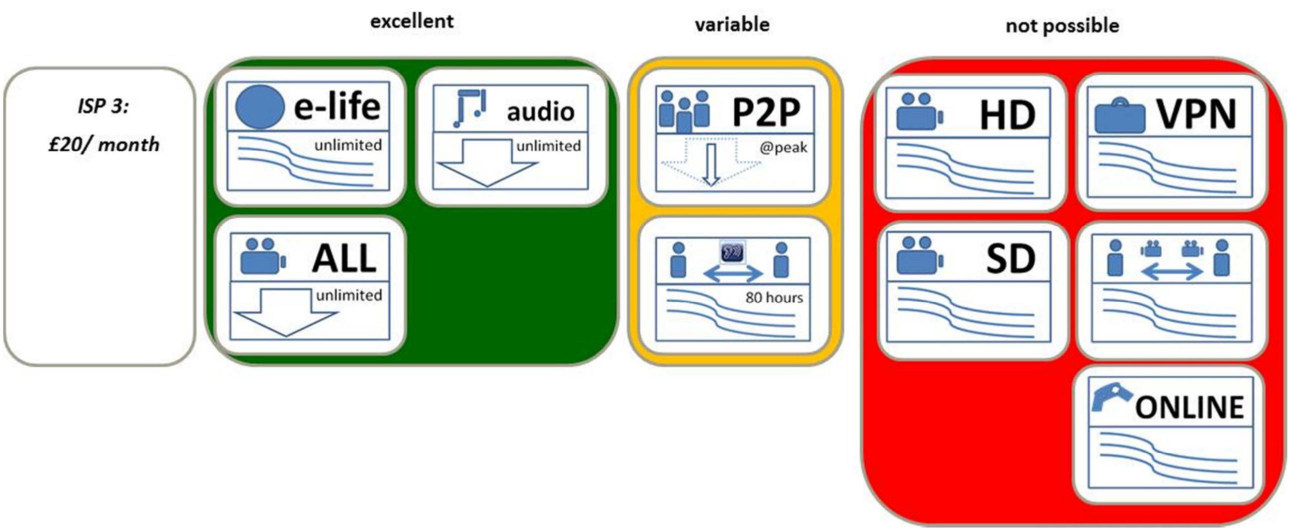
To aid meaningfulness, data rates or volume caps should be given in both bps and indicative units of consumption (e.g. the number of hours of video allowed).

Some information about traffic management cannot be specified purely with reference to an *ex ante* policy, for example, if the amount of traffic management and its effects are statistical (non-deterministic) in nature. In such cases either time series data or estimates should ideally be given in order to provide consumers with greater certainty. However the ability to do this has to be assessed on a case by case basis. Where traffic management is applied in the core, it may be difficult to identify parameters which can give a useful insight into the likely impact of traffic management on an individual customer.

The use of this template is applicable in all traffic management scenarios, though it is conceivable that in the more complex traffic management regimes envisaged in scenario 4, that the template could become large. In that case it might be preferable for consumers to use a 'wizard' though the underlying data in the QoS policy form would still be needed.

3. QoE Summary

To provide consumers with a more meaningful (but less accurate) description when choosing between packages there should be a visual representation of the quality of experience likely to be achieved. An example of such a representation is shown below. The translation between QoS and QoE will need to use standardised values (see box 6 below).

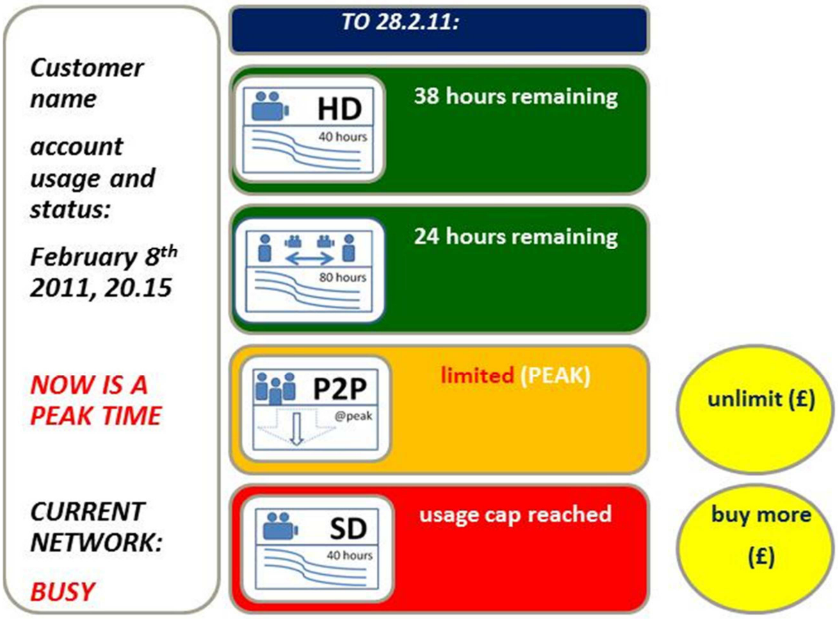


4. ISP-generated in-network measurements and status information (real time and historic)

Data networks have the capability to measure certain traffic statistics. Measurements of performance from within the network can use either software embedded in nodes and interface cards or extra equipment. These are collectively called ‘probes’. ISPs should be encouraged to measure performance and status information, and to provide it both in real time and as historic time series. Where new services are being launched, or there are no measurement data, then ISPs should construct a model or take a series of occasional measurements in order to provide estimates.

5. Real time connection status dashboard

Ideally, consumers should be able to see information about their connection and where they stand in relation to volume limits etc. in real time. One such possibility is illustrated below. Not all network architectures currently support this functionality and it would be costly for some ISPs to implement this. Consumers on tariff packages with simple, high, volume limits are in less need of this information than consumers with complex or low volume limits.



6. Standard QoE thresholds

The translation of QoS into QoE is a necessary step in producing the QoE Summary and will need to be done consistently across ISPs if the QoE Summary is to have value. While such a translation could be undertaken by Ofcom we think that industry bodies such as the Broadband Stakeholder Group should be in a position to agree standards. The translation will need to be updated regularly in line with technical developments and user trends.

7. SamKnows-type measurements

SamKnows[3]. uses monitoring facilities at a user's connection to measure performance and send data back to be aggregated. SamKnows currently only measures QoS for a few traffic types but in principle the same technique could be deployed on different traffic types to give a relative measure of traffic management within the network.

8. Wizard

At present most packages are sufficiently simple that using a 'wizard' to assist consumer decision-making is not essential. If complex packages emerge (such as those suggested in traffic management scenario 4) then wizards could be required. Assuming this is left to third parties, ISPs may be requested such organisations to produce data specifically designed to be input into a wizard.

[3] We will refer to the approach as SamKnows because it is the best example of this approach in the UK, though it is no doubt possible for similar techniques to be used by other companies

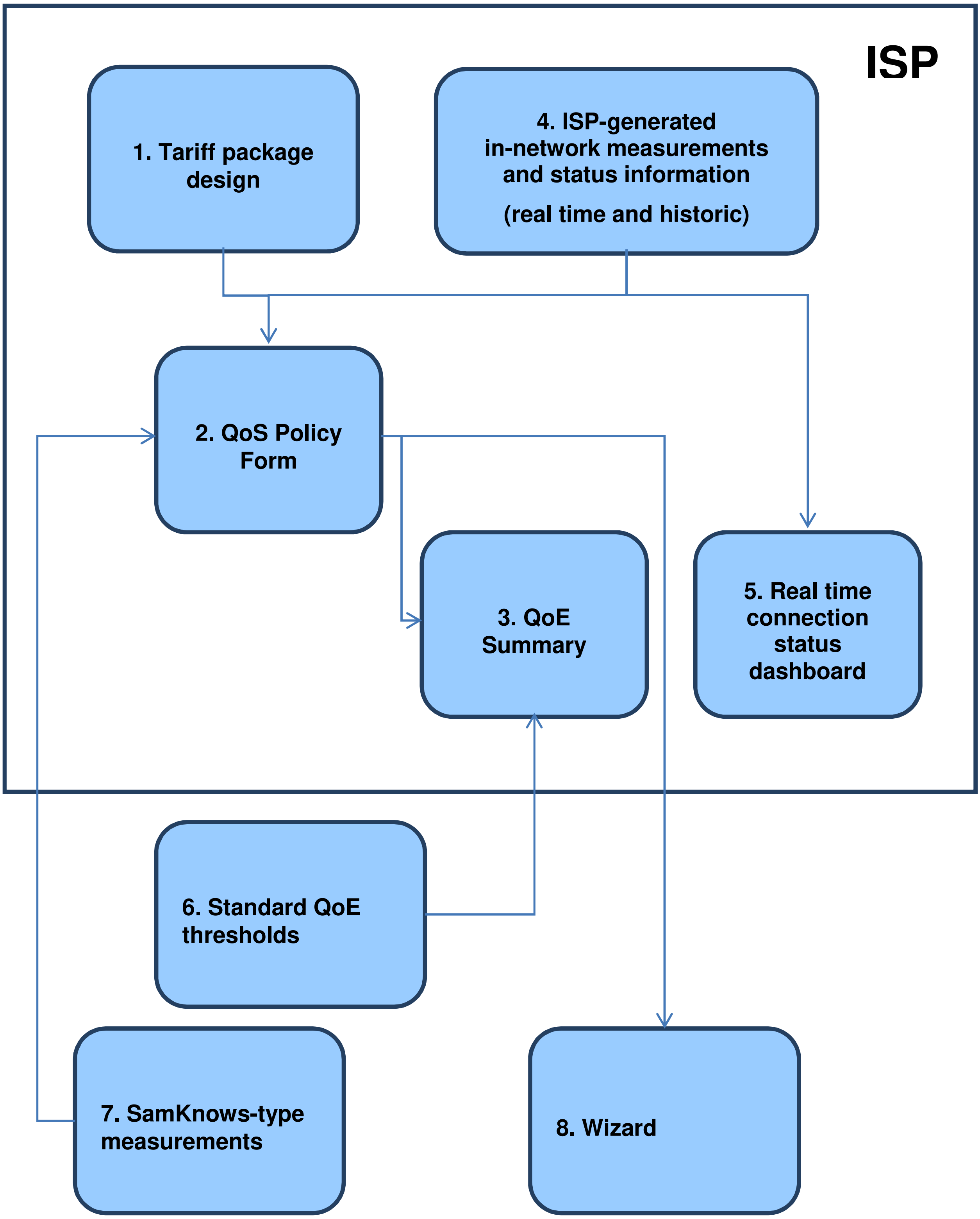


Figure 1: Overall recommendations

1 Introduction

Traffic management has recently risen in importance. There are emerging ‘battle lines’ between those who oppose traffic management and believe that the internet should exhibit strict neutrality, and those who believe that the mix of traffic on the internet has made traffic management a commercial and technical necessity. This study was commissioned by Ofcom to provide detail on the technical aspects of traffic management, to explore the effects of traffic management on consumers’ quality of experience, and to examine ways of measuring and characterising traffic management and connection performance. The study was initiated independently of the Broadband Stakeholder Group’s deliberations on a voluntary code of practice on traffic management transparency[4].

The study has involved desk research and interviews with a representative selection of ISPs, technology providers and content providers. All interviews were conducted on a non-attributable basis. Where reference is made in this report to individual companies, the material has been drawn from sources already in the public domain.

Some respondents asked that their participation should not be disclosed, so we have decided not to list the companies we spoke to. However we are grateful for the assistance provided by all the companies that we contacted during the preparation of this report.

[4] http://www.broadbanduk.org/content/view/479/7/

2 Definitions

There are many definitions of traffic management in use. From a technical point of view there are many aspects of network design and operation that affect QoS, and even affect the QoS of one traffic type compared to another. We think the key to traffic management as envisaged in our brief (Appendix A) is that the discrimination between types of traffic needs to arise out of a purpose. Accordingly, for this project we have defined traffic management as "**purposeful discrimination in access to network resources on the basis of traffic type, origin, or destination"**.

The internet is a packet switched network in which packets are normally transmitted on a best efforts basis. When there is congestion (more traffic than can be handled), data may be delayed or lost. Without traffic management, different data packets are treated more or less equally. Traffic management is a collection of technologies and policies which lead to different types of traffic being treated differently. Under congested conditions traffic management would cause some data to have a greater chance of being delivered than others. Such discrimination between data types would probably affect users' experience; in the extreme some applications would not be able to function. Of course, congestion could also cause applications to fail, but the distinguishing feature of traffic management is that it involves *purposeful* discrimination.

Considerations of whether certain practices are – or are not – traffic management has led us to refine our definition by recognising several sub-divisions within traffic management. The terms describing the sub-divisions are not necessarily in wide use but do, we believe, help delineate the field.

Some traffic management is 'explicit'. **Explicit traffic management** within the open internet involves identifying the class of traffic involved and then allocating data bandwidth or packet priority on a discriminatory basis.

There can also be 'implicit' traffic management. **Implicit traffic management** within the open internet is said to occur when the design and provisioning of a network has the effect of discriminating between traffic classes. For example the capacity of the pipes between content providers and an internet gateway affects the likelihood of it becoming congested. The word 'purposeful' in our definition seeks to exclude situations where discrimination has arisen without a deliberate intention to favour one sort of traffic over another. One form of implicit traffic management is the use of content distribution networks (CDNs). CDNs have a network infrastructure and provide local caching and/or connections to major points of presence. Both these techniques result in improved user experiences with respect to the content carried over the CDN.

The definitions given above focus on the open internet. Recent developments in IPTV and other managed services has resulted in some of the physical infrastructure (mainly the access network) being shared between the internet and managed services. For example, an access link could be physically shared between internet traffic and a paid-for IPTV service such as BT Vision, Virgin Media Player, Sky Anytime Plus and the forthcoming YouView service. These arrangements are not generally regarded as a form of traffic management but we consider that they should be. Accordingly we have also considered what we call '**partitioning**' of bandwidth between the internet and other services (such as IPTV) which may use some of the same physical infrastructure.

There are two sorts of partitioning – ‘static partitioning’ and ‘dynamic partitioning’. **Static partitioning** is where links are shared on a relatively fixed basis. **Dynamic partitioning** is where the sharing is variable, such as when a user pays to watch a movie using a managed IPTV service. Such an arrangement could lead to bandwidth being taken out of a broadband link for the duration of the movie.

3 Technology overview of traffic management

3.1 Introduction

This chapter provides a technical overview of traffic management and its application to the UK internet.

3.2 The principles of traffic management

One challenge in any study of traffic management is that it can be described in many different ways. For example, traffic management can be viewed in terms of:

1. its application across different network types (e.g. fixed DSL, cable, mobile);
2. where it is controlled and enacted within the layers of the ISO 7-layer model of communications;
3. where it is controlled and enacted in the physical, geographic network (e.g. core network, access network);
4. the impact it has on different traffic types (P2P, web browsing, streaming video, etc.);
5. the impact it has on different users, or classes of user;
6. the type of traffic management intervention that is used, and the decision bases for enacting the intervention.

In order to manage the complexity of this topic, we have selected a number of ‘views’, which provide a good description of the types of traffic management in use and the effects these have on QoE for consumers. These views will form the foundation for our QoE work and will help Ofcom to understand the issues and determine policy.

Within the main body of the report we have provided two ‘views’:

- an Intervention View (point 6. above, at section 3.3);
- a Physical Network View (point 3. above, at section 3.4).

We believe that these two views will provide the understanding that this study needs. However, we have also included in the appendix two additional views which may be helpful to some readers:

- an ISO Model View (point 2. above, at Appendix D);
- a Network Type View (point 1. above, at Appendix E).

We believe that these four views together provide the detail required to understand how traffic management in the UK works and to inform the process of setting policy.

Sections 3.3 and 3.4 below describe the two views that underpin the remainder of the project work. Subsequent sections within this chapter look at the way in which traffic management develops as networks expand and mature (section 3.5) and future developments in traffic management (section 3.6).

The chapter concludes with a summary of the differences between traffic management application in fixed and mobile networks (section 3.7), and some observations on the status of traffic management technology today and in the future (section 3.8).

3.3 An Intervention View of Traffic Management

While the implementation of traffic management is far from trivial, there are relatively few underlying techniques available. All traffic management involves a *decision basis* and an *intervention*. For example, exceeding a monthly usage allowance is a decision basis, and the response of cutting data rate according to policy is an intervention.

ISPs use many criteria to decide on what traffic management to apply. However, our study has shown that there are three main inputs used by UK ISPs in reaching decisions on what traffic management to apply, though they can be used in combination, and more complex rules can be set. The three decision inputs are:

- the *user identity* (or profile), specifying a QoS package for that user;
- whether or not a *usage cap* has been exceeded (note that these caps are often set by the user's tariff);
- the particular *traffic type*.

With regard to interventions there are two main types.

- *Packet prioritisation*. Wherever queues occur in a network, higher priority traffic will get through whereas lower priority traffic may be delayed or suffer packet loss. This is typically applied today in the core network, but may in future migrate closer to the access network to increase the effectiveness of traffic management in maximising network utilisation but minimising the effect on most users.
- *Bandwidth allocation*. The bandwidth (or data rate) offered to a user or a type of traffic can be actively controlled. Users can be offered a minimum guaranteed rate or can be limited or capped at a maximum rate. In most cases this is applied at levels 2 and 3 in the scheduler or in the access network which is where most networks have the greatest constraints on bandwidth.

Figure 2 below shows the 6 traffic management interventions that exist in the 2 by 3 matrix which describes all combinations of intervention type and decision input. For ease of reference we have labelled the cells by their row and column number.

Decision Input / Intervention type	1. User identity	2. Usage cap	3. Traffic type
A. Packet prioritisation	A1	A2	A3
B. Bandwidth allocation	B1	B2	B3

Figure 2: Matrix of traffic management approaches

We have reviewed all the traffic management interventions identified in our interviews and have confirmed that they do indeed fall into one or more of the cells in this matrix. Cell A2 is greyed out because we have found no evidence of ISPs currently using this type of intervention (implementing a usage cap via the packet prioritisation route). We found at least one example of an ISP using each other of the interventions identified in the matrix.

The location in the network of the decision and the intervention is not necessarily identical. For example, some traffic management decisions are made at management centres in the core network, but are implemented in the access network.

Similarly, traffic management decisions and implementations often span different layers in the ISO model. Many decisions get made at ISO Layers 3 (Network) or 4 (Transport) even if they are subsequently enacted at ISO Layer 2 (Data Link).

3.3.1 Intervention types

The two types of intervention have different characteristics and it is helpful to understand these when predicting the effects of traffic management on overall network traffic and on consumers' QoE. The main characteristics are listed in Table 1.

.

	Characteristic					
	Possible actions	**Currently applied in**	**ISO Model Level**	**Impact of negative intervention on data type**		**Comments**
Intervention type				**TCP/IP FTP**	**UDP RTP**	
A. Packet prioritisation	Prioritise or De-prioritise	Core network	Layers 3 and 4	Retransmission of packets	Data loss	TCP/IP traffic can be effectively managed by de-prioritising this traffic type
B. Bandwidth allocation	Guarantee or Cap	Access network	Layers 2 and 3	Reduced throughput (Service maintained, but at lower speed)	Reduced quality (Codec may drop to a lower rate)	Video is best managed by prioritising or giving guaranteed bandwidth in the access network

Table 1: Characteristics of intervention types

For each intervention type there is a positive action, which is generally beneficial for the traffic concerned, and a negative action, which may limit or constrain it. The positive approaches are sometimes used by ISPs to support 'up-selling' of consumers onto better packages (that offer greater throughput, or higher usage caps), or as the basis for managed services (which may guarantee performance for a particular traffic type or piece of content).

We explore the two types of intervention below.

Packet prioritisation (ISO layers 3 and 4)

Packet prioritisation makes use of the provision in packet headers of a field to indicate priority level or mark a packet as being of a particular traffic type. Protocols allow for end to end support for priority levels but this is not generally implemented across network boundaries. In practice, ISPs disregard the priority indicated on packets as they enter their networks and reset priorities to match their own traffic management policies.

Networks are made up of switching nodes, routing nodes and transmission links. Nodes switch/route packets from inputs to selected outputs. There are many complicated queuing algorithms used by ISPs to optimise traffic flow, but detailed knowledge of these is not needed in order to appreciate how ISPs manage traffic. The principle is that these nodes operate queues according to packet priority.

Within the ISO model it is important to appreciate that the Layer 3 (Network) does not guarantee the delivery of IP packets. Thus nodes are permitted to delay packets and drop packets that have queued too long. Low priority packets will tend to be delayed and/or dropped. Algorithms implemented at Layer 4 and above will determine whether packets have been lost and arrange re-transmission where necessary.

Different data types fare differently under packet de-prioritisation. In the case of TCP/IP traffic (e.g. P2P file sharing) on an IP network, flow control is implemented via a buffer at the receive end. If TCP/IP data is de-prioritised it suffers increased latency or data loss. High latency causes the flow control mechanisms at either end to reduce data rate. When the buffer is approaching capacity the receiver will tell the transmitter to stop sending data. Transmission errors are overcome by the receiver requesting re-transmission of any lost or corrupted packets. This works well when there is adequate network capacity and not too many packets are lost or corrupted. However, on a congested or poor-quality network, TCP/IP becomes increasingly inefficient and, in extreme conditions, contributes to further congestion through frequent attempted re-tries to send data.

UDP traffic (e.g. video streaming) would also suffer increased latency and data loss. Unlike TCP/IP packets, delayed or lost UDP packets are not re-transmitted. So, unlike TCP/IP, de-prioritisation of UDP packets will not lead to increased network load through attempted re-transmissions. However, depending on the application, users are likely to notice the effects of delayed or incomplete data, so packet prioritisation is not the best way to manage UDP traffic.

Bandwidth allocation (ISO layers 2 and 3)

Bandwidth allocation does not cause packet loss unless it reduces data rate to below that required for a particular application. Adequate data rate is particularly important for codecs, which necessarily operate in real time. Bandwidths can be set either for a connection as a whole or for individual traffic types separately.

The basic methodology for a telecommunication system works on the principle of taking a data stream from the upper layers of the ISO model and transmitting this over a physical interface using a modulated signal. The signals are modulated onto a physical layer medium transmitting and receiving the information sent in the frequency and time domain.

Networks are managed from two fundamental planes - data and control planes. The data contains the information being transmitted / received and some instructions on how this will be dealt with in the network. The control plane dictates how the network is managed including priorities for the information transmitted and potential bandwidth allocated in the pipes carrying the information.

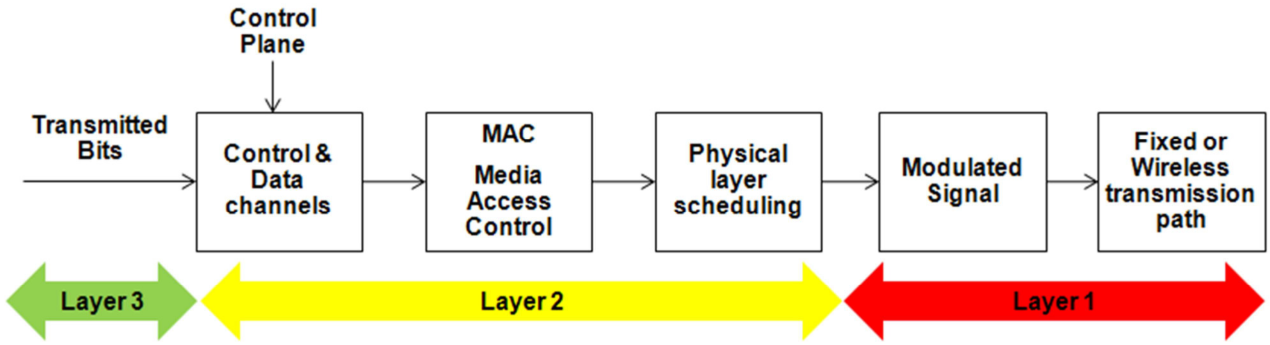


Figure 3 – Basic transmission path, functional blocks

Figure 3 shows the basic functional blocks of how an access node deals with the transmitted bits (data) and uses the control plane signals to create channels. The Media Access Control makes decisions on how the data should be scheduled into transmission medium resource. The physical layer scheduler will schedule information according to service priority, characteristic and available bandwidth determined by the link quality of the physical path, normally determined by the signal to noise ratio (SNR). The signal to noise ratio will determine how fast the link can be operated and is achieved by selecting a higher order modulation scheme which effectively allows more bits to be packed into the same physical allocation.

If no bandwidth control is required then the default mode will be to transmit the information in the highest order modulation scheme, therefore the fastest available bandwidth. In the case where a user is being bandwidth restricted then Layer 2 can control the actual amount of data scheduled to be transmitted over a fixed time period.

The bandwidth available (physical resource) and modulation schemes vary by access technology. For example, 3G wireless systems operate on 5MHz carriers which are shared by all users with a particular carrier/sector. The overall carrier is split into sub blocks which can be allocated depending on the characteristics of the different traffic types.

Cable uses multiple 8MHz channels and DSL networks have a frequency spectrum of approximately 1.1MHz. 3G and cable systems work on the principle of a shared medium and therefore can be used to allocate different amounts of bandwidth to different users. DSL's medium is a physical copper or aluminium wire which is not shared but suffers from high interference due to the poor RF quality of the cabling.

Different methodologies can be used to differentiate traffic and manage bandwidth from Layer 3 to 1 to set up multiple transmission pipes to an individual user. The scheduler will make decisions based on the available or restricted bandwidth.

The transmitted bits, at the layer 3 level, may have already been shaped or policed to determine the priority of traffic entering into the lower ISO layer transmission which will subsequently control the bandwidth (upper limit cap and lower limit guaranteed speed).

Traffic policing results in limiting to a maximum rate, excess traffic is dropped (or remarked) which results in peak traffic being smoothed. By contrast, traffic shaping holds packets back in a queue and then schedules for later transmission. The result of traffic shaping is a smoothed packet output rate.

The UMTS standard for example has four different classes, see Table 2 below.

	Conversational class	**Streaming class**	**Interactive class**	**Background class**
Fundamental characteristics	**Real Time**	**Real Time**	**Best Effort**	**Best Effort**
	Low delay guaranteed bit rate	Guaranteed bit rate	No guaranteed bit rate	-Destination is not expecting the data within a certain time
	- Preserve time relation (variation) between information entities of the stream	- Preserve time relation (variation) between information entities of the stream	- Request response pattern	
	- Conversational pattern (stringent and low delay)		-Preserve payload content	-Preserve payload content
Example of the application	Voice	Streaming video	Web browsing	emails

Table 2 - UMTS QoS Classes, main parameters

DOCSIS and DSL networks can utilise the QoS control defined by IEEE 802.1P, known as class of service (CoS). This is implemented as a 3-bit field called the Priority Code Point (PCP) which specifies a priority value of between 0 and 7 inclusive that can be used by Layer 2 QoS processes to differentiate and schedule traffic.

Table 3 below shows the QoS levels and traffic characteristics.

Network priority	Traffic characteristics	3 Bit PCP
0 (lowest)	Background	1
1	Best Effort	0
2	Excellent Effort	2
3	Critical Applications	3
4	Video, < 100 ms latency	4
5	Voice, < 10 ms latency	5
6	Internetwork Control	6
7 (highest)	Network Control	7

Table 3 - IEEE802.1P CoS network priority classes

Decisions at Layer 2 implementation can be determined by the Network Policy & Control strategy and the information can be programmed into nodes, transmitted over IP or sent separately over a control plane dynamically to change implementation. The IP (layer 3) may have been marked by a

DPI node or by the ISP traffic shaping/policing function for the layer 2 MAC to act upon accordingly.

3.3.2 Decision inputs

In principle traffic management decisions can take account of many different factors. In practice we found three main types of decision basis, though they can be used together. These are:

- the user's identity, providing reference to their tariff and account details. These might indicate such factors as the allowable bandwidth or how other services, such as managed IPTV, will co-exist with general internet activity
- usage caps – implemented in general by comparing the value of a counter with a pre-defined limit
- the type of traffic, identified typically using packet inspection technologies.

The characteristics of these different decision bases are shown in Table 4.

Decision input	Characteristics		
	Input measurement or context variable	**Types identified**	**Example actions**
1. User identity	User account	Consumer / business Tiers of tariff	Allocate bandwidth Prohibit or allow traffic
2. Usage cap	Packet counter	Download or upload amount per period	Warn user of approach to cap Limit bandwidth if cap exceeded
3. Traffic type	Packet inspection	P2P, VoIP, gaming Audio / video streaming	Prioritise or de-prioritise Limit or guarantee bandwidth

Table 4: Characteristics of decision inputs

The most controversial and potentially complex form of decision input is the traffic type. The majority of traffic type identification is initiated by inspecting packet headers and marking[5] them accordingly for transmission across the network. The inspection equipment will investigate the header information being transmitted across layer 3 and, based on criteria set in the Policy and Control node, will implement IP header manipulation. The complexity of this is determined by the ISP policy on traffic management. This ranges from simple blanket prioritisation, such as marking all P2P, through to complex bandwidth allocation by user and/or service intervening by changing packet headers and controlling pipe speeds.

The ISPs told us that they use established techniques to identify different types of traffic on their networks. Typically this is achieved by one or more of the following methods:

- association – by noting sending or receiving IP addresses, physical device types or the ports on which traffic is presented;

[5] Traffic derived from within the network could already be marked by the source according to the traffic management policy and therefore DPI techniques are not required. Traffic shaping and policing can therefore be applied based on the known marking.

- shallow inspection – looking at headers to identify data protocols;
- deep inspection – looking inside packets at the data payload;
- heuristic – looking at the pattern of traffic to determine its type.

In some cases users or content sources attempt to improve the throughput of their data by changing packet headers to disguise their data as another type, which they believe is subject to less management. ISPs told us that they were generally able to identify traffic types using other techniques and were quickly able to minimise the impact of such actions on their networks.

The equipment that identifies traffic is often referred to as a “DPI Box”, regardless of whether traffic is identified actually using deep packet inspection, by shallow inspection or using heuristic methods. ISPs told us that they currently have no need to look in detail at the data payload in order to make commercially- or technically-motivated traffic management decisions.

3.3.3 Interventions used by ISPs

We have analysed the interview responses and published policies from ISPs in order to infer the interventions in use, and we have matched these against the six traffic management interventions defined in Figure 1. One of these interventions (A2) was never used, leaving five remaining.

We found that the DSL ISPs are collectively[6] using all five of the interventions.

We found that mobile operators were using all the interventions except B3 (bandwidth allocation by traffic type). We believe that mobile operators currently have no need to intervene in this way since the bandwidth available in their radio access networks already provides a physical limit which does not need to be augmented. They maintain active control through application of the other interventions. A1 (packet prioritisation by user identity) is the mechanism used, we believe, to block VoIP in some mobile tariffs.

We found that cable providers were using all the interventions except A1 (packet prioritisation by user identity) and B3 (bandwidth allocation by traffic type). We believe that cable operators currently have no need for intervention A1 as they have sufficient bandwidth headroom in their access network that they don’t need to prioritise packets by user; prioritisation by traffic type (A3) provides the control they need. Their access network bandwidth headroom also means they don’t currently have a need to allocate bandwidth by traffic type, meaning that B3 is not required at this time.

We did not find that any of the five interventions in use were exclusive to a particular network type, or that they operated fundamentally differently when used by operators over mobile, DSL or cable networks. The absence of conflicting interventions and characteristics across the network types means that it would be possible to design a single, flexible, traffic management regulatory policy which is applicable to all types of network, and all ‘flavours’ of ISP. Provided that ISPs are allowed to select which of the five interventions they use, they will have the tools they need to manage traffic on their networks and to give their users an appropriate QoE

3.4 A Physical Network View of Traffic Management

To explain the concepts detailed in the previous sections the diagram below, Figure 4, shows how a system using packet inspection can inspect the IP traffic and mark packets accordingly for the network to deal with, driven by the Policy & Control strategy.

[6] This is not to imply that *any individual ADSL ISP* uses all five

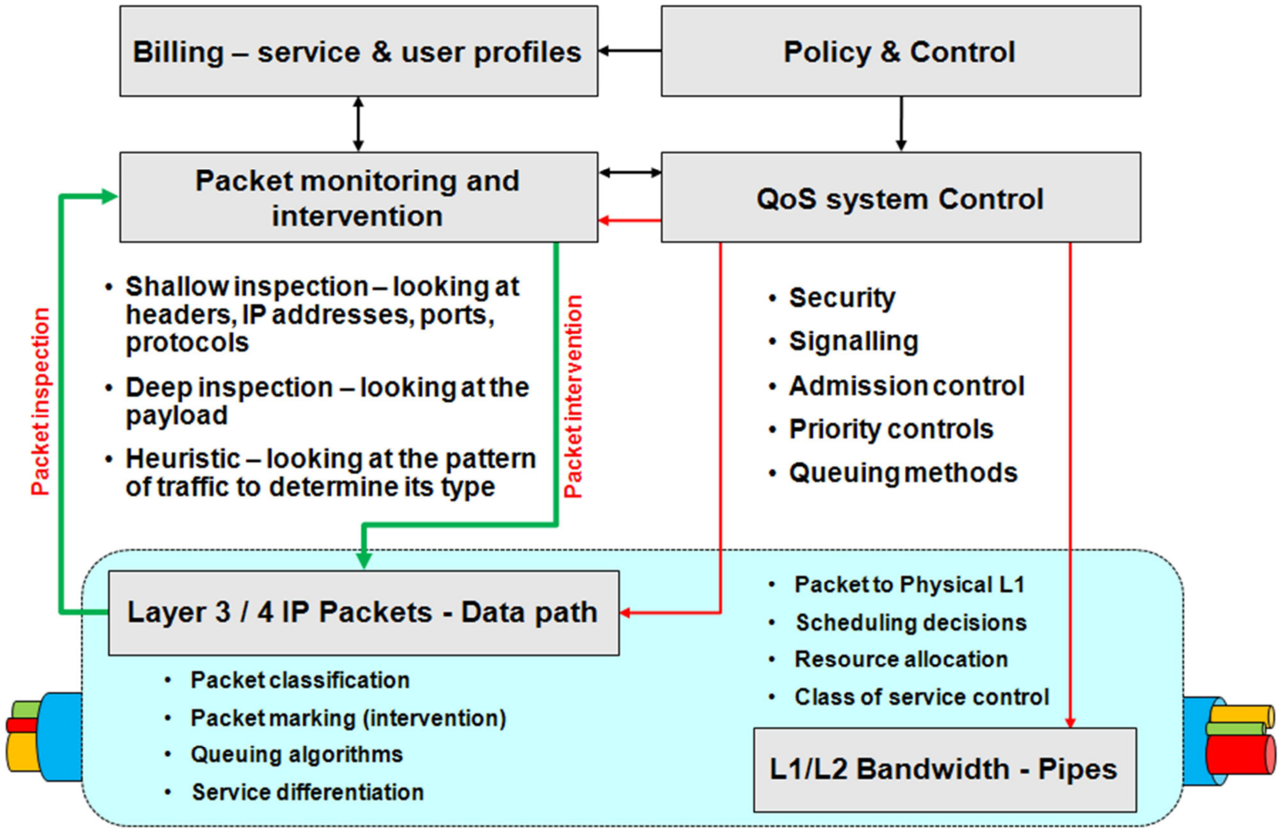


Figure 4 - Generic traffic management architecture

The majority of IP intervention is dealt with by inspecting packet headers and marking them accordingly for transmission across the network. The inspection equipment will investigate the header information being transmitted across Layer 3 and, based on the criteria set in the policy and control unit, will implement IP header manipulation. The sequence of events is denoted by the green lines on the diagram. The complexity of this is determined by the ISP policy on traffic management, from simple blanket prioritisation such as marking all P2P through to complex bandwidth by user and/or service intervention by changing packet headers and controlling pipe speeds. The control is implemented over the control plane of the network and is denoted by the red lines on the diagram.

To further explain this methodology for traffic management Figure 5 below shows a generic network architecture with the functions associated with particular nodes.

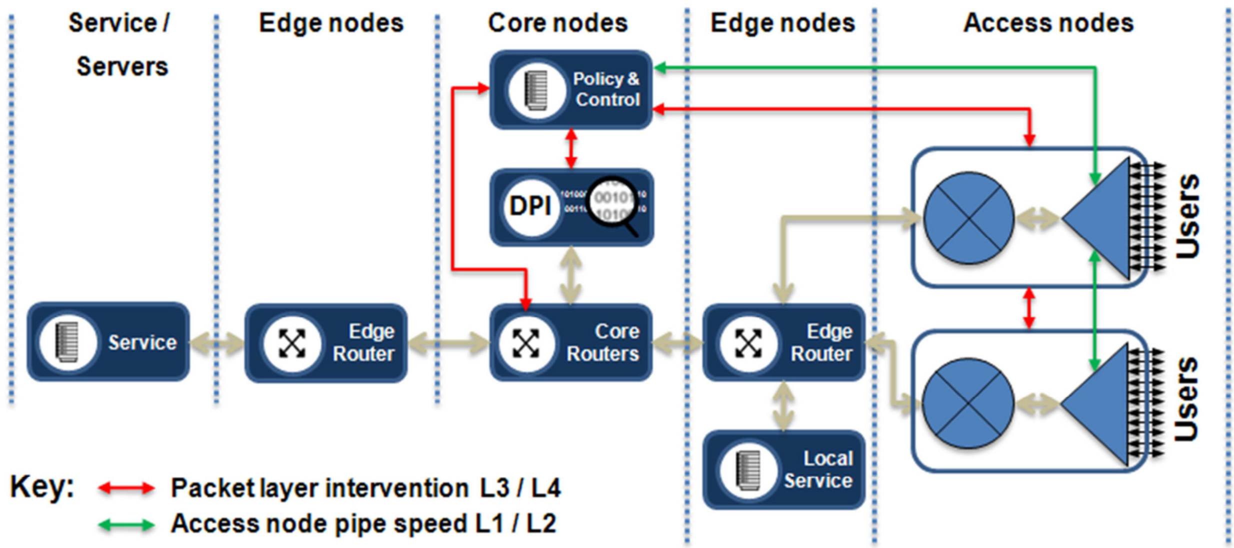


Figure 5 - Network agnostic traffic management example by node

The key aspects of Figure 5 are as follows:

- DPI boxes are deployed in the core network nodes to inspect the packets to determine traffic types. Information is passed to the policy and control node.
- Policy and control units typically contain the traffic management policies and, based on the information received from the DPI box, send control signals to the respective nodes on how to deal with the traffic.
- The red lines indicate the packet based intervention where the core nodes re-label packets based on the priority decided by the traffic management policy, and the access nodes treat them accordingly. As most networks today have the ability to inspect packet headers then all packet header fields in theory could be manipulated.
- The green line indicates control of the access node Layer 1 and Layer 2. For example the DPI box could monitor a monthly usage cap and when the limit is reached could apply a reduction to the pipe speed by allocating less resource to the end user in the access node.

Traffic management architectures differ by access technology but the fundamental principles remain valid for all network types. Appendix E includes some generic network architectures by access types: DSL, Cable and 3G.

3.5 Traffic management versus capacity expansion

From the analysis conducted it appears that traffic management extent and complexity is associated with how much congestion is being experienced or how close the network traffic is to the limit of the network capacity (see Figure 6)

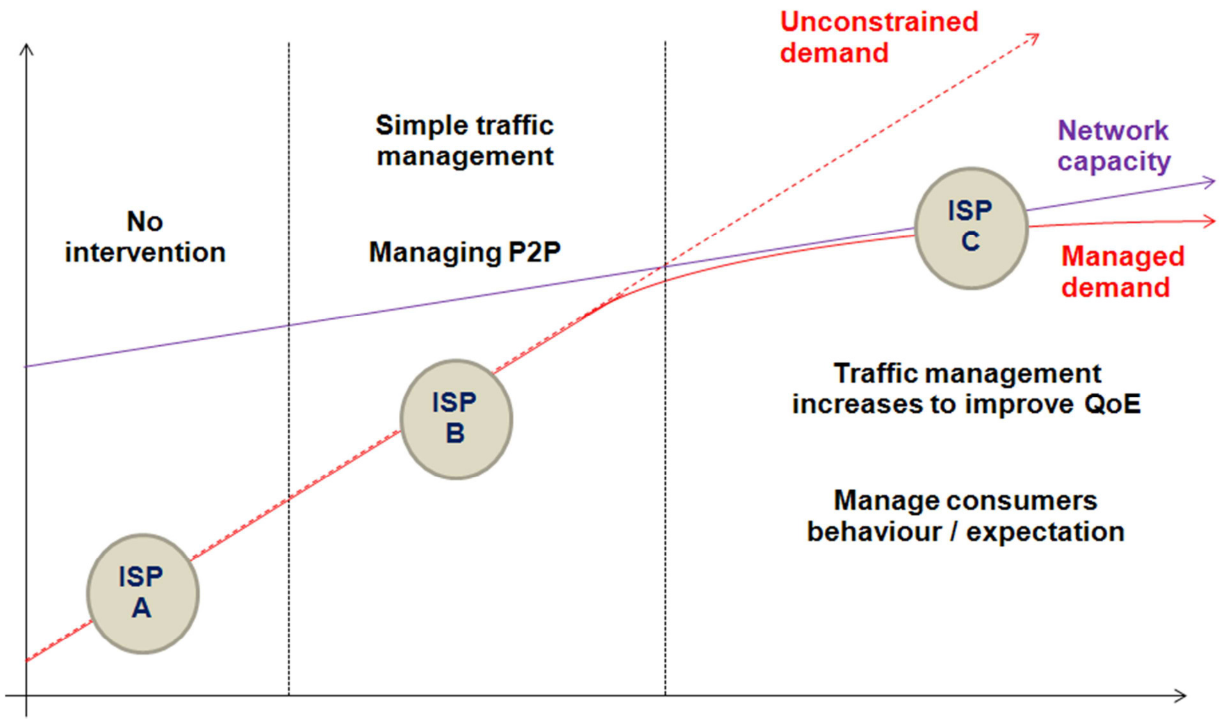


Figure 6: Traffic management as a response to capacity limitations

Traffic management in general is a response to an emerging need to manage the network resource as it approaches capacity limits. This results in congestion which in turn impacts the Quality of Experience (QoE) for the end users. Figure 6 shows three hypothetical ISPs. The x axis represents the stage of development of the ISP. The y axis represents data demand and available network capacity available to that ISP.

- ISP A has plenty of capacity and can deliver all services at a high QoE as no nodes are congested. Even heavy users do not impact the network.
- ISP B has less headroom and a minority of heavy users are causing congestion. In this case the ISP identifies that P2P users are the root cause and implements a policy where P2P traffic is managed therefore bringing the congestion under control and maintaining a high level of QoE.
- ISP C is reaching its capacity limit, and multiple traffic types are causing congestion. In this case the ISP starts to differentiate between multiple traffic types and/or users. The intervention is based on QoE characteristics of the service. For example, video is prioritised over a non real time application. The complexity of intervention has increased with multiple dimensions and service packages are differentiated to manage the end user expectation and behaviour.

Increasing capacity is often preferred to managing traffic. In practice this may not be possible due to the economic implications or because the fundamental technology of the network is at its limit, in terms of capacity or pipe speeds. Traffic management may be deployed as a ‘stop gap’ before new capacity comes on line. Traffic management can allow ISPs to delay the point at which new capacity is installed. This improves asset utilisation whilst maintaining an acceptable QoS for users until the business case for new capacity is strong enough to justify investment.

Figure 6 can also be treated as a single ISP (hypothetical) evolution over time. The ISP starts at position A, where there is plenty of capacity. It ends at position C, where congestion has to be managed at a finer granularity to retain the QoE and stop potential churn.

3.6 Where is traffic management heading over the next five years

From a purely technical point of view, traffic management will remain as a response to congestion and a mechanism to maintain the highest level of QoE. The most important changes will be that packet inspection capabilities will migrate outwards from the core, enabling a more finely graded and user-specific form of traffic management. Figure 7 below shows a potential scenario based on a hypothetical single network architecture evolution over time.

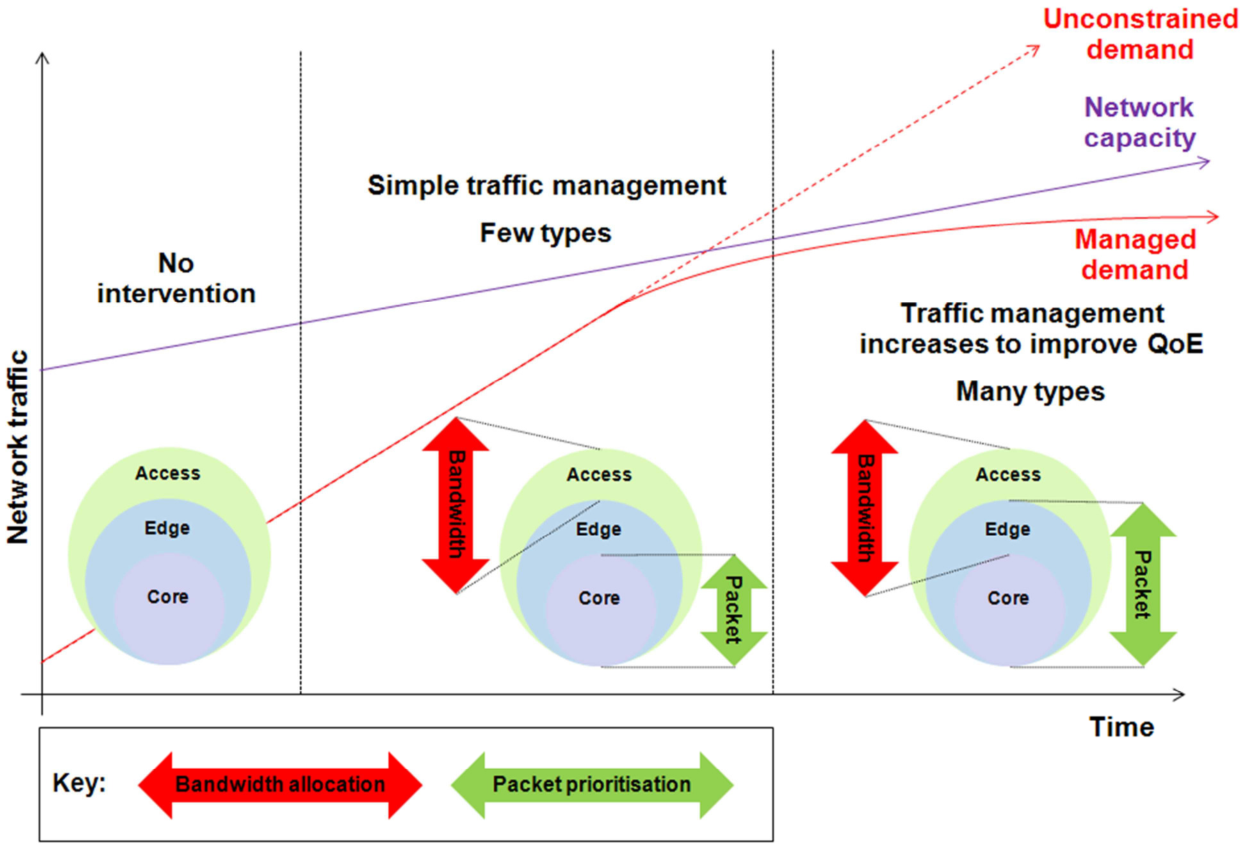


Figure 7 - Traffic Management evolution in relation to network technology

- **No intervention** – Network has plenty of capacity and can deliver all services at a high QoE as no nodes are congested. Even heavy users do not impact the network.
- **Traffic management few types** – Network has less headroom. A small number of users and/or traffic types are the root cause and the operator implements a policy where traffic is managed, therefore bringing the congestion under control and maintaining a high level of QoE. Basic intervention occurs through packet prioritisation in the core and managing bandwidth at the access through traffic shaping and policing intervention.
- **Traffic management many types** – Network is reaching its fundamental capacity limit and multiple traffic types are causing congestion. In this case the network technology starts to differentiate multiple traffic types and/or users. The intervention is based on QoE characteristics of the service. The intervention now has multiple dimensions and service packages are differentiated to manage the end user expectation and behaviour. DPI functionality and packet prioritisation is becoming more complex and migrating from the core to the edge in order to

provide maximum control. Bandwidth control due to fundamental limitations is being applied at the access and edge

3.7 Differences between fixed and mobile networks

We have included in Appendix E a set of generic network architecture diagrams and brief descriptions for each of the three main types of network - DSL, cable and mobile. We have annotated these with information about how traffic management is currently carried out on these networks.

Our analysis indicates that there are some detailed differences in the location, control and enactment of traffic management between the network types. However, these differences primarily reflect the variations in network architecture, rather than any fundamental difference in approach to traffic management.

Each network type features a centralised Policy and Control function, which stores the network management policy and converts this into instructions for specific management actions. Each network type features DPI boxes[7] in the core network which mark traffic types and enable packet prioritisation to be enacted in the core and access networks as determined by the policy. Bandwidth allocation is mostly implemented in the access network, where bandwidth is more scarce and there is greater potential for the actions of individual consumers to impact the QoE of others.

ISPs told us that there is a general trend to move packet inspection and prioritisation towards the edge of networks, where it can provide finer levels of control and thus an improved QoE for consumers. However, the cost of doing this is high, so it will only happen if it can be shown to bring significant benefits. It may be more cost-effective in many cases for ISPs to address congestion by installing additional network capacity rather than by increasing the number of packet inspection points in the network.

The main difference in implementation occurs in the case of mobile networks, which uniquely have a free space radio link as their access medium. Mobile operators also have to manage the challenge of users who move around both within and between cells, and of large changes in local demand, for example during major sporting or entertainment events.

Since radio spectrum is a limited and costly resource, mobile operators always work to make best use of it. This means that the radio access network is likely to remain as the limiting point in terms of traffic capacity on most mobile networks for the foreseeable future. Its limited capacity also acts as a natural traffic management function, which means that mobile operators generally don't have to apply bandwidth allocation for different data services in the access network. They do however apply rules that allocate radio network resources between voice and data traffic, to ensure that each gets a fair share. This isn't an issue for DSL or cable, as in these cases voice traffic is carried on a reserved part of the available spectrum (DSL) or on a separate copper pair (cable). So coexistence with voice is far more of an issue in radio networks, and voice still tends to have priority.

On the basis of our knowledge of the different network types, and the information gained from the interviews, we do not see the need for any fundamentally different regulatory policies for traffic management being required for ISPs using DSL, cable or mobile networks. There will continue to be traffic management policy differences between operators, but these will reflect the demand for their services, the development status of their networks and differences in the capacity of their access networks, rather than any fundamental difference in approach driven by network type.

[7] Though not necessarily using deep packet inspection

3.8 Traffic management today and in the future

ISPs told us that traffic management technology, as applicable to IP networks, is already relatively mature. All ISPs have access to packet inspection techniques that allow them to implement their desired packet prioritisation policies. Collectively they have the ability to measure traffic by user, by traffic type and apply prioritisation by user, by traffic type and by time of day. They can apply usage caps and can control the bandwidth consumed by individual users or types of traffic. They are able to follow and react to trends in the market where some users attempt to disguise a traffic type as another to reduce the impact of traffic management. They can identify and control problem users and prevent them from degrading the service of other customers.

Real-time packet inspection is currently an expensive activity. Like most electronic goods, DPI products will increase in performance and reduce in cost over time. Increased affordability will present an opportunity for operators to move inspection out from the core towards the network edge. The result is likely to be more user-specific traffic management policies in the future, giving operators additional controllability of high-use consumers and traffic types.

Operators will be in a position to offer differentiated packages, for example featuring a guaranteed performance for certain types of traffic (e.g. games or streamed video). Current traffic management technology has the ability to support the delivery of such offerings in principle but the cost of deployment is the issue.

ISPs do not expect any major dislocations or significant changes in the way they manage traffic over the next five years. They see a steadily increasing demand for bandwidth being matched by investment in new capacity, which they may choose to fund partly through differentiated service offerings. This will continue to be supplemented by traffic management to optimise network utilisation, manage peak loadings and ensure that the actions of a few heavy users don't impact the QoE of the majority.

The increasing popularity of streaming video will continue to cause stress for most operators. It demands relatively high bandwidth, tends to need low jitter, and consumers expect it to be delivered at ever-increasing quality and to operate without interruption for several hours (e.g. when viewing a movie). Unlike P2P (which can be de-prioritised or bandwidth throttled) streamed video cannot easily be managed, other than by a co-operative process with content providers to employ more efficient codecs, or to buffer or cache content. ISPs can be expected to develop their traffic management policies to cater for the growing demand for video, whilst maintaining an acceptable QoE for other users.

In mobile, the radio access network (RAN) can additionally be managed through the allocation of codes (effectively, the relative allocation of spectrum).

4 ISPs policies and the current use of traffic management

4.1 ISP policies

The ISPs we interviewed all claim that they use traffic management strictly according to published policies. These policies are either constituted as separate documents on their websites or are effectively incorporated into the details of tariffs.

4.1.1 Fixed ISPs

In Table 5 we have summarised the published policies of a selection of the main[8] fixed ISPs. We have included the top six UK ISPs together with Plusnet which is widely cited for its use of traffic management. The entries in the table show that a policy applies to at least one of an ISP's tariffs but not necessarily all of them.

	Volume limits	P2P policy	Video streaming policy	Comments
BT fixed	Yes	Yes	No	No other traffic management specified on website
TalkTalk fixed	Yes	Yes	No	No other traffic management specified on website
Virgin fixed	Yes	Yes	No	No other traffic management specified on website
Sky fixed	Yes	Yes	No	Heavy users monitored and restricted. Sky "may slow down the speed that all Sky Broadband Connect customers can get on applications such as peer-to-peer networking and newsgroups, which we consider use up a lot of bandwidth and have a negative effect on other customers.[9]
Orange fixed	Yes	Yes	No	"Traffic management is where we sometimes apply restrictions to the amount of network capacity a customer can use, which can affect your throughput speed. We do this to stop a small number of customers who excessively download during peak times (6pm to midnight), as this affects the quality of service we provide to all other customers. It also means that we are able to prioritise certain types of internet traffic on time-sensitive applications, such as our second line phone service or gaming." [10]
O2 fixed	Yes	Yes	Yes	P2P and streaming video allocated bandwidths according to tariff. No other traffic management specified on website
Plusnet fixed	Yes	Yes	Yes	11 traffic types independently treated through both a combination of prioritisation and rate limiting

Table 5: Summary of fixed ISP traffic management policies

---

[8] According to http://www.ispreview.co.uk/review/top10.php
[9] http://www.sky.com/helpcentre/broadband/set-up/sky-broadband-product-information/
[10] http://shop.orange.co.uk/broadband/broadband-explained#traffic-management

Explicit traffic management

The majority of fixed ISPs employ relatively simple policies which seek to mitigate the impact of 'heavy' or 'problem' users. The policies are based on a volume limit and some restriction of peer-to-peer (P2P). Most ISPs state that only a small number (typically 1% to 5%) of customers are 'caught' by this sort of traffic management. The rationale is always that the excessive use of resources by a minority is unfair on others – hence the policies often include the expression "fair use" in the title. In some cases there are 'application agnostic' usage limits set for each tariff. The lower priced tariffs have limits that would certainly be restrictive to some users, while the higher priced tariffs allow 'unlimited' use, subject to a fair use policy. In the case of Virgin, usage limits are tied to line speed.

There are two notable divergences from the norm among fixed ISPs.

- O_2 has bandwidth limits on both P2P and streaming video. These limits vary according to tariff. These 'limits' can also be read as guarantees.
- Plusnet, which was cited by several ISPs in their responses to Ofcom's consultation document on net neutrality, has a range of tariffs which differ according to the QoS offered on eleven traffic types. Two methods are used - traffic prioritisation and rate limiting. The details given for each tariff are the priority levels (described as platinum, titanium, gold & gold plated, silver, bronze, and best effort) and the rate limit, if applicable, in kbit/s at different times of the day.

The interviews revealed that there was some traffic management being applied which was not apparent from policies. For example, one ISP gives priority for gaming and VoIP traffic in order to improve QoS. Because of the traffic volumes involved, this would probably not have any detrimental effect on other users.

Implicit traffic management

No ISPs told us of any deals with content providers, content delivery networks (CDNs) or disclosed anything that would distort consumers' access to content. However BT wholesale has subsequently announced BT Content Connect which is a form of CDN. CDNs are not seen by ISPs as a form of traffic management. However they are promoted to content providers as a means to improve their connectivity with ISPs, thereby improving QoS for consumers. The fact that CDNs are not generally owned or controlled by ISPs explains their positioning within an ISP's frame of reference.

Partitioning

Managed IPTV services (e.g. BT Vision, Virgin Media Player, Sky Anytime Plus and the forthcoming YouView service) share the same physical access pipe as broadband internet traffic. Whether this affects usable broadband bandwidth depends on the service concerned. When such traffic is 'guaranteed', it limits the bandwidth available for 'best efforts' internet traffic within the finite limits of shared physical resources. Managed IPTV services are typically not included in ISPs' traffic management policies, and this sort of traffic often does not count towards volume quotas. In the case of ADSL we understand that ISPs will check a consumer's connection to ensure that there is sufficient residual capacity for internet traffic before allowing a managed service to be provisioned.

In interviews it was pointed out that IPTV may drive an incremental investment in bandwidth within ISP access networks, creating bandwidth which wouldn't be there if IPTV didn't exist. It was argued that the aggregate effect on other internet traffic would be minimal once this extra capacity is taken into account.

.

4.1.2 Mobile ISPs

In Table 6 we have summarised the published policies of a selection of the main mobile ISPs.

	Volume limits	P2P policy	Video streaming policy	Comments
T-Mobile	Yes	Yes	Yes	"If you exceed our fair use policy, you can still use the internet for the things you love most - like email, Facebook and news sites - and we won't charge you any extra. But we may restrict video streaming, peer-to-peer downloading and other things that affect other people's use of the internet at peak times." [11]
Orange mobile	Yes	See comment	See comment	"Orange may additionally manage customers' data connection at peak times to preserve the best experience for the greatest number of users" [12]
O_2 mobile	Yes	No	No	No other traffic management specified on website
Vodafone mobile	Yes	No	No	VoIP blocked under some tariffs
Virgin mobile	Yes	No	No	No other traffic management specified on website
Three	Yes	No	No	Three does not have a fair use policy but guards against excessive use through volume charging (except in the case of the One plan which has unrestricted access to the Internet)

Table 6: Summary of mobile ISP traffic management policies

Explicit traffic management

All mobile operators have volume charging and/or 'fair use' policies, though some have current or legacy tariffs which are headlined as 'unlimited'. Where low volume limits (around 1GB) are in use, there is little need to differentiate between different traffic types in order to limit heavy users. However some fair use policies do make a distinction between traffic types, and will allow browsing and emails, but not video streaming, once usage limits are exceeded.

Some ISPs block VoIP under some tariffs; this is presumably to account for the potential for VoIP to cannibalise telephony revenue.

Implicit traffic management

No mobile ISPs told us of any deals with content providers, content delivery networks (CDNs) or anything that would distort consumers' access to content.

[11] http://www.t-mobile.co.uk/shop/mobile-broadband/about-mobile-broadband/
[12] http://www.orange.co.uk/images/editorial/Orange-mbb-Animals-Terms-20101101b.pdf?linkfrom=%3C!--linkfromvariable--%3E&link=box_main_pos_1_1_link_1&article=termsofusemobilebroadbandcurrent

Partitioning

As the radio access layer carries voice in addition to broadband, the bandwidth available to broadband can be affected. None of the published policies give any insight into how bandwidth and priority is managed between voice and data.

4.2 Observations on the current situation

Currently, all fixed ISPs use some form of traffic management. Most use it in a minimalist way. They adopt a simple approach of having relatively high volume quotas and only actively restrict P2P. This is essentially an engineering response to a small minority of very heavy users. In principle there is a trade-off between expanding capacity and more extensive traffic management. We infer from ISPs' behaviour that, in practice, investing in capacity is considered preferable to investing in traffic management. Only where capacity expansion is particularly costly does traffic management feature more strongly in ISP strategies.

The mobile ISPs vary in their approaches. Some attempt to limit usage through volume quotas alone whereas others combine volume quotas with traffic type discrimination. While the original tariffs were often unlimited, reflecting the practice in fixed broadband, the rapid growth in penetration of smartphones mean that most ISPs now offer packages that differ in the volume of data allowed. T-Mobile appears to apply traffic discrimination only once the 'fair use limit' has been reached in order to continue to allow access to certain services even if the limit has been exceeded.

The technology exists to create finely differentiated services by applying different priorities and bandwidths to different types of traffic. In principle this allows QoE to be managed more directly, and tariff packages to be targeted to different user segments. Currently this approach is being used by Plusnet in the UK. In interviews with ISPs we explored whether similar approaches might be adopted more widely. The general view was that any market segmentation advantages would be offset by the increase in network complexity and cost, and the difficulty of communicating policies to consumers.

In interviews we did not find an appetite among ISPs for using substantially more traffic management.

While the net neutrality debate has sparked concerns about the growth of 'covert' traffic management, we found that the use of traffic management in the UK is reasonably overt. This is probably because traffic management is implemented in order to support a process of consumer behaviour change – e.g. reducing use of P2P, or encouraging consumers to trade up to a more expensive package.

None of the ISPs interviewed indicated that they treat traffic differently according to its source. Indeed, as yet, the content-supply side of the two-sided market is hardly developed. No deals between ISPs and content providers were disclosed, and while CDNs have been implemented, they are not being implemented by ISPs, and they are not seen by ISPs as traffic management *per se*[13].

[13] While BT Wholesale has launched Content Connect, a form of CDN, BT Wholesale is not strictly an ISP (ie it does not provide an internet service to end users).

4.3 Current approaches to transparency

All the ISPs we spoke to expressed a commitment to transparency. In practice, the policies can be difficult to locate[14], and some ISPs do not provide great detail in their policies. For example, while a policy may be clear that P2P is moderated or restricted during certain periods, the extent of the impact is not indicated.

This limitation in published information may be a result of the statistical nature of traffic, causing P2P to be restricted only as much as is necessary at any given time. We call such policies 'non-deterministic'. Whereas deterministic policies can be fully described in policies, non-deterministic ones cannot.

During the course of this project there has been some public discussion of T-Mobile's fair usage policies, prompted by the reduction in fair usage quotas. It appears from press comment that T-Mobile is relatively sophisticated in its use of traffic management to distinguish between, say, the content of an email message and an attached file. This level of sophistication is not described in the company's published policies.

[14] Our experience concurs with the mystery shopper exercise reported in the Ofcom discussion document http://stakeholders.ofcom.org.uk/binaries/consultations/net-neutrality/summary/netneutrality.pdf

5 Future scenarios for traffic management in the UK

Our interviews with ISPs suggested that they do not see a need for a radical change in their traffic management policies in the foreseeable future. The perceived problem of a small number of very heavy users consuming disproportionate capacity is largely being controlled by a combination of usage caps and restrictions on peer-to-peer traffic.

This situation may not persist, however. Video streaming is widely expected to grow substantially, and will need to be managed. Network capacity will continue to be added, and will have to be financed. The bandwidth for internet applications may be squeezed from both managed IPTV services and over the top services.

To focus attention on the ways traffic management may evolve we have developed five core scenarios. Apart from the first scenario, the other scenarios could, in fact, be combined.

1. Fair use Minimal, with limitations targeted at excessive users only
2. Traffic management evolves to facilitate congestion-sensitive traffic An engineering response to improve QoS for gaming, VoIP and live video streaming
3. Traffic management evolves in response to growth in video streaming Likely to involve a combination of restriction and monetisation of demand
4. Traffic management evolves as a business/marketing tool Promoting coordination across the two-sided market, using traffic management to define and implement new service packages.
5. Managed services become the norm Users' access connections are routinely partitioned to allow for managed services such as IPTV.

Table 7: Core traffic management scenarios

These five scenarios are discussed further below.

5.1 Scenario 1: Fair use

This scenario assumes that capacity expansion is considered in practice to be preferable to traffic management. This strategy is only likely to be feasible for those ISPs with a low marginal cost of capacity and/or the ability to finance capacity expansion through convincing plans to increase revenue.

5.2 Scenario 2: traffic management evolves to facilitate congestion-sensitive traffic

Current 'best efforts' delivery tends to produce poor QoS under congestion conditions. Some types of traffic – such as VoIP and gaming - are particularly sensitive to latency. Live streaming video is also sensitive to QoS. This scenario assumes that QoS-sensitive traffic (real-time) is identified and prioritised. Provided such traffic is not a large proportion of total traffic, it is assumed that there would be minimal offsetting impact on the non-prioritised traffic.

5.3 Scenario 3: traffic management evolves in response to growth in video streaming

Video streaming is widely regarded to be the next growth area for the internet. There will be wide fluctuations in demand and at peak periods video streaming will need to be limited.

Mobile ISPs may tackle this through volume quotas and/or specific limitations on video. On fixed line ISPs, simple restrictions similar to those applied to P2P are unlikely to be appropriate. This is for two reasons. Firstly, because video streaming is not associated with unlawfulness in the same way, it cannot be throttled back without an impact. Secondly, whereas P2P has been characterised by a small number of very heavy users, streamed video is characterised by a large number of moderate users. We expect ISPs to use a combination of restrictions and premium packages with guarantees. These are more likely to be in the form of bandwidth guarantees (eg an allowance of a certain number of HD streams) than a general de-prioritisation policy which would be unpredictable in its effects.

There is already dialogue between major content providers and leading ISPs regarding optimum coding rates for streamed video traffic. This coupled with further improvements in codecs (leading to better quality at lower bitrates) will allow existing networks to handle some increase in demand.

The architecture of different ISPs' networks affects what can – and cannot – easily be done. However we think that bandwidth guarantees for video will be easier to sell and communicate than differential priorities. The peaks in demand are not completely predictable, so policies may therefore not be fully deterministic.

5.4 Scenario 4: traffic management evolves as a business/marketing tool

This scenario puts traffic management in the hands of the marketeers, not the engineers. We envisage a combination of restrictions and paid-for services, monetising consumers' willingness to pay for content/service bundles.

Some hypothetical examples are:

- a content-led ISP offers a guaranteed QoS for its own content but only offers a 'best efforts' QoS for rival content;
- an ISP offers a tariff package that guarantees QoS within working hours (including 'professional P2P/file transfer applications') which might appeal to home workers;
- an ISP offers a priority service to a small number of very popular websites by caching these sites locally.

5.5 Scenario 5: Managed services become the norm

This scenario assumes that users' broadband connections are routinely partitioned to allow for managed services such as IPTV; this scenario features:

- a managed video conferencing service;
- pushed content – delivery outside peaks.

These five scenarios are used later in the report for assessing the types of consumer information and measurements required. Section 7.4 explores the information requirements by scenario.

6 The effects of traffic management on QoE

There are a number of almost philosophical issues that need to be addressed when looking at the effect of traffic management on quality of service and quality of experience.

Traffic management is not fully observable directly

From the forgoing technical discussion it is apparent that traffic management can act on traffic in different ways. These include:

- guaranteeing delivery of data or reserving bandwidth for that data;
- prioritising certain types of data in the event of queuing;
- de-prioritising certain types of data;
- restricting certain types of data or the bandwidth allocated;
- blocking certain types of data.

At an individual connection or device, a user cannot necessarily observe traffic management directly. He or she can observe the *performance* of an application and decide whether the performance is acceptable or not. If the application works as expected one can infer that the data have arrived in a timely manner. But it is impossible to tell whether the data have arrived only because they have been prioritised, or whether they have arrived because best efforts are perfectly adequate. Conversely, where the performance of an application suggests that data have not arrived in a timely manner it is impossible to observe directly whether this is the result of them having been deprioritised, or of congestion.

Despite the lack of certainty on the above points, a user may however make inferences based on the behaviour of applications compared to previous performance and possibly compared to other applications running at the same time. Some examples are set out below.

- A user may see a disparity between the apparent performance of different applications or websites, or between one user in the household sharing the same connection and another. Some applications are known to be more sensitive to QoS than others[15].

[15] In a previous study for Ofcom we were able to place applications into three categories according to how resilient they are to reductions in QoS:

- **QoS tolerant applications.** Downloading files from iTunes, the casual online multiplayer game, iPlayer Live, iPlayer SD and the on line interactive application continued to work well with packet loss of 1% and latency of 100ms. These represent the worst case conditions for a UK ISP except when there are network problems.
- **QoS sensitive applications.** Skype (VoIP) and YouTube worked well with packet loss of 0.25% and latency of 50ms. These represent average conditions for a UK ISP. Under worst case conditions, these applications continue to work but exhibited problems.
- **QoS critical applications.** The MMOG, iPlayer HD and the VPN started to exhibit reduced QoE with packet loss of 0.25% and latency of 50ms. These represent average conditions for a UK ISP. Under worst case conditions, the

- Users may find applications being blocked or not being usable due to reduced bandwidth. There could be subtle effects such as some types of email attachments not downloading properly whereas others do.
- A user may also see dramatic reductions in the overall performance of their connection compared with previous performance, perhaps because the data rate has been reduced as a result of crossing a usage limit.
- Users may observe variations in performance at different times of day, either because 'peak hour' traffic management is being applied to certain services, or because of simple congestion.

It follows that a user can observe performance of applications and infer the role of traffic management in producing that performance, but not know for certain. The status of traffic management being applied to the network or to a user's connection can only be known for certain by the ISP. The more prevalent traffic management becomes, the more consumers will tend to explain performance problems with reference to the use of traffic management. This implies the need for diagnostic tools to help users understand whether and in what way traffic management is affecting them.

Users will tend to be more alert to 'negative' effects of traffic management, such as applications performing poorly, than they will to 'positive' effects, which will tend to be taken for granted.

Traffic management on a user's connection versus traffic management in the network as a whole

There are debates over whether the effects of traffic management are positive or negative. To help bring clarity to this debate we have drawn the distinction between traffic management happening to other peoples' traffic (the network as a whole), and traffic management on one's own data (see Figure 8).

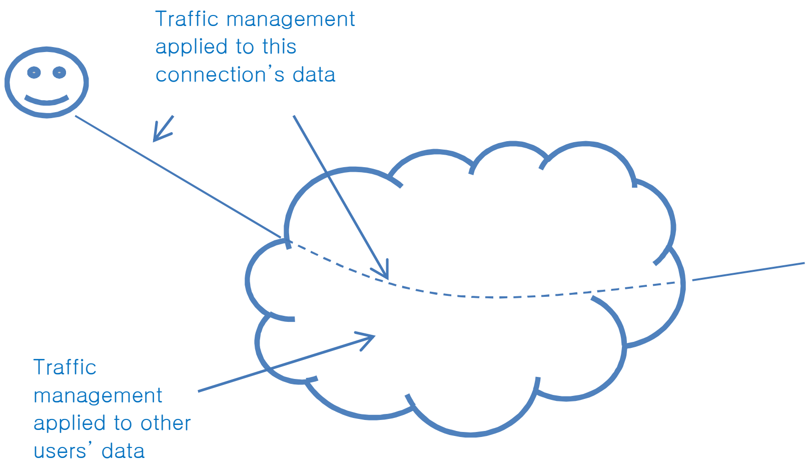


Figure 8: Traffic management of own and other users' data

---

iPlayer HD and the VPN were unusable according to our coding of QoE, and the MMOG exhibited noticeable problems.

Traffic management can have both positive and negative effects, as Table 8 shows. While the lower row in this table summarises the effect of traffic managing other peoples' data, measuring the effects would be virtually impossible[16]. We have therefore restricted our discussion to the effect on QoE of the traffic management applied to a user's own traffic.

	Positive effects on QoE	Negative effects on QoE
Traffic management applied to a user's own traffic	Can guarantee or prioritise data for sensitive applications	Can restrict or block certain applications
Traffic management applied to other people's traffic	Can reduce congestion to manageable levels, allowing fair use for all	Other people's traffic may take priority

Table 8: Positive and negative effects of traffic management

As stated above, the positive effects are unlikely to be as easy to observe as the negative effects, and unfortunately this has the effect of painting traffic management in a more negative light than perhaps it ought. Traffic management is justified by ISPs in terms of fairness. They consider that a relatively small number of users consume a disproportionately high share of resources, and that statistically more users benefit from traffic management than are disadvantaged.

[16] Even with perfect information on the traffic management is being used in the network as a whole, it would be logically impossible to determine the effect of this traffic management on a specific user.

7 User information requirements

The purpose of this chapter is to identify the information that consumers are likely to want.

7.1 When and why would consumers want information on connection performance?

In order to identify specific information requirements we have considered the situations in which consumers might want information on their internet connection performance. There are two basic consumer information situations, "prospective" and "in use", as shown in the flowchart (

Figure 9) and described in the following text:

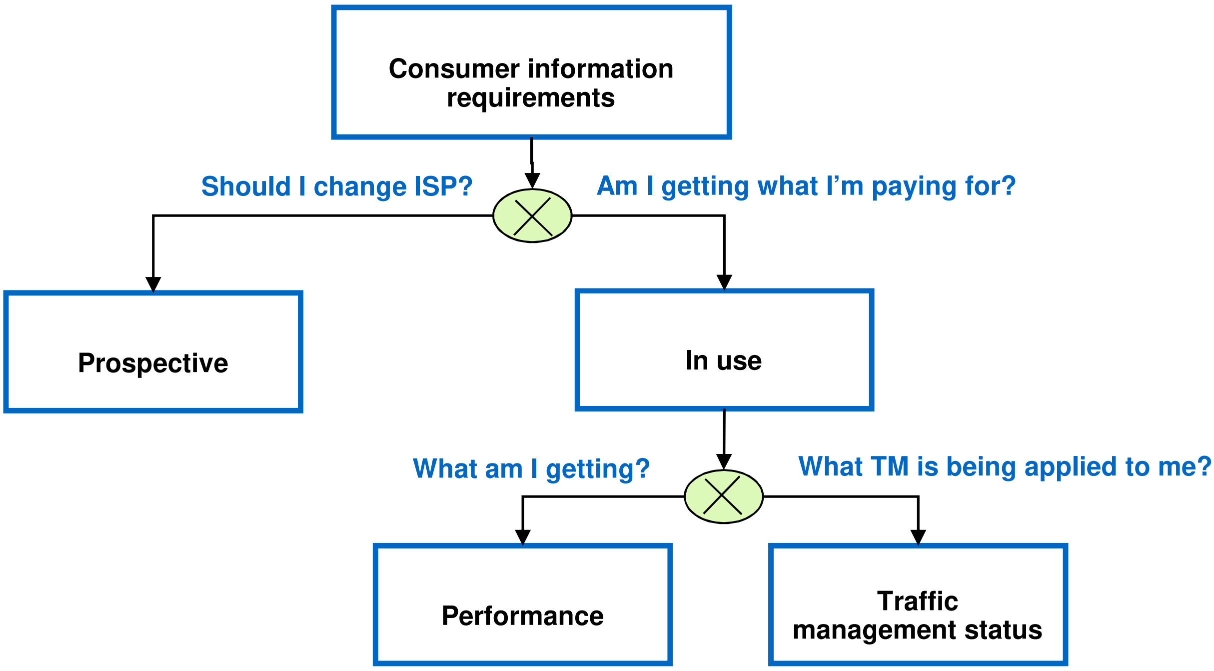


Figure 9: Consumer information contexts

Prospective: In the situation where consumers are deciding between ISPs or between packages, they will be making choices where at least one of the options is a prediction. The information required is whatever is necessary to check the suitability of a package, estimate costs and make comparisons of performance.

In-use: A consumer may wish to have information about the performance and the status of an existing connection. There are two classes of information in this category as explained below.

- **Performance checking:** Consumers may want to check at any time whether an ISP is delivering what was promised, or they may want to diagnose an apparent problem. Performance checking is always based on QoS measurements.

- **Traffic management status checking**: If traffic management becomes more prevalent consumers will increasingly want to determine whether they are being traffic managed. And if usage-based tariffs become more prevalent then consumers will want to find out how they stand relative to caps and thresholds.

7.2 Types of information: pull and push

We can distinguish between two types of information:

“**Pull”:** that which is requested by the consumer about an ISP or several ISPs services. “Pull” information requests are initiated by a consumer, with the prime functions of enabling consumers to check: (a) how their connection is performing at the time of the information request; (b) whether any generic (network level) traffic management is in place at the time of the information request; (c) whether any user-specific traffic management is in place at the time of the information request; (d) service levels and costs of different services from different ISPs, to enable an informed choice of provider; and

“Push”: that which is pushed by an ISP to customers of a service about their current service, or an alternative service which may better address the customers’ needs. “Push” information sessions are initiated by an ISP, and can carry information relevant to: (a) the network as a whole; (b) a customer’s connection performance; (c) a customer’s service usage; and (d) whether customers are on the most suitable package for their service usage. “Push” communications either enable an ISP to instruct/ control/ persuade a customer to behave in a certain way, or to appear helpful and open to the customer.

In section 7.3, we specify whether each of the items of information consumers might want about their connection performance could be served by Pull, Push or either mechanism.

7.3 What information would consumers want on connection performance?

Below we present a list of what kind of information consumers may need to know about their connection performance, split across prospective, performance and status. For each information item, we also state whether it could be met by Pull, Push, or both types of communications. This list is based on experience and logical analysis as we have not conducted empirical research into consumer concerns. We cannot vouch for the prevalence of each question. The numbering of the questions is for reference purposes and does not imply any priority order.

Prospective Information	
1.	To know which internet applications and services will be guaranteed (Pull);
2.	To know which internet applications and services will work (Pull);
3.	To know how well particular internet applications and services will work (even for internet applications and services known to be particularly sensitive to latency or line speed) (Pull);
4.	To be able to anticipate which internet applications and services are unlikely to work reliably or at all (Pull);
5.	To know which internet applications and services will be limited/restricted, and the details of such restrictions (Pull);
6.	To know which internet applications and services will be blocked (Pull);
7.	To estimate total costs of obtaining a service level to meet usage requirements (Pull);
8.	To judge what is the best package/ service level to meet usage requirements considering cost (Both Push and Pull).
9.	To judge the effect of managed services on the broadband connection (Pull)
Performance Information	
10.	To understand whether ISP is delivering what being is paid for at a discrete time (Pull);
11	To understand which network characteristic (QoS) may be stopping any internet application or service working when they do not work (Both Push and Pull);
12.	To know whether the problem can be bought/upgraded around (Both Push and Pull)
Status Information	
13.	To understand which traffic management function may be enabling or stopping any internet application or service working when they do or do not work (Both Push and Pull);
14.	To know whether the problem can be bought/upgraded around (Both Push and Pull)

Figure 10: Detailed consumer information requirements

7.4 Information requirements by scenario

In the table below we list the information types (those listed above) which could be useful or relevant to consumers in each of the five future traffic management scenarios (as presented in Chapter 5).

	Scenario				
	Scenario 1 Fair use	**Scenario 2** Congestion management	**Scenario 3** Video streaming	**Scenario 4** Business tool	**Scenario 5** Managed services
Prospective Information					
1 What's guaranteed?	✓	✓	✓	✓	
2 What will work?	✓	✓	✓	✓	
3 How well will it work?	✓	✓	✓	✓	
4 What will not work	✓	✓	✓	✓	
5 What's limited?	✓	✓	✓	✓	
6 What's blocked?	✓	✓	✓	✓	
7 Cost?	✓	✓	✓	✓	✓
8 Best for me	✓	✓	✓	✓	✓
9 Effects of managed service?		✓	✓	✓	✓
Performance Information					
10 As promised?	✓	✓	✓	✓	✓
11 What QoS is wrong?	✓	✓	✓	✓	✓
12 Can I buy better?	✓	✓	✓	✓	✓
Status Information					
13 What traffic management in action?	✓	✓	✓	✓	
14 Can I pay around?	✓	✓	✓	✓	✓

Table 9: Summary of relevance of information in the five traffic management scenarios

As can be seen from the table, two points are of particular note. First, nearly all the information items may be of interest to some consumers in most of the scenarios. Second, while there are some differences, the scenarios do not differ substantially in terms of the information that may be required. The measurement implications of the above analysis are picked up in Section 9.3.1.

8 Designing a communication approach

8.1 Principles of information transparency

Through our research activities and from synthesising our findings, we have identified three general principles which should underlie any transparent communication about traffic management, so that it is:

- **meaningful** (e.g., have utility),
- **accurate** (e.g., be valid/true), and
- **comparable**.

While these concepts have been derived from the team's understanding of what consumers need, based on experience of consumer decision making in other contexts, similar concepts have been identified and discussed in current academic and applied research literature[17].

Two of the dimensions (meaningfulness and accuracy) are extensively referenced in the literature on information quality. As described below, given sufficient knowledge on what information is needed for and by whom, in the majority of instances it is relatively straightforward to generate guidelines or rules to support the dimensions of meaningfulness and accuracy for transparent communication.

In the context of information about traffic management, comparability as a dimension of transparency is a somewhat more complex principle. Comparability will become increasingly dependent on user context as more complex traffic management regimes are implemented. In essence, to make informed comparisons based on multiple complex rules requires the processing of a lot of rules simultaneously. Solutions to such difficulties in comparability can include the development of representative user contexts or scenarios (e.g., "average family household", "online gamer"), or automated advisers or wizards to recommend a service offering best suited to different usage criteria input by a user.

Here we discuss each of the three dimensions introduced above in more detail.

Accuracy

Any information provided to consumers should be as accurate as possible, whether in a written policy, or real-time status or usage updates. In complex domains with simultaneous variation in multiple dimensions, perfect accuracy can be impossible – there will always be some irresolvable error. In fact, there is always a trade-off between the cost and benefit of seeking accuracy. In the domain of information about the performance of an internet connection and effects on that performance of any traffic management, there are two potential sources of inaccuracy.

- **Predicting performance of access into an ISP**. There is a potential source of inaccuracy in relation to information about the quality of service any particular phone line is able to support. Without performing a test of line quality, it is impossible to know whether a specific

[17] references: http://www.epa.gov/oei/symposium/2010/worthington.pdf; http://asbbs.org/files/2010/ASBBS2010v1/PDF/C/Caratelli.pdf; http://www.google.co.uk/url?sa=t&source=web&cd=13&ved=0CCwQFjACOAo&url=http%3A%2F%2Fpublications.accion.org%2Finsight%2FIS24EN.pdf&rct=j&q=dimensions%20of%20transparency%20of%20consumer%20information&ei=r1BLTcGrAsOxhQeA9vy_Dg&usg=AFQjCNFFwVXPdS8I6V10R_knvBV7hO8jZg

ADSL connection will be able to support, say, an 8Mbps connection, or substantially less bandwidth. There is therefore a limit to the accuracy of prospective information about what services can be supported for a specific customer. Descriptions of the statistical distributions of what services an ISP's customers receive are of course possible, in general terms. But these are different to direct, specific predictions about a prospective customers' line.

- **Forecasting**. It is not possible for an ISP to predict with perfect accuracy how congested its network will be in the future. Given this, where traffic management is triggered by congestion it is not possible to know what levels of traffic management will be in operation at any time of the day, or the exact impact of the traffic management, in advance of its operation and impact. Despite the occurrence of big unexpected events, past performance is as good a guide as there is, and enables intelligent predictions.

Additional accuracy in predictions of performance of access into an ISP, and in forecasting, may be technically feasible, for example through the addition of probes to the network (see section 9.2), but the costs and benefits of obtaining such increased accuracy need careful consideration.

In the context of this discussion, considerations of accuracy as a dimension of the transparency of communications to consumers about traffic management and its effects must centre on whether:

- available measures are communicated honestly;
- the output of future forecasting is communicated honestly;
- complex statistical information is represented meaningfully to users..

Meaningfulness

Any information provided to consumers about traffic management, whether it describes a policy, the effects of implementation of a policy, or a usage allowance, should be meaningful to the target recipient. Important elements to consider in ensuring information is meaningful, are:

(a) that it is **relevant** to the consumer's situation (i.e., that they will be motivated to access the information, a feature of the "what"); and

(b) that it is represented using **easy to understand units, concepts and terminology**, appropriate to the technical literacy of target recipients (the "how").

In relation to relevance, a key consideration relates to what consumers want to use their internet connection for. We assert that consumers generally use the internet to access specific functioning applications and services (for example: websites, video streams, games, social networking sites, email, social interaction, VoIP and so on). This assertion in relation to relevance provides some useful indications or pointers in relation to the second component of meaningfulness, the use of appropriate terminology, units and concepts with which to communicate.

Target recipients of information (about what services are supported, how well and for what price by different ISPs) obviously constitute a diverse audience, with a broad range of technical literacy. Given this broad distribution, a challenge for ensuring communications are meaningful is getting the right balance between excessive simplicity and baffling complexity. A related balance to achieve is between providing too little and too much information. An effective way to address these challenges is to provide different levels of information to different types of user. For example, one view could provide the information that any non-expert user would want/need, with an alternative detailed/experts' view available for those interested. Using this approach, a small minority of people may need additional assistance in understanding the information. This solution is not ideal, but it may ensure optimal meaningfulness for the most people.

With regard to the impact of traffic management on QoE, the most meaningful information possible to communicate to interested users is whether they can use an application or service they wish to; and, if the answer is uncertain, to inform them of their options to ensure access. This would tend to suggest communicating directly about specific applications and services (such as video, photos, video conversations, gaming) in meaningful units (e.g., time – hours, minutes; or units – for example episodes, calls).

It is of course the case that technical literacy develops over time in a population, whether incrementally through the accretion of technical knowledge or experience, or in generational step-changes. As technical literacy develops it is important to ensure that communications remain meaningful to target recipients. It is worth noting here that making metering transparent can be an effective means of supporting the implicit learning of the resource consumed by a device or activity.

Comparability

For information provided to consumers about traffic management to be of use in their decision making requires that different ISPs provide comparable information, i.e., information about the same (potential) variables or features of a service, in comparable units, concepts and terminology. In relation to comparability, a key consideration relates to whether any or all ISPs are able to provide the required information to enable an interested consumer to compare ISPs. Another key consideration relates to the volume of data presented – less information can be easier to digest and compare, and therefore more transparent.

- There are some limits to the comparability of information about traffic management. For example, as described earlier in this report, some ISPs deploy traffic management at specific times of day, whilst others use monthly usage caps. To compare the likely impact of these policies directly is not easy – as there is no automatic translation between the two dimensions.
- To address the fact that different approaches to traffic management are used by different ISPs, the following are potential approaches to support comparability:
    - provide the basic information but do not force any comparability
    - force commonality of practices (data/month or hours/day) to enable direct **price/ quality** comparisons;
    - use generic scenarios to create synthetic comparables (e.g., a typical household) on **price/quality**;
    - personalised scenarios (based on your data/service use over time – this is what it would **cost** you with ISP A, B C...);

Understanding consumer preferences is also important in supporting comparability. Passive consumers may want to be told which suppliers can meet their needs at the best prices. More active/ informed consumers may have a rough wish list and want to evaluate potential trade-offs they can make to get as near to obtaining their wishes as possible whilst saving as much money as possible.

8.2 Prospective information approach

The requirement for prospective information can be met by three elements as listed and detailed further below.

- A QoS Policy Form which gives a complete description of the policies in operation and performance data where policies alone are inadequate'
- A QoE Summary which gives meaningful and comparable information for the popular services'
- A 'wizard' which is capable of helping a consumer choose packages that are appropriate to their needs.

8.2.1 QoS Policy Form

The QoS Policy Form characterises the connection as a whole, the baseline for ordinary traffic, and specific information on any traffic type that is separately managed.

With respect to the policy, the options range from giving a general impression ("We may additionally manage customers' data connection at peak times to preserve the best experience for the greatest number of users") through to being completely specific ("Between 4pm and 11pm your bandwidth for streaming video will be limited to 1Mbps").

For a policy to be 'transparent', it should ideally inform the user precisely what the performance of the connection will be. Some forms of traffic management – such as restricting bandwidth - do allow for predictability. Unfortunately, this is not always possible because performance of a network is generally dependent on the level of traffic, which is not completely predictable.

The default treatment of traffic is 'best efforts'. That is, the ISP attempts to deliver a packet but makes no guarantee to do so, especially if there is congestion. The policy should specify all departures from this default. Thinking ahead to more complex traffic management policies than we currently have, all policies involve applying a traffic management action either unconditionally or in a way that is triggered by some other factor. The policies should be as specific as possible in both describing the condition and describing the traffic management action and its effects.

Fortunately, the technical survey of traffic management suggests that there are relatively few different types of traffic management and that the different options can be represented using a common template as shown below.

For the overall connection, the table below sets out essential information.

Data rate	
Achievable rate, taking into account as much of the user's situation as is reasonable *	

Volume	
• Unlimited	Choose between these and provide the information
• Capped (specify the cap , the alerting procedure, and the consequences of going over the cap)	

For each traffic type, the following must be described. The two techniques of traffic management are covered. Priority level' refers to packet prioritisation and 'data rate' refers to bandwidth allocation.

Traffic type	
Specify the type of traffic	

Priority level	
• Guaranteed	Choose between these and specify the criteria (e.g. times of operation)
• Accelerated (specify criteria) *	
• Normal *	
• Restricted (specify criteria) *	
• Blocked	

Data rate	
• Guaranteed minimum (specify)	Choose between these and specify the data rate in both bps and indicative units of consumption (e.g. type of video supported)
• Maximum, not specially limited	
• Capped or restricted (specify)	

Volume caps	
• Unlimited	Choose between these and specify the data rate in both bps and indicative units of consumption (e.g. hours of video)
• Capped (specify the cap , the alerting procedure, and the consequences)	

Historic data	
If starred above, provide performance graphs based on past measurements. For a new service the performance graphs must be estimated.	

In the above table:

- All classes of traffic not treated on a best efforts basis should be identified.
- There is no requirement to describe how traffic is identified but the description should be sufficiently complete that a user would know whether their traffic is included.
- For each traffic class, the way in which it is handled should be described in as much detail as possible.
- There is no requirement to provide absolute certainty where the impact is affected by demand – which is not fully predictable.
- A reasonable level of certainty should however be offered through historic data or models of expected performance.
- Data on the QoS achieved is more meaningful than data on the level of traffic management employed.
- Where the use of one service (e.g. managed IPTV) affects the use of other services, then the dependency should be explained.
- Particular attention must be paid to services which are, in practical terms, blocked.

We have illustrated an example QoS Policy Form in Figure 11.

ISP Tariff Example

Data rate	5 Mbps
Volume	40 GB/month Email warning at 75% and 90% of limit. Above cap, user can choose to pay a supplement or upgrade. Residual 100kbps is still provided to allow browsing and email

Traffic type	P2P Recognised P2P such as BitTorrent, eMule, Gnutella and newsgroup applications
Priority level	Restricted Between 6pm and 12pm, P2P on our network is deprioritised.
Data rate	Capped All users apart from our 'Mega' tariff are restricted to no more than 2 Mbps upload speed
Volume caps	None
Historic data	See data rates as measured for 3 months Jan 2010 to Mar 2010

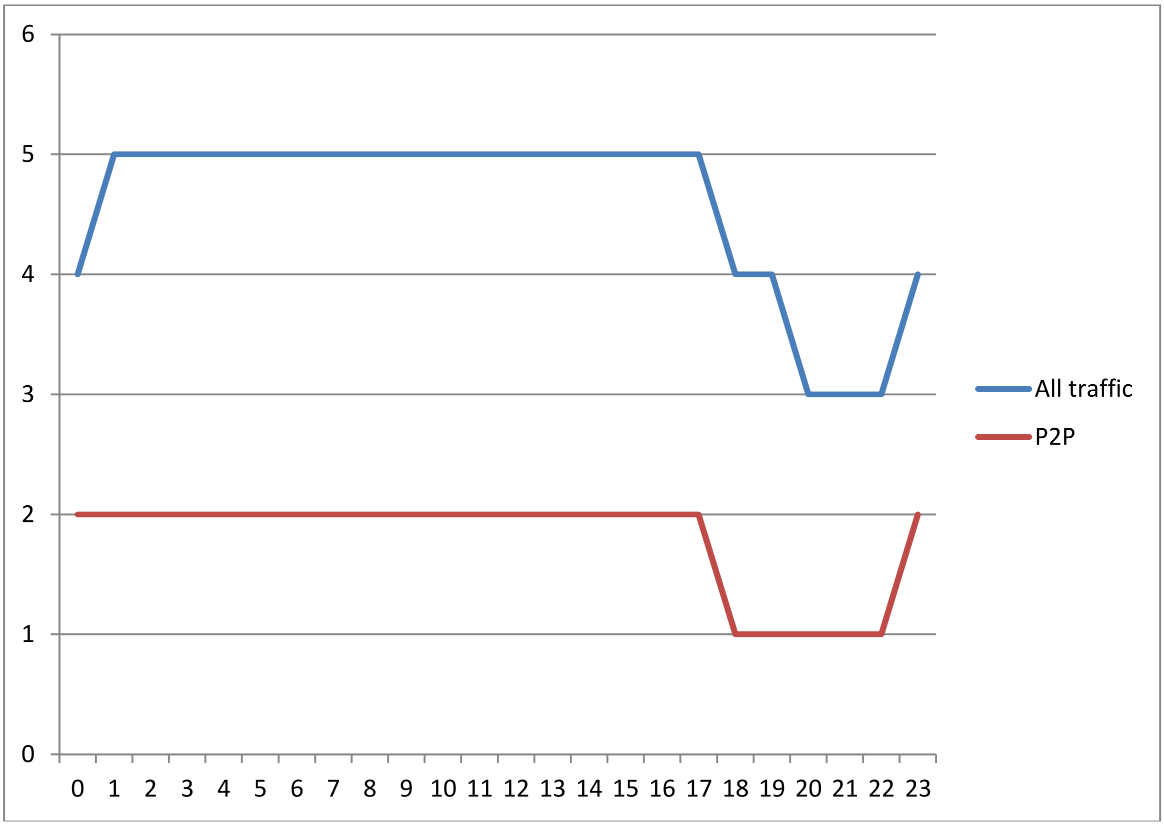


Figure 11: Illustrative example of a QoS Policy Form

8.2.2 QoE Summary

The requirements for each type of information (prospective, performance and status) can be described by different sources, with the two in-use information types most amenable to direct display of current status. In the case of performance, this can be described by real-time quality of service measures (of bandwidth, latency, jitter and packet loss) – which is effectively an aggregate of the inherent network QoS with any QoS effects of traffic management overlaid, at the time of measurement. In the case of status, this can be described simply as a summary of what traffic management, if any, is being applied to a particular connection (or the network as whole) and what proportion of any periodic allowances have been used at the time of measurement.

The most complex question is how to represent prospective information transparently, given the range of relevant variables of which it is constituted.

Transparent representations of Prospective QoE information

Whilst, as discussed further below, we anticipate the development of any consumer-facing representations and graphics describing prospective QoE information to be delivered by the market, according to criteria agreed collectively by ISPs, we have demonstrated in the figure overleaf how it is possible to categorise the service levels of different (hypothetical) ISPs transparently (so that they are meaningful, accurate and comparable).

In the examples here, several service groupings relevant to consumers are represented: Standard Definition video streaming, High Definition video streaming, online gaming, VPN, video conferencing over IP, voice over IP, music streaming, music downloads, video downloads, P2P file sharing, and day to day online activities, labelled e-life: comprising web search/ browsing and transactions.

We have developed, and present here, simple logos – not as suggestions for formal communications, but to test whether to do so is possible. The logos are shown in Figure 12. Logos/graphic representations are presented for downloading, streaming and P2P data methods, and to illustrate where a service's bandwidth is throttled. We also present logos/ graphic representations for a number of service groups: video, audio, gaming, conferencing, P2P. And we suggest how usage limits or caps could be communicated, again in meaningful and relevant units to consumers. The illustrations are not meant to be exhaustive nor complete, but a demonstration of what is feasible.

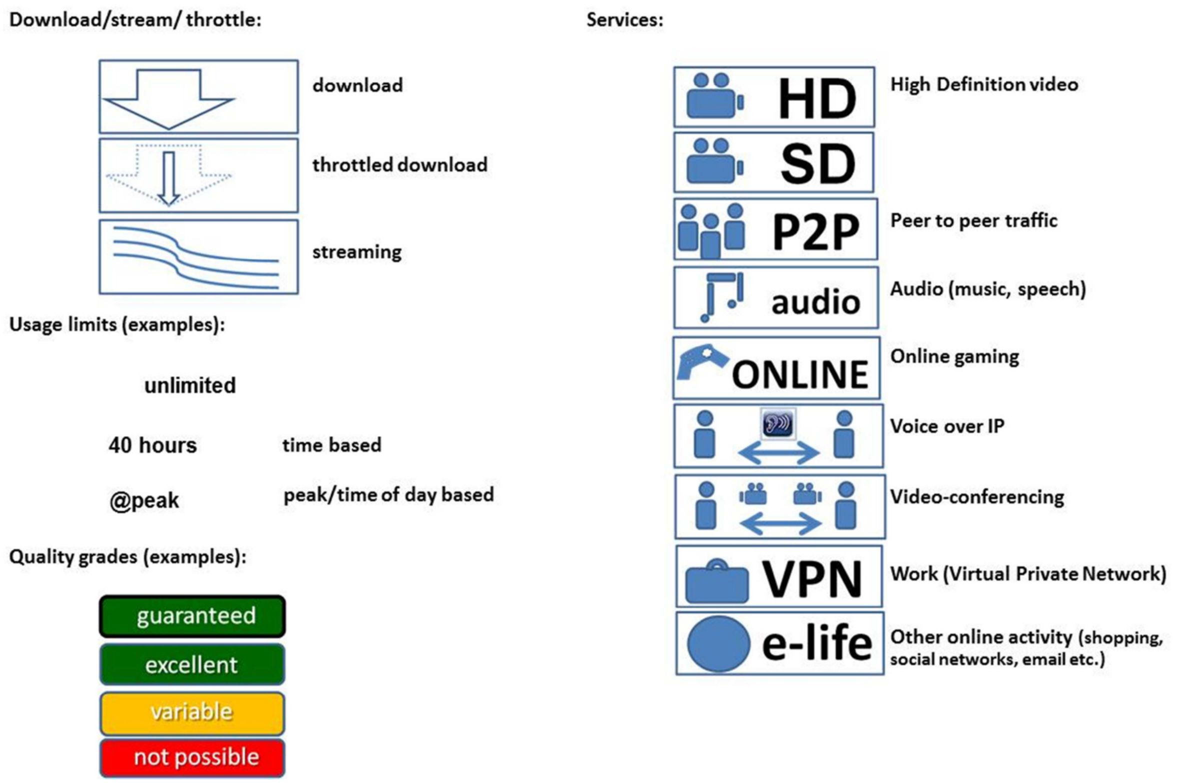


Figure 12: Key to logos

In Figure 13 we have represented the services for three hypothetical ISPs using these logos. We have then grouped the services according to whether the services are: (a) guaranteed (effectively managed services), labelled ‘guaranteed’; (b) strongly expected to work well, labelled ‘excellent’; (c) expected to fail at some points of the day (typically during peak contention and congestion), labelled ‘variable’; or (d) are blocked or firmly expected not to work, labelled ‘not possible’. The three hypothetical ISPs have different package profiles which are used here to illustrate the use of the logos to produce meaningful graphic representations.

The universal implementation of a similarly transparent approach should allow non-expert consumers to identify easily whether a specific ISP’s service/tariff is likely to meet their needs, and to compare at a glance between different ISPs/tariffs. These are most meaningful if the majority of the QoE seen by the consumer is the result of deliberate actions by the ISP, or limitations known to the ISP (e.g. bandwidth restrictions). If the QoE is dominated by other factors, such as radio propagation for a mobile consumer, then the graphic won’t be such a good guide to performance.

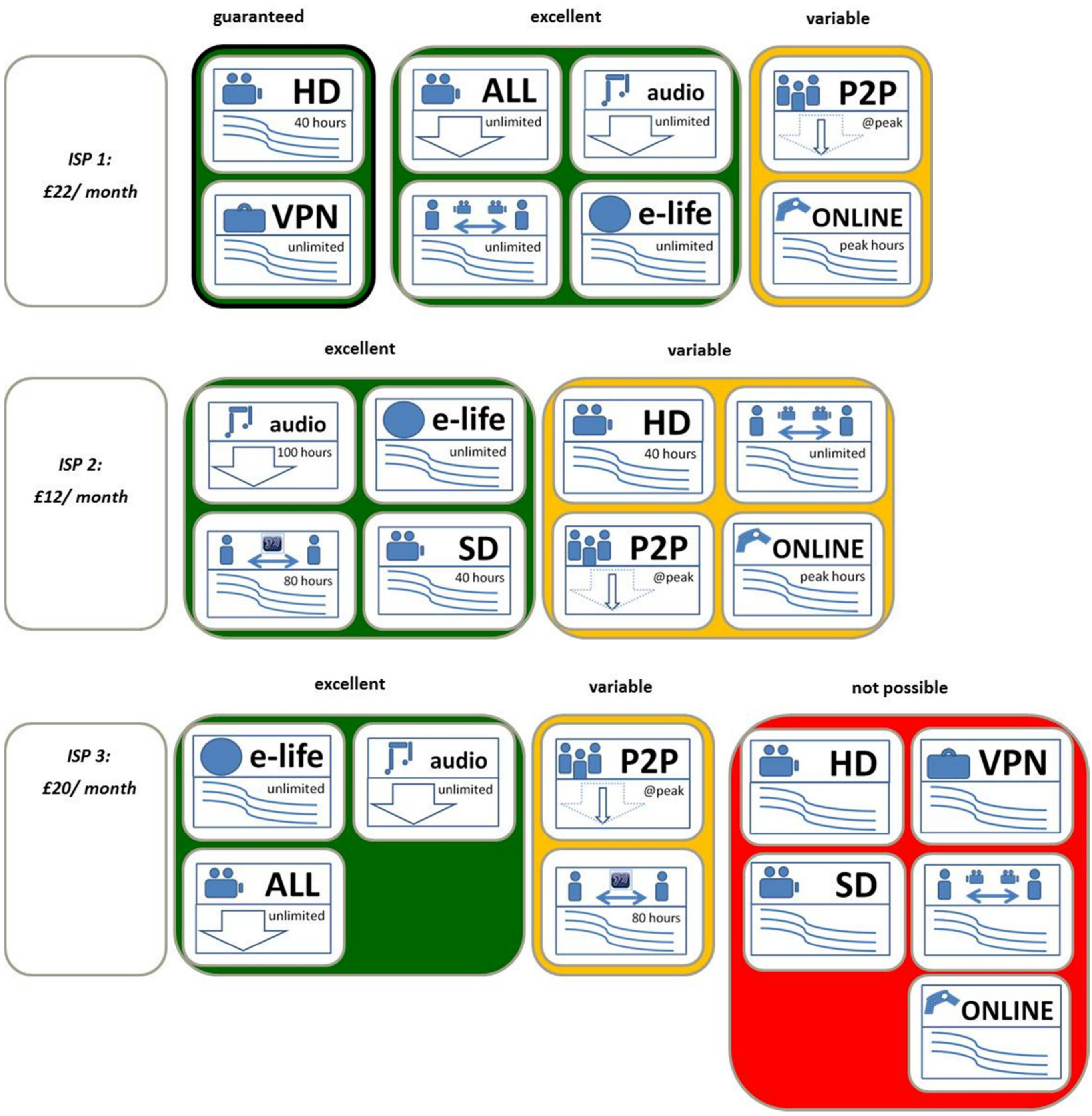


Figure 13: Illustrative QoE Summary for 3 hypothetical ISPs

8.2.3 Thresholds

The implementation of an information provision system such as that outlined above of course requires QoS thresholds to be agreed and set in order that the information provided in such a summary table is as accurate as is possible. Consideration will need to be given in particular to the thresholds between the categories “excellent” and “variable”, and between “variable” and “not possible” (where “not possible” is because of a QoS limitation rather than traffic management policy based blocking). A reasonable approach could be that responsibility for agreeing and setting these thresholds be devolved to ISP industry bodies such as the Broadband Stakeholder Group.

8.2.4 Wizard

For some consumers, a wizard approach may be preferable to the display of a lot of information. Here, consumers enter data describing their usage of services and a wizard recommends an ISP service package appropriate to these needs. A wizard may be most relevant in scenario 4. The design and implementation of any wizard is, again, best left to the market but in order for the wizards to be useful all ISPs must make data available describing the services supported by any of their broadband packages.

8.2.5 Applicability to scenarios

All five of our scenarios can be represented using the QoS Policy Form and the QoE Summary. The more complex traffic management scenario (4) may additionally require the use of a wizard to help people choose. In all cases, we think that ISPs could be asked to agree to provide data to third party wizards if requested.

8.3 Performance and status information approach

8.3.1 Performance

Once threshold QoS parameters are agreed for any service to work, translating a real time network performance test to a working/not working indication for any service is a relatively straightforward task. In addition to the straightforward QoS information, a potential transparent approach to communicating the results of a QoE performance check is shown in Figure 14 below.

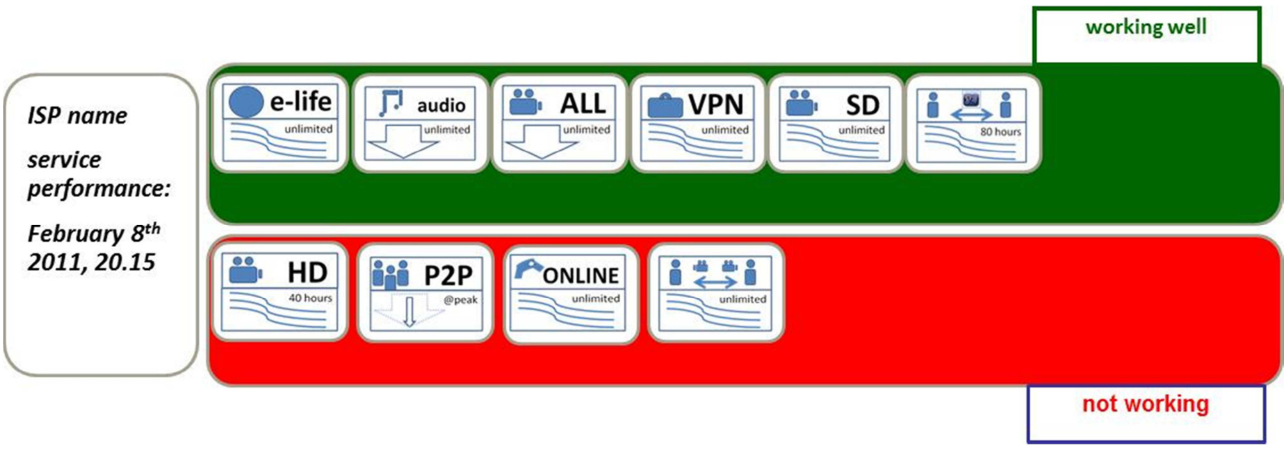


Figure 14: Illustrative transparent QoE performance representations

8.3.2 Traffic management status checking

We defined traffic management status checking above as a process enabling consumers to find out how they stand relative to usage caps and thresholds, and, if possible, to check what traffic management is being applied at the time they make the check.

A potentially transparent approach to communicating the results of a traffic management status check is shown in Figure 15. Whilst simple text descriptions of time remaining are used here, many similar representations could work (e.g., egg timer, pie chart, % used illustrated on a bar) as would simple text statements describing usage relative to caps and thresholds.

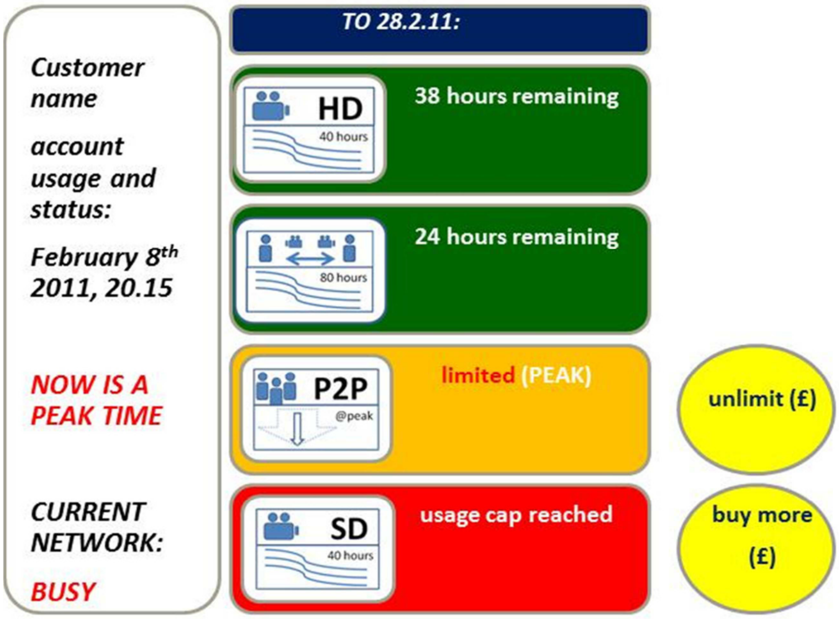


Figure 15: Illustrative transparent traffic management status representations

It is also worth noting that traffic management status updates are those most amenable to "Push" information sessions initiated by an ISP. Examples of traffic management status which could be communicated by an ISP via Push notifications include information relevant to: (a) the network as a whole; (b) a customer's connection performance; (c) a customer's service usage; and (d) whether a customer is on the most suitable package for their service usage.

9 Measurement approaches

9.1 Introduction

This chapter examines the options for providing the data necessary to support the user information requirements introduced in chapter 7 and developed in chapter 8.

Figure 9 in chapter 7 described the reasons why information might need to be provided. In Figure 16 we have identified that the need for prospective information gives rise to the need for time series data capable of predicting future performance, and real time data capable of being used to check performance and the status of a user's usage relative to any caps that may exist.

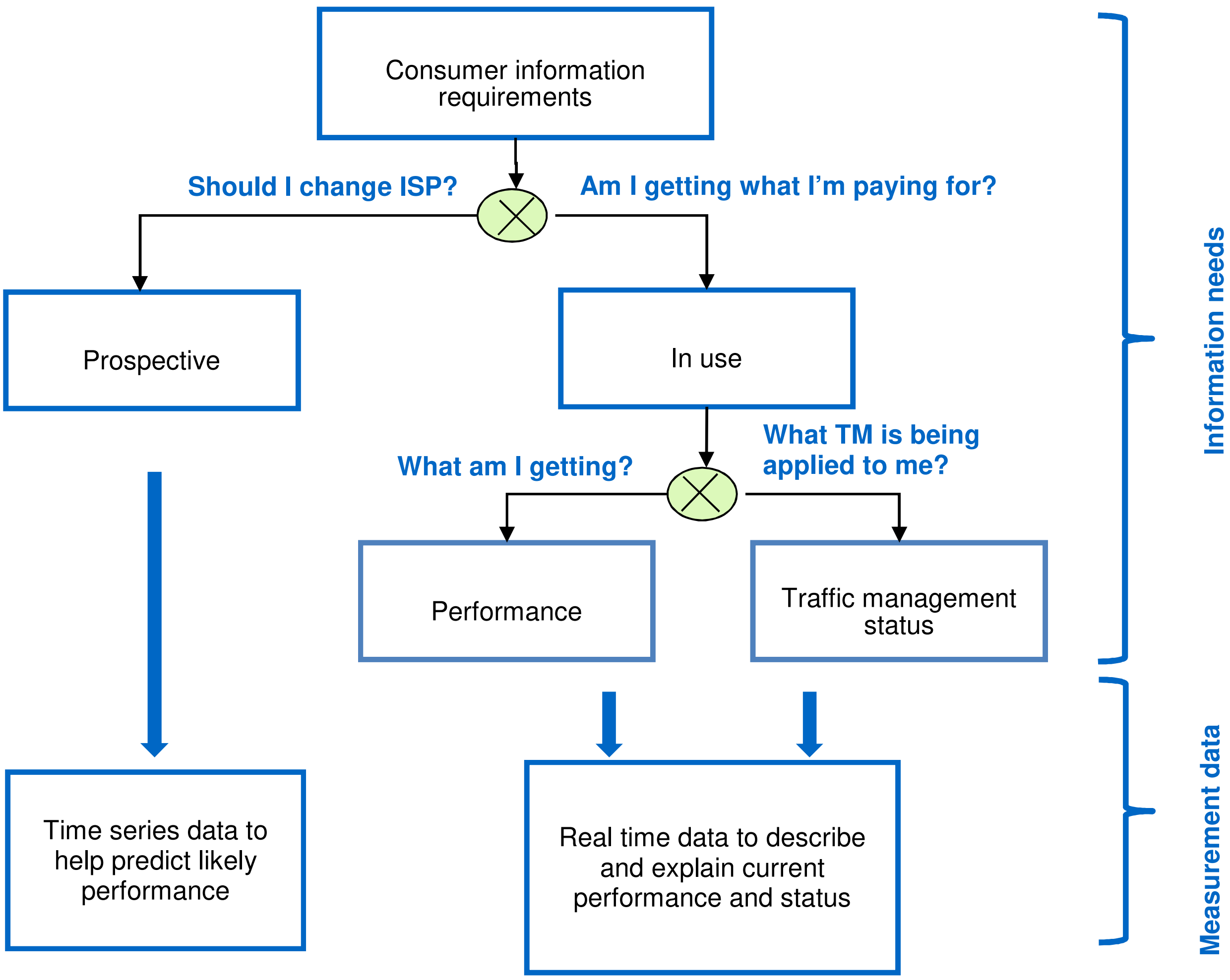


Figure 16: Measurement data requirements arising out of user information needs

It is helpful to consider the two types of measurement data separately, though in practice it is possible that time series data could be generated by collecting and aggregating the real time data.

9.2 Measurement options

There are a number of different dimensions to measurement and we start by discussing these.

Method

Quality of service can be measured either on existing traffic (typically called 'passive' measurement) or on specially injected traffic (typically called 'active' measurement). For the purposes of detecting traffic management at a user's connection or device, the active approach of injecting traffic allows the QoS of specific types of traffic to be detected. However, if traffic volumes are capped, such an approach will eat into allowances.

Location

There are many places within a network and its attached devices where measurements can be undertaken. Principally, these are:

- the user's device, typically a PC or phone;
- a special purpose monitoring device at a user's connection;
- within the ISP's network: at nodes or line cards - network equipment will often incorporate the ability to produce traffic statistics[18] though this ability may not in practice be used;
- at a content provider.

Temporal

The options range from occasional spot tests through to continuous monitoring in the background. Some equipment is able to detect quiet periods at a user's connection and run tests at that time.

Initiation

Tests can be initiated or controlled by a user, an ISP or a third party.

Having researched the field and spoken to ISPs we consider that there are five main options which are illustrated in Figure 17 and summarised in Table 10.

---

[18] This DSLAM incorporates stats - http://www.huawei.com/broadband_access/products/dslam/ip_dslam.do?card=2

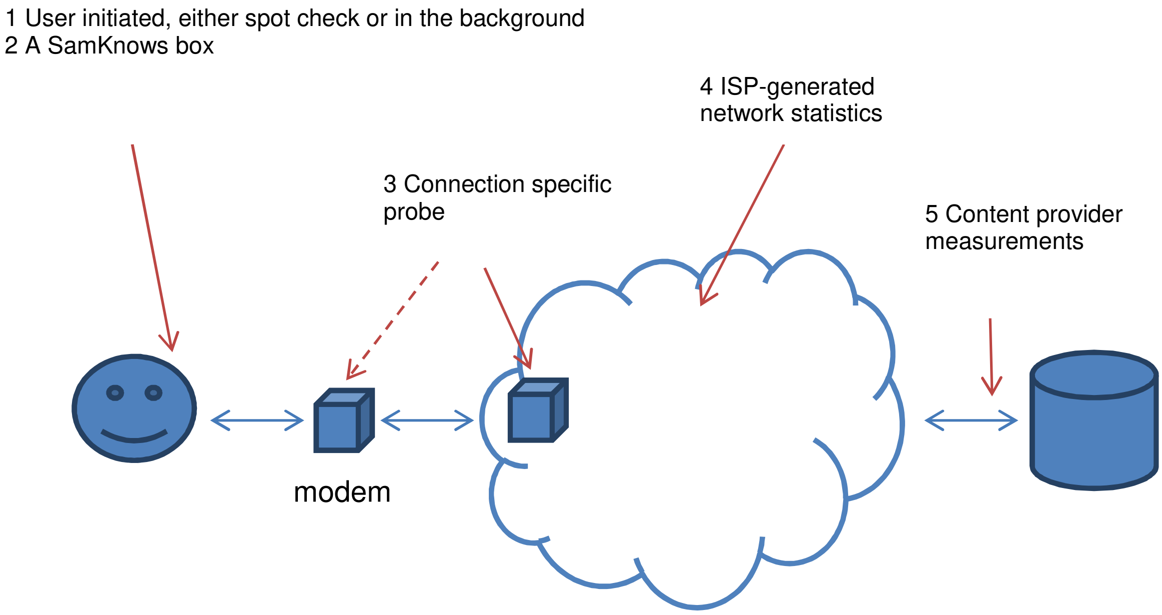


Figure 17: Options for data capture

1 User initiated

The principle of running broadband measurement applications on a user's device is established. Sites such as pingtest.net and speedtest.net interact with remote servers to provide basic QoS data. An application called glasnost[19] performs similarly but sends and receives different data types in order to calculate whether there is any traffic management in operation. Such tests are useful for troubleshooting but there is the possibility that ISPs could detect the packets and prioritise them in order to manipulate the result (though we emphasise that we have no indication that such practices are occurring).

It is also possible to have software measuring performance as a background process or as part of an application. These may not easily be possible in some situations, especially on mobile devices where they may impose a significant overhead.

These applications typically send data to a site to be aggregated. Whether this gives a realistic overall picture is questionable because they will tend to be used mostly when problems have been experienced.

2 SamKnows box

Here, software resides either in a special purpose device or in a router or modem, and runs tests initiated and aggregated externally. The company SamKnows has developed this technique and currently monitors six ISPs for Ofcom. The technique is explained in a paper on the SamKnows website[20]. We will refer to the approach as SamKnows because it is the best example of this approach in the UK, though it is no doubt possible for similar techniques to be used by other companies[21].

---

[19] http://broadband.mpi-sws.org/transparency/bttest.php
[20] http://www.samknows.com/broadband/pm/PM_Summer_08.pdf
[21] References to SamKnows should be seen as generic to the class of technique rather than specific to that company

The SamKnows approach may not be feasible in all circumstances. The software works by generating and sending extra data. This causes an overhead, which may not matter in most fixed broadband contexts, but could eat up allowances in mobile broadband. If a traffic management intervention only applies to a small number of users then the effect may not be noticeable in the sorts of sample size typically used. It may also be inconvenient for the user if, say, volumes of data are injected purely to trigger a volume-related intervention.

In the case of home networks it is necessary to run the tests from a device with priority over other traffic on the network.

3 Connection-specific probe

This option involves measuring aspects of quality of service at a connection from within the network, or from the modem. The ability to measure some parameters is incorporated into some network equipment but may not be used in practice. Such measurements need to be made near the edge of a network where the user context is known. There would be considerable additional cost and complexity in adding a workable measurement capability in some networks.

This option is particularly relevant to the reporting of usage statistics. The architecture of traffic management has a big effect on what is feasible. If traffic management is centralised then data will only be available in an aggregate way. Not all ISPs can offer all details related to an individual connection at present.

4 Network statistics

This technique puts the onus on ISPs to make measurements on their network that characterise QoS and the effect of traffic management. How the measurements are made can depend on the architecture of the network. Network nodes are often already able to report the behaviour of queues, and therefore latency and loss statistics. It may also be possible to aggregate connection specific measurements. Alternatively, in the absence of specific data, it may be possible to model QoS for the expected network conditions.

5 Content provider

In principle, measurements can be performed from a content provider's server. It is likely that from this point it would be possible to expose any differences between ISPs, or how aggregate QoS varies over time. However content providers do not have access to information on users' tariff packages and will not be in a position to determine how much an ISP is actively managing traffic on a particular tariff which, we think, is what consumers would want to know.

9.2.1 Summary

	Method	Location	Temporal	Initiation
1 User initiated	Existing traffic or Injected traffic	User device	Spot test or continuous in background	User
2 SamKnows box	Existing traffic or Injected traffic	User device	Quiet periods	External
3 Connection-specific probe	Existing traffic	Modem or DSLAM/CMTS/Node B	Continuous	ISP
4 ISP-generated network statistics	Existing traffic	Network nodes	Continuous	ISP
5 Content provider	Existing traffic	Server	Continuous	Content provider

Table 10: Summary of characteristics of the main measurement options

9.3 Evaluation

In principle, with enough investment, all the desired measurements can be made, and all the desired data can be generated. In practice we think that such investments should be considered from a cost-effectiveness standpoint. We have considered the applicability of each of the measurement approach to each of the scenarios, and taking into account the scenario and the measurement options we suggest the following.

9.3.1 Prospective

	Scenario				
	Scenario 1 Fair use	**Scenario 2** Congestion management	**Scenario 3** Video streaming	**Scenario 4** Business tool	**Scenario 5** Managed services
Prospective Information					
1 User initiated					
2 SamKnows box	✓ (not to detect TM)	✓ (not to detect TM)	✓	✓ (not with complex tariffs)	✓
3 Connection-specific probe					
4 ISP-generated network statistics	✓	✓	✓	✓	✓
5 Content provider					

Table 11: Options for prospective measurements

The two main approaches relevant to a consumer looking to evaluate ISPs other than his/her own are the SamKnows approach and ISP-generated network statistics. Because scenarios 1 and 2 would tend not to affect enough users to make a SamKnows approach cost effective, we consider that data will be better provided from within the network. For the other scenarios both of these techniques can be used. However the SamKnows approach may not be feasible with complex tariffs (scenario 4) because of the need for a large enough sample of each tariff. In the case of mobile networks, the SamKnows approach may impose an unacceptable overhead, suggesting that network statistics are the only solution. Not all QoS parameters are easily measured from within a network, however.

9.3.2 Performance checking

	Scenario				
	Scenario 1 Fair use	**Scenario 2** Congestion management	**Scenario 3** Video streaming	**Scenario 4** Business tool	**Scenario 5** Managed services
Performance Information					
1 User initiated	(✓)	(✓)	(✓)	(✓)	(✓)
2 SamKnows box					
3 Connection-specific probe	(✓)	(✓)	✓	✓	✓
4 ISP-generated network statistics	✓	✓	(✓)	(✓)	(✓)
5 Content provider					

Key: (✓) indicates that the option may be appropriate but with some qualification

Table 12: Options for performance measurements

There are more options in this case. ISPs may not recognise user-initiated measurements as valid, so the best option is ISP measurements. We consider that the lower variability in scenarios 1 and 2 will tend not to require connection specific probes. SamKnows boxes are not appropriate to this application because only a small proportion of users have these boxes.

9.3.3 Traffic management status checking

	Scenario				
	Scenario 1 Fair use	**Scenario 2** Congestion management	**Scenario 3** Video streaming	**Scenario 4** Business tool	**Scenario 5** Managed services
Status Information					
1 User initiated					
2 SamKnows box					
3 Connection-specific probe	✓	✓	✓	✓	✓
4 ISP-generated network statistics	✓	✓	✓	✓	✓
5 Content provider					

Table 13: Options for status measurements

The information is essentially only held by the ISP so it needs to be provided from within the network, either connection-specific or for the network as a whole.

9.4 Conclusion

There is no one solution to the provision of measurement data.

In the case of prospective data capable of giving greater certainty over the details of how a traffic management policy is working in practice there are two main options – the SamKnows approach and ISP generated network statistics. Both of these approaches are capable of providing consistency and repeatability. The SamKnows approach is based on sampling, so where the tariff or circumstance has low prevalence, the SamKnows sample will possibly be too small to provide statistical reliability. Alternatively, where there are numerous different tariffs and situations, the number of distinct SamKnows samples will tend to proliferate. Using ISP generated data can potentially get around these problems but such data has its own limitations and there might still be a need to audit data produced by ISPs.

For real time performance information, there is always the possibility of users initiating tests using third party software and servers. Alternatively, ISP data can be used. Once taken in conjunction with the need for status information, it becomes clear that ISP data (either connection-specific probes or network statistics) would be a strong option. We envisage that the ISPs would provide a performance and status interface in real time. This should be connection specific if possible, but some aspects could be representative if connection specific data are not available because of the architecture.

Mobile broadband is more difficult in this context than fixed. The susceptibility of mobile broadband to all sorts of factors that affect the radio access network means that past data may not have much predictive power. Data can be collected from within the network but will have to be aggregated and will not be very relevant to what is essentially a location-sensitive service.

10 Conclusions

The current approach adopted by most UK ISPs to traffic management can be characterised as ‘minimalist’. This form of traffic management is mainly designed to promote ‘fair use’ so that heavy users are not able to consume such a disproportionate level of network resources that they degrade the service available to moderate or light users. The information that ISPs currently provide to consumers on their traffic management describes their policies in broad terms but often not in sufficient detail that a user can predict the precise level of traffic management they will experience. This is partly because the level of traffic management at any one time can depend on the level of congestion, which in turn is subject to statistical variations.

This report has described a series of future scenarios that would involve greater use of traffic management. The scenarios have been developed by considering commercial and technical trends. These scenarios show traffic management either affecting more users or being more complex, or a combination of the two.

Assuming that consumers should have ‘transparent’ information about the traffic management envisaged in these scenarios, the question then arises as to what this information should be, and how it should be obtained. ‘Transparency’ can be considered to include meaningfulness, accuracy and comparability.

Meeting the requirement for transparency in these scenarios would involve more detailed information being given to consumers than at present. This information would be necessary to describe the packages on offer and to allow consumers to check that the delivered services accord with the package descriptions.

The representation of the effects of traffic management on performance is reasonably straightforward. This study has shown that common templates are feasible, given that traffic management interventions are broadly only of a few basic sorts. Applying the transparency criteria of meaningfulness, accuracy and comparability leads to three ‘representations’:

- a detailed policy, supplemented by time series data, which describes QoS
- a summary which presents the QoE of popular applications
- data for wizards (only necessary in the more complex scenarios).

There remains a need to translate between QoS and QoE in a consistent way. For example, the same threshold for the data rate necessary to support an HD video stream should be applied by all providers. The production of these criteria and thresholds could be undertaken by Ofcom or may possibly be devolved to an industry body such as the Broadband Stakeholder Group.

Even now, some tariffs make real-time status information desirable. Applying concepts of transparency to the traffic management scenarios could make such information essential in the future. Where usage limits are easily reached, then consumers will arguably need real time status information in order to regulate their usage patterns. If status information were provided it would fit comfortably alongside the provision of comprehensive real time information on QoS and any applied traffic management interventions.

By its nature, some traffic management is non-deterministic. For example, the effect of packet prioritisation is relative to other traffic and therefore depends on network conditions. In such cases, a policy alone cannot fully describe the QoS on offer: the level of traffic management and the effect on an individual connection may be impossible to predict with certainty. This facet of traffic management sets a limit to the achievable transparency.

Greater certainty for consumers when trying to compare packages could be provided through the use of time series data to show how much traffic management had been applied in the past. For example, it might be possible to state that on a typical weekday evening, P2P traffic was reduced to an average data rate of 5 Mbit/s on a connection capable of supporting 20 Mbit/s.

Providing such data is not straightforward. Though there is no single ideal way in which performance can be measured, on balance ISP-generated data will probably be the best long term solution. However, the architectures of ISP networks differ, and some ISPs are in a better position to gather such data from within their networks than others. For ISPs with centralised traffic management, imposing a requirement to measure and publish performance would be particularly costly. Accordingly we conclude that transparency will be enhanced by ISPs providing time series data from within their networks, and that the provision of such data should be encouraged. However, the case for mandating such data will need to be assessed carefully. The imposed costs would need to be evaluated relative to consumer benefits, and both sides of this trade-off will vary according to the traffic management scenario in operation. At present, the ‘fair use’ approach to traffic management is probably sufficiently minimal in its impact on most consumers that time series data would only be of benefit to a small proportion of consumers.

The Broadband Stakeholder Group has published its voluntary industry code of practice on traffic management transparency for broadband services. The code of practice includes a key facts indicator (KFI) which is similar in intent to the QoS Policy Form described in this report.

Appendix A Terms of Reference (from Ofcom ITQ)

Background
The internet is increasingly central to the lives of citizens, consumers and industry. It is a platform for the free and open exchange of information, views and opinions; it is a major and transformative medium for business and e-commerce, and increasingly a mechanism to deliver public services efficiently. As such it provides access to a growing range of content, applications and services which are available over fixed and wireless networks.

Many of these services, particularly those which contain video content, require high capacity networks to deliver them. Some networks are already experiencing congestion problems as consumers use ‘bandwidth hungry’ services. Even in the longer term, as next generation networks are deployed, there may continue to be congestion problems, particularly in wireless networks.

In response, network operators and internet service providers (ISPs) are making greater use of traffic management techniques. These can allow them to handle traffic more efficiently, to prioritise traffic by type, to charge for guaranteed bandwidth or to block or degrade the quality of certain content. It is important for ISPs to be able to clearly communicate the impact of their traffic management techniques to consumers, so that consumers can make informed choices about their broadband services.

Objective
This technical study will review the (current and likely future) traffic management techniques used by ISPs and the options for their characterisation, such as the Quality of Experience (QoE) provided to different types of online services. We have drawn up a list of questions to be answered by this study, divided into two categories depending on whether they relate to traffic management techniques or QoE:

Traffic Management	Assessment of QoE
• Which traffic management technologies and approaches are used that are under the direct control of ISPs and how are these likely to develop in coming years? • Which traffic management technologies and approaches are used that are not under the direct control of ISPs and how are these likely to develop in the coming years? • What role with Content Delivery Networks play? • Will these technologies be (or continue to be) network specific, or deployable across multiple network types? Will there be any changes to where traffic management technologies are deployed in the network? • What impact (positive and/or negative) does traffic management employed by a particular ISP have on: 1. Access to the internet by consumers; 2. Access to the internet by content providers; and 3. Other ISPs that may be part of an end-to-end exchange of data? • Are there any barriers to achieving end-to-end traffic management? Is co-ordination required between different traffic management technologies operated by interconnected ISPs? • Is there the potential for an *arms race*? For example, will it be possible for content owners to categorise their content in such a way to circumvent traffic	• What are the different options for measuring and characterising the impact of traffic management on consumer internet connections? • Which of these approaches (if any) would enable a meaningful repeatable (quantitative) comparison of performance across different ISP providers? • What are the relative merits of using a QoS and QoE approach for characterising connection performance and establishing a potential minimum level of required internet connection performance? • What effect would connection sharing, e.g. using a WiFi access point, have on these measures? • To what extent could these measurements of connection performance be communicated in a meaningful way to consumers? • How will measures of QoS and QoE stay relevant over time, i.e. can they be technology and application neutral? Will it be possible to update them as networks evolve?

management approaches? • Can traffic management cause further congestion in core networks? For example, if packets are dropped by ISP routers (rather than delayed), can this result in excessive retransmission by servers? Are traffic management techniques “polite” to the rest of the Internet?	

This study is intended to provide some of the technical input into Ofcom’s policy activity on Net Neutrality and Traffic Management. We therefore require this study to be completed within 3 months.

Deliverables

We suggest the following deliverable schedule:

1. A draft final report, at month 2
2. A final report, at month 3
3. A presentation to project team members, at month 2 or 3 as appropriate;

Appendix B Ofcom's technical questions

The technologies and approaches are covered in detail in Chapters 3 through to 5. Our key conclusions are bullet pointed below:

B.1 Which traffic management technologies and approaches are used that are under the direct control of ISPs and how are these likely to develop in coming years?

- The majority are under the direct control of ISPs
- The technologies and approaches are reasonably well established. ISPs use what they call "DPI boxes" to identify and categorise packets. Despite the name, Deep Packet Inspection is not necessarily used
- Subsequently packets may be treated differently according to their priority. At nodes packets may queue for longer or may be dropped according to their priority.
- Traffic may also be restricted in data rate
- No account is taken of packet priorities as signalled in the headers of incoming packets
- Packet priorities are generally stripped out on leaving the ISP's network
- In radio access networks, there are ways to prioritise within the radio layer. However it is technically difficult to integrate radio and IP management
- The technologies are not predicted to change. However there are general moves towards more distributed forms of traffic management (moving from the core towards the edge of networks) which would allow more finessed approaches.

B.2 Which traffic management technologies and approaches are used that are not under the direct control of ISPs and how are these likely to develop in the coming years?

- The ISPs we spoke to did not consider the technologies and approaches that are not under their control. Fundamentally ISPs restrict their interpretation of traffic management as helping them manage their network.
- Managed services and CDNs are not conventionally thought of as traffic management but they do have an effect on QoE. Both are set to become more prevalent. They may not even be visible from the ISP's perspective.

B.3 What role with Content Delivery Networks play?

- CDNs are in existence but are not perceived as being part of traffic management. ISPs did not highlight the role of CDNs and regarded them exclusively as a matter for content providers. The recent launch of BT Content Connect might be seen to counter this, but BT Wholesale is not an ISP.
- CDNs are marketed as offering end users a better experience. They can always be justified on network efficiency grounds.
- The launch of BT Content Connect as, in effect, a CDN within an operator's network is a new development. This should benefit consumers in helping to provide a better QoE for content-hungry applications, such as video streaming. Ofcom may wish to look into any impact this may have in the market for independent CDNs.

B.4 Will these technologies be (or continue to be) network specific, or deployable across multiple network types?

- In principle packet priorities could be retained across network types. In practice there is no interest in end to end traffic management within the internet. Managed services can cut across network types, but may or may not be regarded as services on the internet.
- We found that ISPs operating the three main types of network (DSL, cable and mobile) used broadly the same traffic management principles.

- There were differences in emphasis between network types – for example we found that DSL operators used all five of the intervention types described in section 3.3. Mobile operators used all except bandwidth allocation by traffic type (our reference B3), and cable operators used all except packet prioritisation by user identity (our reference A1)
- There are good reasons for these differences – see section 3.3.3 for details
- In view of the above findings we do not see any need for regulation of traffic management to apply differently to operators delivering service over different types of network

B.5 Will there be any changes to where traffic management technologies are deployed in the network?

- Most ISPs observe a trend to put traffic management closer to the edge and closer to users.
- This allows more finessed and individualised packages.
- But the extent of this trend does depend on cost. There needs to be a business benefit to justify the additional investment that would be needed.

B.6 What impact (positive and/or negative) does traffic management employed by a particular ISP have on: 1. Access to the internet by consumers; 2. Access to the internet by content providers; and 3. Other ISPs that may be part of an end-to-end exchange of data?

- In summary, most ISPs regard traffic management as a way of delivering an acceptable QoS for all, and penalising users who use more than their fair share.
- Very heavy users may be throttled back and/or removed
- There is no evidence in our interviews of ISPs using traffic management to affect access from content providers or other ISPs.
- Specifically:

1. Consumers that are particularly heavy users of certain types of traffic (e.g. P2P) can expect restrictions to be applied to their service to prevent undue impact on the service of other consumers. At peak times (typically evenings) users may see a reduction in overall internet speed. Some users may also experience a usage ‘cap’ which, if exceeded, will result in their service being restricted in speed for a period of time.

2. We have been told of an instance in the past where an ISP limited the bandwidth available for a particular type of traffic from a specific content provider. This was intended to provide consumers overall with a better service by limiting the coding rate of all data streams from this provider. This restriction has now been removed and we were not told of any similar restrictions currently in place in the UK. However, current regulation would not prevent such a restriction being applied again by an ISP.

3. ISPs told us that they did not take account of any packet prioritisation information in the headers of traffic entering their network. They apply their own traffic management policies to the data as it travels across their networks. They didn’t expect that ISPs receiving data from their networks took account of their prioritisation either. We may therefore conclude that the overall QoE enjoyed by a consumer will reflect the summation of the traffic management policies being applied by all the operators in the chain between the source and destination of a link. For each traffic type the result will reflect the most restrictive management applied to that traffic type by any operator in the chain.

B.7 Are there any barriers to achieving end-to-end traffic management? Is co-ordination required between different traffic management technologies operated by interconnected ISPs?

- Currently ISPs are not looking to achieve end to end traffic management.

- An end to end approach would effectively allow the ISP closest to the customer to detect and set the priority of each packet, and for this priority to be recognised by subsequent networks. In practice this is not done.
- There is also a bandwidth allocation mechanism built in to the standards that would allow end-to-end PVCs to be created. Such techniques are not used in the open internet.
- We are not sure if end-to-end traffic management is an achievable goal, in the short term at least. One ISP might apply a particular traffic management policy which is appropriate to the architecture and capacity of its network. The next ISP in the end-to-end chain might have a different architecture or capacity restriction (e.g. having a mobile access network) and could therefore be unable to honour the packet prioritisation or bandwidth allocation applied by the first ISP.

B.8 Is there the potential for an arms race? For example, will it be possible for content owners to categorise their content in such a way to circumvent traffic management approaches?

- In practice a variety of techniques are used by ISPs to identify content types. The vendors of packet inspection equipment send out updates to ensure their equipment retains the ability to detect content, even if attempts have been made to 'hide' it
- The ISPs did not tell us of any difficulties in identifying content
- There is no reason to believe that content identification is a losing battle – ISPs and vendors are confident that they can keep up with the development and use of new traffic types.

B.9 Can traffic management cause further congestion in core networks? For example, if packets are dropped by ISP routers (rather than delayed), can this result in excessive retransmission by servers? Are traffic management techniques "polite" to the rest of the Internet?

- Not all traffic management techniques, and not all protocols, will result in retransmission.
- However, reducing the priority of TCP/IP data can cause data loss, which in turn requires retransmission.
- The sender will reduce the data rate accordingly, but retransmission is still likely
- If retransmission emerges as a problem, then bandwidth restriction is an alternative which does not cause the same problems.

Appendix C Ofcom's QoE questions

The technologies and approaches are covered in detail in Chapters 6 through to 9. Our key conclusions are bullet pointed below:

C.1 What are the different options for measuring and characterising the impact of traffic management on consumer internet connections?

The options all involve a combination of a written policy, a set of measurements, and a way of representing the information. The main ones are:

- A QoS Policy Form giving full details
- A QoE Summary giving highlights and indicating applicability
- A wizard to help in complex choices
- A real time connection status dashboard.

C.2 Which of these approaches (if any) would enable a meaningful repeatable (quantitative) comparison of performance across different ISP providers?

All have their place, and it depends how far traffic management moves from its current 'fair use' paradigm. We identified five types of measurement approach:

1 User initiated

2 SamKnows box

3 Connection-specific probe

4 ISP-generated network statistics

5 Content provider.

Of these, 2, 3 and 4 are most relevant. The SamKnows approach (2) can be used for validation but may not be practical to give complete characterisation of all tariffs. In-network measurement (3&4) is constrained by architecture and by the sorts of information that can be provided from within a network.

C.3 What are the relative merits of using a QoS and QoE approach for characterising connection performance and establishing a potential minimum level of required internet connection performance?

Transparency involves three factors which cannot always be simultaneously satisfied. These are:

- accuracy
- meaningfulness
- comparability.

The QoS and QoE approaches are both necessary. QoS is more accurate and QoE is more meaningful. A method to convert QoS to QoE is needed, and this involves establishing minimum performance levels for applications. The levels must be consistent between ISPs.

C.4 What effect would connection sharing, e.g. using a WiFi access point, have on these measures?

ISPs deal with connections, not individual devices. Thus in general connection sharing is not well accommodated within the frameworks suggested. A particular issue is usage limits where one user of the connection may cross thresholds to the detriment of other users of that connection.

C.5 How will measures of QoS and QoE stay relevant over time, i.e. can they be technology and application neutral? Will it be possible to update them as networks evolve?

Both applications and traffic management will evolve. QoS is more likely to stay relevant than QoE. There is no problem in principle in updating as required. The QoS Policy Form should remain relevant over time as it is technology-neutral and application-neutral. The QoE Summary is necessarily application-specific (in order to be more meaningful) so it will need to be updated to reflect application developments.

Appendix D An ISO Model View of Traffic Management

D.1 Introduction

The ISO model is the reference model for IP networks so we have included a simple explanation here. In order to describe traffic management we will look at:

- The way in which IP networks work, based on the ISO model
- The way in which congestion and QoS are managed
- The way in which applications and protocols work
-

D.2 IP Traffic and the ISO Model

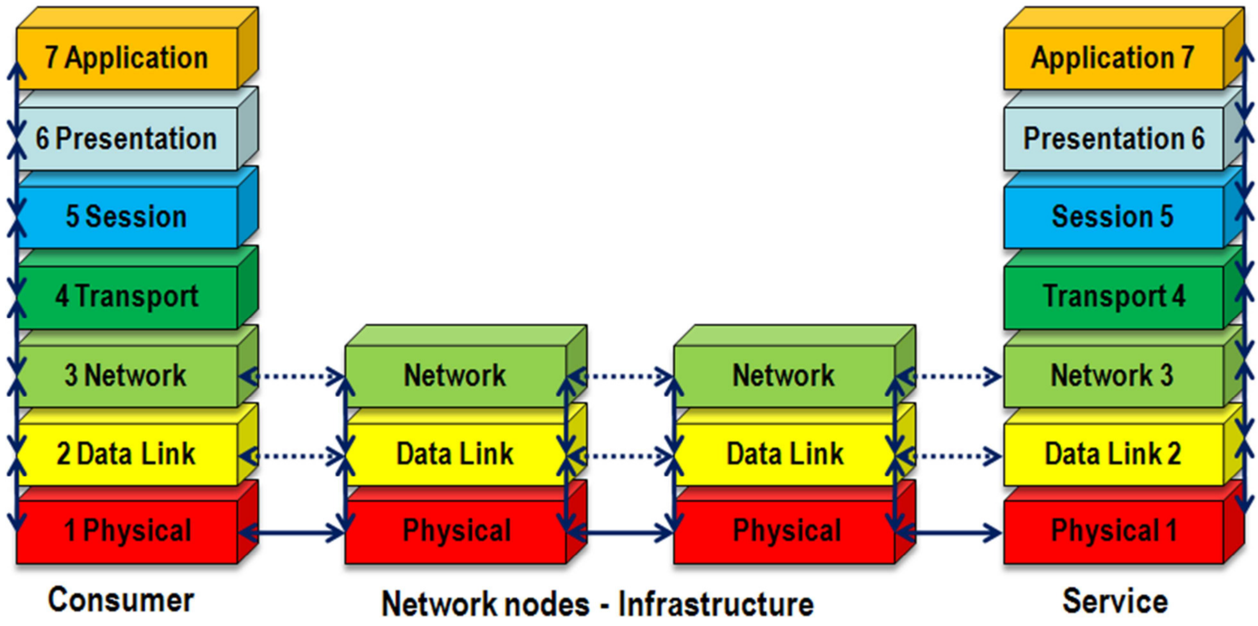


Figure 18: The ISO 7 Layer Communications Network Model

Figure 18 above summarises the arrangement of the well-known ISO 7 Layer model for communications systems.

With respect to the internet model, 5 main layers are used:

- Layer 1: Physical layer – defines the electrical / interface technology.
- Layer 2: Data link (Media Access Control – MAC) task is to take the Layer 1 transmission and convert this into a stream of error free scheduled data over the specific technology being deployed.
- Layer 3: Network layer – Forwards and routes packets based on a priority implemented in the 'Internet Protocol – IP'.
- Layer 4: Transport – Uses protocols such as TCP/IP and UDP to determine the format which the data is transmitted over the Network Layer.

- Layer 5-7: Application – Uses high level protocol such as HTTP, DNS etc. The application layer is not an application as such, it provides the ability to have a 'network transparent' common system for resource allocation and partitioning.

In the IP world, traffic is typically managed and controlled at Layer 3 and above. In contrast, in the telecoms world, resource allocation controls apply to Layers 1 and 2. This means that these networks operate differently under overload conditions.

In an IP network, congestion occurs at Layers 1, 2 and 3. Services are managed at Layer 3 and above, whilst congestion in Layers 1 and 2 is managed by schedulers running in Layer 2. These schedule data and prioritise access to the physical resources available in Layer 1.

D.3 Data Network Protocols and QoS

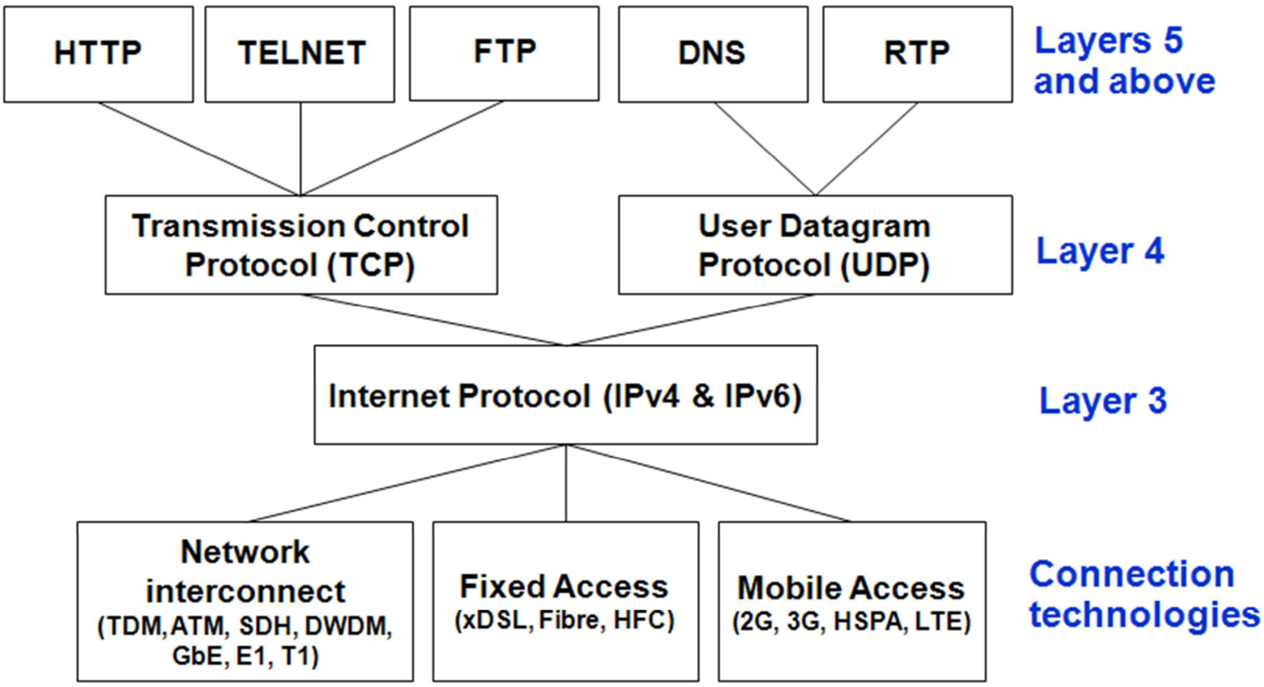


Figure 19: Data Network Protocols

Figure 19 above shows the main protocols used in data networks, and their relationships to each other.

In summary, HTTP is the protocol used for web browsing, Telnet is a bi-directional interactive text-oriented protocol and FTP is used for file transfer. These protocols all feature flow control, with the ability to re-transmit lost or damaged packets. TCP supports these protocols by including flow control to provide a reliable, ordered delivery of IP traffic over the network.

On the right hand side of the figure, DNS provides the internet 'phone book' and RTP is the standard protocol for sending audio and video content over IP networks. UDP delivers these protocols via a service that emphasises low latency over reliability of transmission.

TCP and UDP traffic have different transmission requirements and constraints. Ideally, they would be managed in different ways.

UDP provides a transport layer with the source and destination ports (the IP addresses are actually in the IP layer (Layer 3). UDP does not support flow control and retransmission and is typically used for streaming applications.

TCP is used to transmit data reliably. It is less suited for real time streaming due to the overhead required to implement the retransmission of lost packets.

IP is the basic protocol of the Internet which operates at Layer 3 and in isolation is unreliable for guaranteed delivery of data. The transport layer 4 deploys the techniques used to create the data packets which are delivered over IP layer 3, therefore layer 4 processing determines the QoS for layer 3 to process.

D.4 Application of QoS in the ISO Model

In many IP networks there is no connection between QoS management in the upper layers and physical provisioning in Layers 1 and 2. Operators rely on supplying adequate physical network capacity, and constraining traffic at Layer 3 or above, in order to avoid overload. Connection speed can be controlled via Layers 1 and 2, but systems at these levels have no knowledge of traffic type. This means that all forms of traffic running over the connection are impacted by any changes to capacity.

In the past, networks were usually designed for a single application, for example, voice or data. The evolution of networks to carry multiple different traffic types has led to the management of traffic, and hence the control of QoS, being distributed more widely across the ISO stack, depending on the traffic management objective, for example:

- IP QoS control – managed at Layer 3 and above
- Physical connection speed – managed at Layers 1 and 2, and driven by prioritisation from layer 3 and above.

Mobile operators, whose networks are often heavily loaded, face special challenges when more capacity is needed in the access network. They may not have access to additional radio spectrum, and moving to smaller cells to improve frequency re-use may require new base sites to be build and backhaul to be installed. They therefore make extensive use of resource scheduling in Layers 1 and 2 in order to make the best use of available capacity. This provides a more graceful degradation of performance under overload conditions than the alternative of allowing IP rules to battle inefficiently for capacity in a constrained radio access network.

However, the mobile community is also following the same evolution as fixed and currently implement a hybrid model across 2G – 3G – HSPA, where services such as voice take priority over data. Data services are generally treated equally with little QoS differentiation. As mobile networks evolve to handle more complex data services it is highly likely that the evolution experienced in the fixed world will be deployed in the mobile world by implementing more complex IP level differentiation.

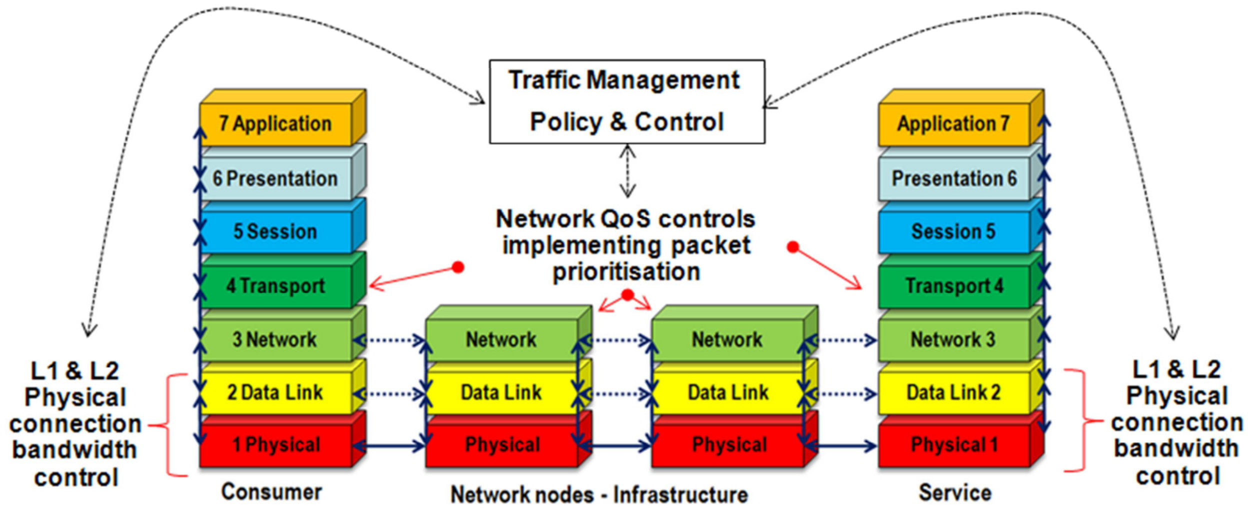


Figure 20: The ISO 7 model and traffic management controls

Figure 20 summarises the main elements of managed system where policy and control implements the ISP profile through managing bandwidth at layers 1 and 2, and QoS mechanisms (packet prioritisation) at layers 3 and 4.

Packet prioritisation can be conducted by identifying packets through techniques such as DPI and marking them accordingly and/or traffic shaping/policing for known traffic types within control of the network.

Operators believe they have the information and technology to understand where congestion is beginning to be a problem and to manage this. In the short term they do this by limiting the capacity available for certain traffic types, and in the longer term this is achieved by installing additional capacity.

Appendix E A Network Type View of Traffic Management

This appendix shows schematically how the principal components of traffic management can be implemented in different network types. The diagrams have been compiled from public sources describing networks internationally and are not intended to depict the networks in use in the UK.

E.1 DSL

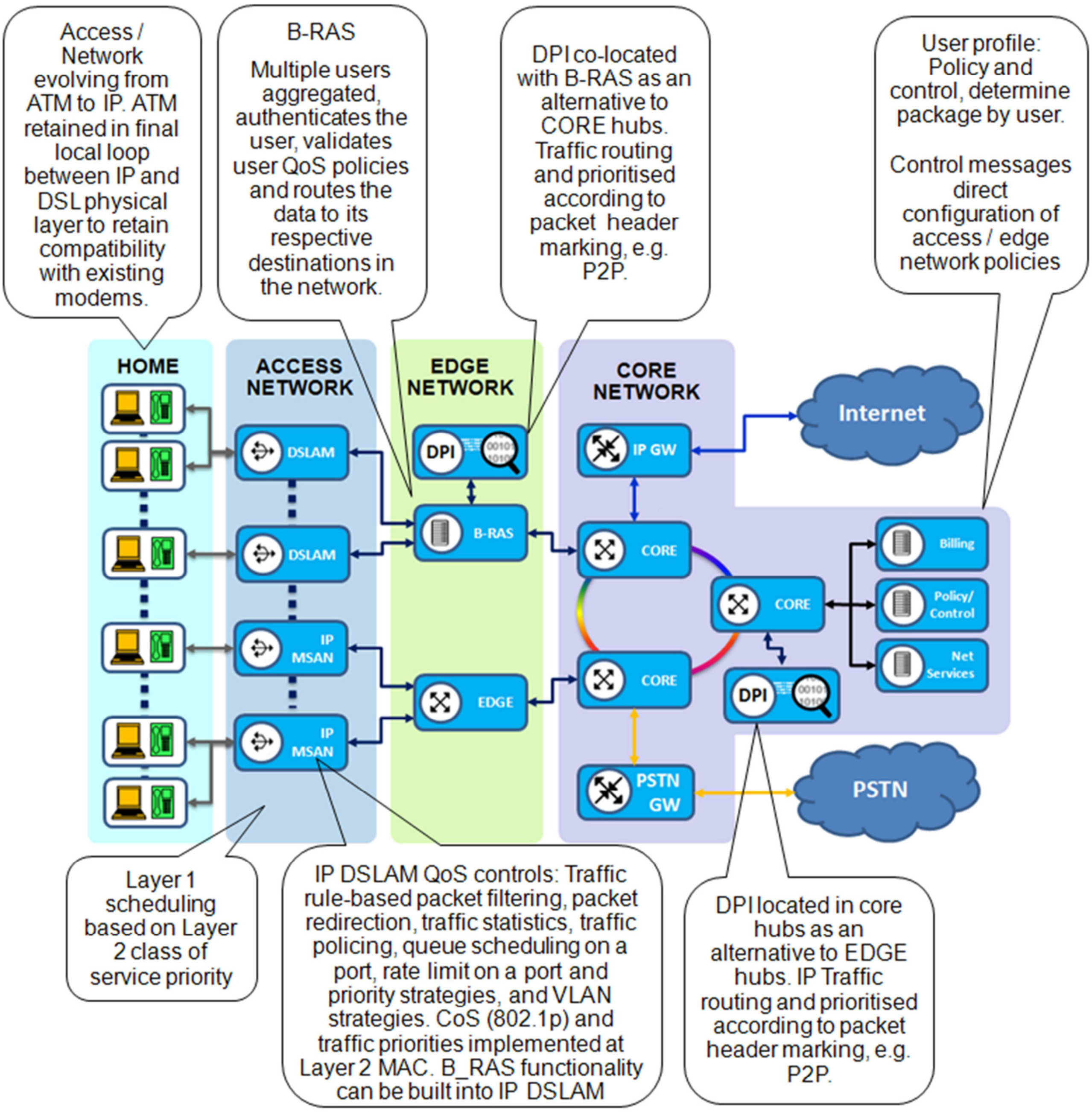


Figure 21: Typical Structure of ADSL ISP Connectivity

E.2 Cable

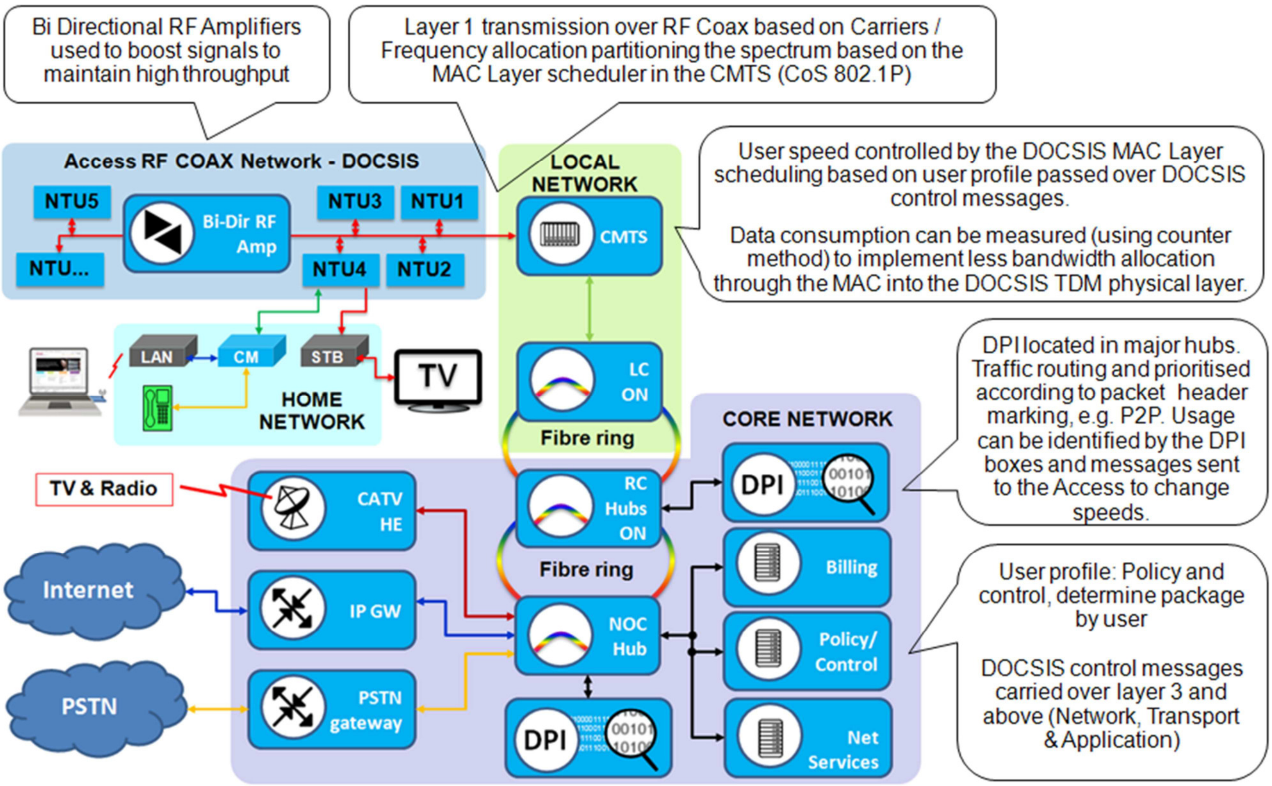


Figure 22: Typical Structure of Cable Access Network

Figure 22 shows an architecture implementation based on cable networks using Hybrid Fibre Coax (HFC) running the Data Over Cable Service Interface Specification (DOCSIS). This modulates data onto carriers which fit within the 8MHz PAL TV channel allocations of European CATV systems. DOCSIS supports a maximum downstream (to the consumer) data throughput of 55.62Mbit/s per channel, using up to 256-level QAM modulation. The upstream throughput is a maximum of 10.24 or 30.72Mbit/s, depending on DOCSIS version.

The Cable Modem Termination System (CMTS) is connected on the network side to redundant fibre rings, which in turn connect to the operator's fibre backbone network. Consumers have Cable Modems (CMs) which provide access to one or more DOCSIS channels. Contention is specified by the number of consumers connected to each channel.

Cable operators have good control over contention and congestion, and are less limited by available bandwidth in the access network than ADSL operators. They can change allocations of consumers to channels, or allocate additional channels to meet changing demand. The DOCSIS 3.0 specification provides greater upstream throughput per channel and allows channels to be configured together to provide downstream throughputs of 222.48Mb/s (4 channels) or 444.96Mbit/s (8 channels).

Cable operators can apply traffic management on a per-user basis in the access network and de-prioritisation of selected traffic types in the core network, typically at one of a small number of major backbone hubs.

As with other network technologies and topologies, DPI is used to identify traffic and packets are marked according to a traffic management strategy controlled centrally (Policy & Control). Bandwidth allocation will be implemented by using the information provided by the packet marking and the user profiles instructions also derived from the traffic management strategy implemented through the Policy and Control centre.

E.3 Mobile

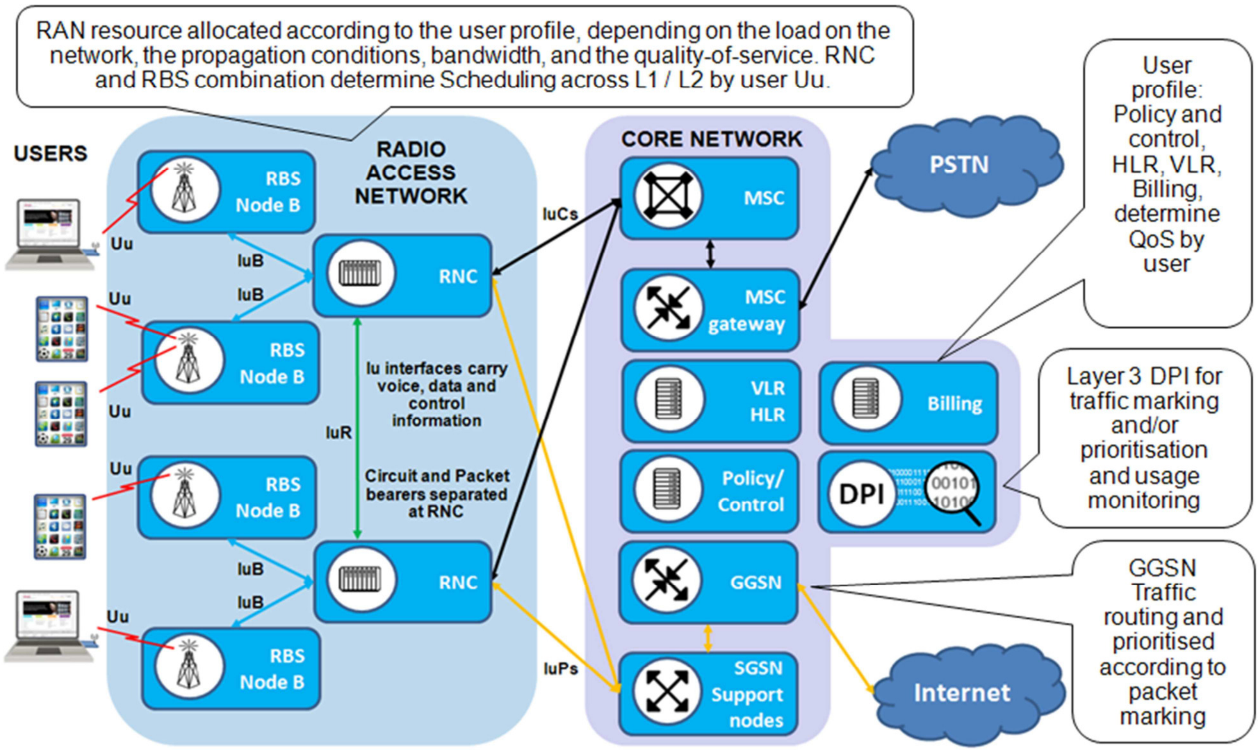


Figure 23: Typical WCDMA Mobile Network Architecture

Figure 23 above shows the architecture of a typical WCDMA (3G) mobile network. The main determinant of QoS in mobile networks is the available radio bandwidth. This will be constrained by:

- The amount of radio spectrum available to the mobile operator
- The intensity of installed infrastructure (i.e. how large the cells are)
- The transmission protocol used (number of bits/Hz carried)

In 2.5G (GPRS) networks, operators will typically pre-allocate a small proportion of radio network capacity to GPRS data. This will be shared amongst data users on each cell. In 3G networks, voice calls are often given priority over data, which means that data capacity will vary with voice loading on each cell. On some networks voice traffic may restrict data capacity to the point where data users are offloaded onto a GPRS cell in the same area.

Typical methods of traffic management employed by mobile operators include use of 'traffic consumed' counters, which measure the amount of data generated and consumed by users. These reside in the core network (e.g. on the GGSN) and communicate with RNCs which instruct the RAN to drop data rates when consumers approach or exceed pre-defined limits.

Mobile operators also have the ability to limit usage of traffic types (e.g. P2P) within their core network and to limit data entering their networks.

DPI located in the core is used to identify traffic and packets are marked according to a traffic management strategy controlled centrally (Policy & Control). Bandwidth allocation will be implemented in the RAN by using the information provided by the packet marking and the user profiles instructions also derived from the traffic management strategy implemented through the Policy and Control centre.

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄四

Ofcom: Measuring mobile voice and data quality of experience

Measuring mobile voice and data quality of experience

Call for Input

Publication date: 23 January 2013

Closing Date for Responses: 1 April 2013

Contents

Section 1

Introduction

Purpose of this document

1.1 Consumers and citizens are growing increasingly dependent on mobile networks to make phone calls and access data services. The performance of these networks can vary between operators, by location and time of day and may not always meet the expectations of consumers. In this document we use the phrase 'quality of experience' ('QoE') to describe the technical performance[1] of the services delivered to consumers.

1.2 As an economic and competition regulator Ofcom primarily relies on market mechanisms to drive performance improvements in networks (thereby improving QoE). However, the market can only operate effectively when consumers are able to compare the quality of the services on offer and this in-turn requires the availability of accurate and comparable QoE information.

1.3 For fixed broadband services, Ofcom has, for several years, collected information on broadband speeds. This information has enabled consumers to improve their purchasing decisions, and appears to have driven improvements in service quality by operators. In this Call for Input, we wish to explore whether there is similar information that we might provide in the mobile arena. Specifically, we want to identify what network and/or service performance information Ofcom could gather which accurately reflects the consumer QoE and which we could publish in a way that would assist consumers in making informed choices about the mobile service they purchase.

1.4 We are seeking views from all interested stakeholders on:

- What information would be valuable to consumers when purchasing mobile services;
- What data would be required to produce this consumer information, and
- How we could best collect it.

1.5 Alongside this Call for Input, where appropriate, we will seek to engage directly with the mobile network operators to ensure that any other relevant information can be taken into account.

Ofcom's role and market context

1.6 As the regulator for the communications sector, our principal duty is to further the interests of consumers and citizens in relation to communications matters.[2]

1.7 We also have general duties to consider, amongst other things:

---

[1] By technical performance, we are referring to the operation of the network and services (i.e. the coverage, speed, capacity and reliability) rather than customer service related aspects of a mobile service such as billing, call centres and sales.
[2] Section 3 of the Communications Act 2003

- the desirability of encouraging the availability and use of telecommunications services throughout the UK;
- the desirability of promoting competition in relevant markets; and
- the interests of consumers in respect of choice, price, quality of service and value for money

1.8 Our 2012 Communications Market Report[3] found that 94% of UK adults use a mobile phone and through the rise in smartphones and tablets, consumers increasingly rely on mobile networks to provide a mobile data connection as well as a voice and text service. Our Infrastructure Report 2012 Update found that the capacity of the UK's communications infrastructure is changing quickly. This is in response to a rapid increase in consumers' take-up and use of communications services and the resulting investment by operators.[4] Data via mobile devices more than doubled between 2011 and 2012.

1.9 In a time of such rapid change, it is all the more important that consumers have access to timely and accurate information on the quality of services available in the market.

1.10 There is also increasing Government interest in ensuring that UK consumers are able to access mobile services which meet all their needs and expectations. The UK already has a high level of mobile signal coverage; based on figures derived from operator predicted coverage models[5] we estimate that 99.7% of UK premises receive an outdoor 2G signal[6] from at least one operator and 93.6% of premises receive a signal from all operators. The government has announced an initiative aimed at extending existing mobile voice coverage further still through its mobile infrastructure project[7] in recognition of the importance of mobile services to citizens and the economy.

1.11 Expressed as a percentage of geographical area, coverage figures are lower, because mobile masts are more commonly installed near centres of population. At present, we estimate 12.8% of the UK landmass is not covered by any 2G signal. Extending coverage to these more remote areas is challenging because the costs of doing so are high relative to the potential revenues.

1.12 Coverage of 3G services is lower than 2G, but improving, with a recently increased (outdoor) coverage obligation placed on operators to reach 90% of UK premises due to be met by June 2013. Furthermore, Vodafone and Telefónica (O2) have recently begun to share their radio access networks, a development which has the potential to reduce materially the number of partial not spots[8] in the UK.

1.13 Additionally, we will require that one of the 800MHz licensees in the current 4G spectrum auction deliver a high speed, mobile data service indoors to 98% of UK premises and 95% of premises in each of the Nations by the end of 2017. We

[3] Figure 5.55 - http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr12/CMR_UK_2012.pdf
[4] http://stakeholders.ofcom.org.uk/binaries/research/telecoms-research/infrastructure-report/Infrastructure-report2012.pdf
[5] Each mobile operator uses planning tools to predict signal strength in different areas.However, as with any planning tool, these predictions are subject to an error of margin and do not necessarily account for all the factors that can affect the quality of a mobile voice call or data session.
[6] i.e. the signal is predicted to be sufficiently strong to make and sustain a call while outside.
[7] http://www.culture.gov.uk/what_we_do/telecommunications_and_online/8757.aspx
[8] We define partial notspots as areas with coverage from 1 or more MNOs, but not all MNOs

anticipate that the resulting outdoor coverage will be materially higher than the indoor requirements.

1.14 These developments are expected to bring about significant mobile coverage improvements for consumers over the next few years, but it will be important to keep track of how these improvements progress.

Coverage vs quality

1.15 The reach and coverage of mobile signals is just one part of delivering a mobile service. Consumers sometimes experience misalignment between predicted coverage and their day-to-day experiences of using their mobile phones. Concerns about consumers' ability to make and receive calls or use the internet on their mobile phones have been expressed directly by consumers. These complaints have been received by Ofcom[9] as well as by MPs, our own Advisory Committees and through reports in the media.

1.16 We know that a number of factors can affect consumers' QoE. This can sometimes be the result of localised low signal quality caused, for example, by 'signal shadowing' by buildings, how many people are using the network in a particular area or the performance of a particular handset. While some of these factors may be outside the control of the mobile operators (for example, handset performance), the technical performance of each operator's network does represent the key differentiator in the consumer QoE delivered by different networks.

1.17 There are signs mobile network operators (MNOs) are increasingly competing on issues of service quality and reliability, particularly as they ready themselves to offer new 4G services (subject to the outcome of the 4G auction in 2013). For example, Orange provides a Network Performance Promise[10] which compensates consumers for dropped calls, and Vodafone recently referred to the depth of its network as 'deep pan' pizza[11] - able to provide service deeper into buildings.

1.18 The extent to which operators are incentivised to improve their consumers' QoE is in part related to the competitive advantage that they can gain from offering the higher quality. However, unless consumers are able to take the QoE offered by different operators into account when making purchasing decisions, there is less incentive for operators to invest in improving it.

4G Auction

1.19 2013 is likely to see rapid change in the mobile market as the 4G auction concludes and operators roll out 4G networks. We are currently considering the potential scope of initial research into 4G QoE with a likely initial focus on connection speeds and coverage.. The lessons from any 4G-specific research will be combined with the feedback we receive to this Call for Input and will inform Ofcom's longer term research objectives in the area of mobile QoE across 2G, 3G and 4G networks."

[9] We estimate that mobile coverage and quality issues represent approximately 5% of all mobile complaints received by Ofcom

[10] http://help.orange.co.uk/orangeuk/support/personal/480099/2

[11] https://www.vodafone.co.uk/our-network-and-coverage/what-makes-a-great-network/index.htm

Structure of this document

1.20 In section 2, we examine the results of recent consumer research we commissioned to update our understanding of consumers' QoE. We then consider what information would be most useful to consumers in making purchasing decisions.

1.21 In section 3 we set out the type of data that would be necessary to produce the relevant consumer information and how this data could be collected.

Next steps

1.22 We welcome feedback from all stakeholders on this Call for Input. We intend to review responses in April before deciding on how best to proceed.

1.23 We are particularly keen to get the views of stakeholders representing the needs of consumers in different parts of the UK to ensure we have a clear view of the information that consumers would find useful when purchasing mobile services. If there is sufficient interest from stakeholders, we propose to host a workshop in March to facilitate the development of ideas and options. To register your interest in a workshop, please contact us by 15 February 2013.

1.24 Details on how to contact us and how to respond to this Call for Input are provided in Annex 1.

Section 2

Understanding the consumer experience

How and why we gather information

2.1 Ofcom has a statutory duty under the Communications Act 2003 ("the Act") to collect and publish certain types of data.

2.2 Under section 14 of the Act we are required to make arrangements to find out about the experiences of consumers using electronic communications services and the way they are provided, and we do this by carrying out research into their experiences of these services. Under section 15 of the Act we have a duty to publish the results of our research and to take account of it in carrying out our functions; for example we do this through our annual Communications Market Reports, and our Consumer Experience Reports.

2.3 We may also inform our thinking by conducting economic or technical research, and/or by engaging with consumer groups and industry. We gather data directly from industry on a regular basis.

2.4 In addition, and in keeping with our duty to consider the interests of consumers and citizens, we also seek to provide advice and information to help consumers make better and more informed decisions about their telecommunications services. Consumer information plays a critical role in ensuring competitive communications markets, and we noted this in our Customer Service Satisfaction report in December 2012[12]. A lack of information may lead consumers to make poor purchasing decisions, or inhibit them from switching provider. If such information is not readily available or is presented in a complex way, there may be a case for Ofcom to intervene to address issues in the interests of and to protect consumers.

Research findings on the consumer QoE

Methodology

2.5 To update our understanding of consumers' experience of using their mobile phone and to help us keep track of improvements in consumers' QoE, we carried out a consumer survey in November 2012. We expect to conduct research of this kind annually. This research helps us understand whether and to what extent mobile phone reception issues affect consumers and, if so, what types of problems are most prevalent and of most concern.

2.6 Our research also sought to examine whether there are differences in consumers' QoE in urban and rural areas and in each of the Nations. We have published a report of our findings alongside this Call for Input[13] and provide highlights of the results relevant to consumers' QoE below.

[12] http://media.ofcom.org.uk/2012/12/04/latest-customer-service-satisfaction-levels-revealed-2/
[13] Mobile Coverage Report: http://stakeholders.ofcom.org.uk/binaries/consultations/mobile-voice-data-experience/annexes/usage.pdf

Quality of mobile experience is important to consumers

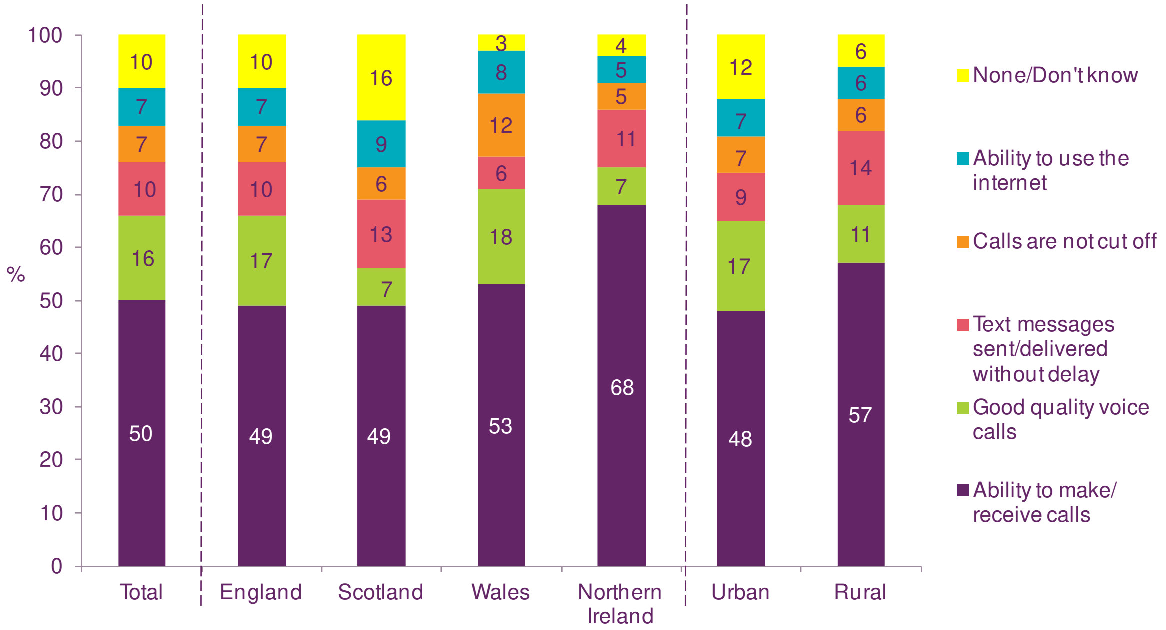
2.7 The ability to make or receive calls or texts is consistently selected as the most important feature when thinking about their mobile operator, followed closely by the price of the service (38% and 34% respectively for the UK as a whole – Figure 1). In Wales and Northern Ireland, and in rural areas, the ability to make or receive calls is particularly important when selecting an operator. Mobile users in Wales (53%) and Northern Ireland (47%) are significantly more likely than those in England (37%) or Scotland (34%) to say that this is the most important factor when choosing a provider and users living in rural areas are significantly more likely than those in urban areas to say this (45% v. 37%). This may be a reflection of a poorer consumer experience in those locations, although we do not have sufficient information to determine this for certain.

Figure 1: Most important element when considering mobile provider, by nation and urban/rural

Source: Kantar Media omnibus, (14th – 20th November 2012)
Base: All who use a mobile phone (N=2136/1743/195/95/103/1757/379)
Q.10 And which is the ... important to you when thinking about your mobile operator? Most important.

2.8 Mobile users were also asked about the importance of the ability to make or receive calls alongside other aspects of mobile reception (Figure 2). The ability to make and receive calls remains the most important for mobile users when thinking about their mobile provider by a considerable margin (50% of UK mobile users). This is particularly so for those in Northern Ireland (68%). Quality of voice calls is the next most frequently cited aspect among UK users, with 16% saying this is most important.

Figure 2: Importance of different elements of mobile reception, by nation and urban/rural locations



Source: Kantar Media omnibus, (14th – 20th November 2012)
Base: All who use a mobile phone (N=2136/1743/195/95/103/1757/379
Q15: Which of the following aspects is most important to you when thinking about your mobile operator?

The majority of consumers are satisfied with their mobile service

2.9 Our survey found that in the UK as a whole overall satisfaction with mobile providers was 81%[14] (6% reported they are dissatisfied). There are no differences in levels of overall satisfaction by urban or rural location or by nation.

2.10 When considering mobile functions and services, illustrated in figure 3, the highest level of satisfaction is with the handset, with 78% of users either somewhat or very satisfied. This is followed by satisfaction with the ability to make or receive calls or text messages (74%).

2.11 The number of people satisfied with the speed or reliability of internet is lower, with 47% either somewhat or very satisfied. However, when filtered by those who use the internet on their mobile phone the proportion saying they are either somewhat or very satisfied increases to 70%.

[14] Another recent Ofcom survey found that overall satisfaction with mobile phone services was higher than this http://stakeholders.ofcom.org.uk/market-data-research/market-data/consumer-experience-reports/consumer-experience/ at 89%. The difference may be explained by question ordering; in our November 2012 survey the question about overall satisfaction was positioned immediately after several questions about individual aspects of service, which may have has some influence over what the respondent was considering when rating the 'overall' service.

Figure 3: Satisfaction with different elements of mobile phone functions or services (percent)

	Very dissatisfied	Somewhat dissatisfied	Neither satisfied /dissatisfied	Somewhat satisfied	Very satisfied
Ability to make/receive voice calls/ text messages	5	9	13	35	39
Cost of service/ contract	2	7	22	34	35
Internet reliability	6	6	45	27	20
Handset	2	3	17	33	45
Customer service	3	5	28	33	32

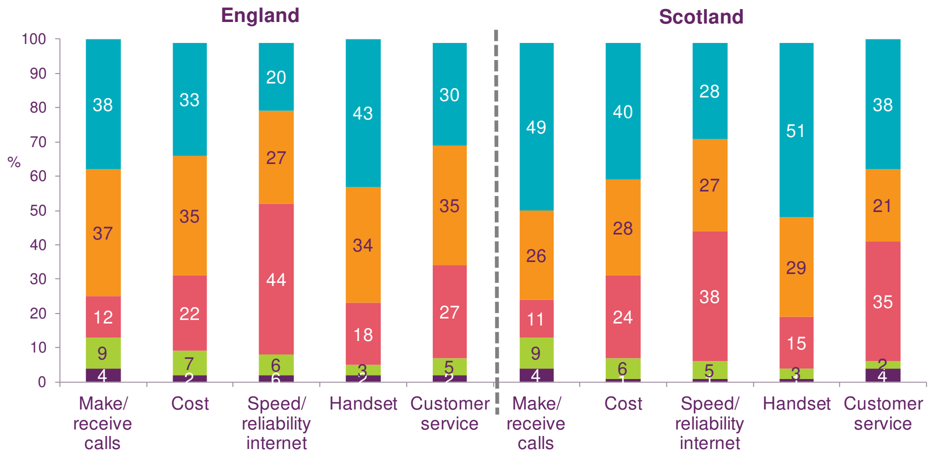
Source: Kantar Media omnibus, (14th – 20th November 2012)
Base: All who use a mobile phone (N=2136)
Q.11: Thinking about these functions, how satisfied do you feel with each in relation to your mobile phone and mobile services with ...?

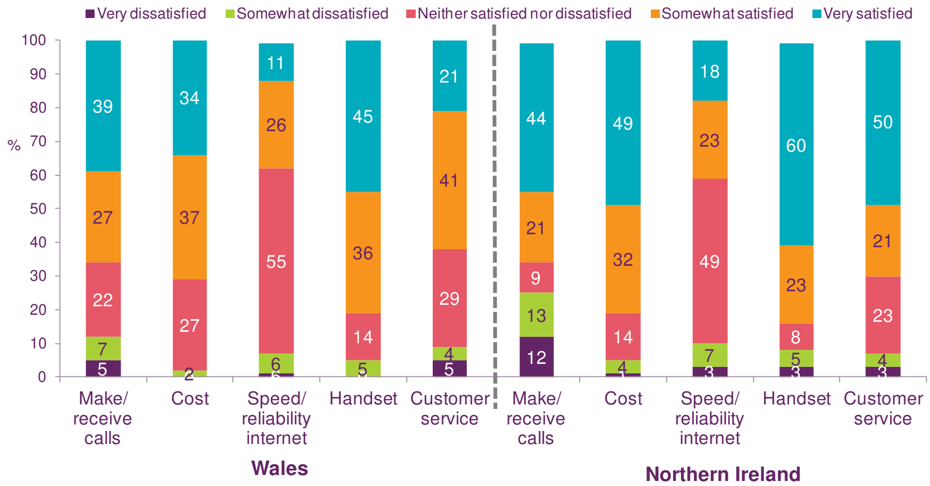
2.12 There are also some differences between the nations, shown in Figure 4, below.

2.13 Users in Scotland are the most satisfied with speed or reliability of the internet (55%), with those in Wales being the least likely to be satisfied (37%).

2.14 Users in Northern Ireland are the most likely to report dissatisfaction with their ability to make or receive calls or text messages. This is almost double the proportion who are dissatisfied with this aspect of service in England (13%).

Figure 4: Satisfaction with mobile phone functions or services, by nation





Source: Kantar Media omnibus, (14th – 20th November 2012)
Base: All who use a mobile phone (N=1743/195/95/103)
Q.11: Thinking about these functions, how satisfied do you feel with each in relation to your mobile phone and mobile services with ...?

The majority of consumers are satisfied with different aspects of mobile reception

2.15 We also asked respondents about their satisfaction with various aspects of mobile reception (Figure 5). The element with the highest level of satisfaction among UK mobile users is good quality voice calls (78%). This is followed by calls not getting cut off (75%), mobile reception (74%) and text messages sent/delivered without delay (also 74%). Figure 5 shows that just under half (48%) said that they were satisfied with using the internet, though this rises to 71% when filtered to include only those who use the internet on their mobile.

Figure 5: Satisfaction with different aspects of mobile reception (percent)

	Very dissatisfied	Somewhat dissatisfied	Neither satisfied /dissatisfied	Somewhat satisfied	Very satisfied
Ability to make/receive voice calls	3	8	15	38	36
Good quality voice calls	2	6	14	39	39
Texts sent / no delay	2	5	19	35	39
Calls not cut off	2	7	15	32	43
Ability to use internet	2	5	44	26	22

Source: Kantar Media omnibus, (14th – 20th November 2012)
Base: All who use a mobile phone (N=2136)
Q16: How satisfied do you feel with each in relation to your mobile phone reception in the UK with ...?

Satisfaction with aspects of mobile reception is lower in rural areas and of the four UK nations is lowest in Northern Ireland

2.16 There are some differences between urban and rural users. Rural users are more likely than those in urban areas to be very dissatisfied with their ability to make or receive calls (6% vs. 3%). Figure 6, below, shows differences in satisfaction between the nations. Users in Northern Ireland appear to be the least satisfied with their ability to make or receive calls (18% report that they are dissatisfied). They also have the highest levels of dissatisfaction with the quality of voice calls (12%) and calls not getting cut off (12%).

Figure 6: Satisfaction with different aspects of mobile reception, by nation

Source: Kantar Media omnibus, (14th – 20th November 2012)
Base: All who use a mobile phone (N=1743/195/95/103)
Q16: How satisfied do you feel with each in relation to your mobile phone reception in the UK with ...?

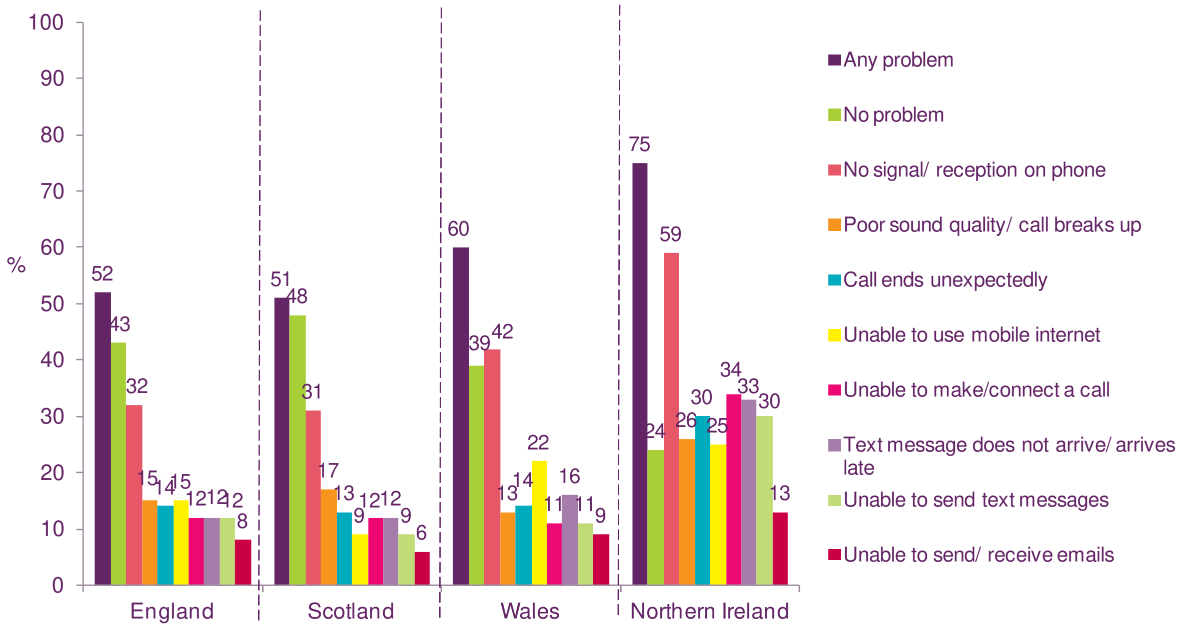
Over half of UK mobile users say they have experienced problems with reception – this rises to six in ten in Wales and three-quarters in Northern Ireland.

2.17 Just over half (53%) of UK mobile users have ever experienced any issues with mobile reception with 12% experiencing four or more problems.

2.18 The most common problem is having no signal/reception on phone (34%), followed by poor sound quality/sound breaks up, call ending unexpectedly and being unable to use the mobile internet (all 15%), being unable to make/connect a call even though the phone shows “bars” present and text messages not arriving or arriving late (both 13%), being unable to send text messages (12%), and being unable to send or receive emails (8%).

2.19 There are no statistically significant differences between users living in rural and urban locations. However, among the nations (Figure 7), mobile users in Northern Ireland are significantly more likely than those in England, Scotland and Wales ever to experience a problem (75% vs. 52%, 51% and 60%). Around a third (32%) of people in Northern Ireland say they have experienced four or more of these problems.

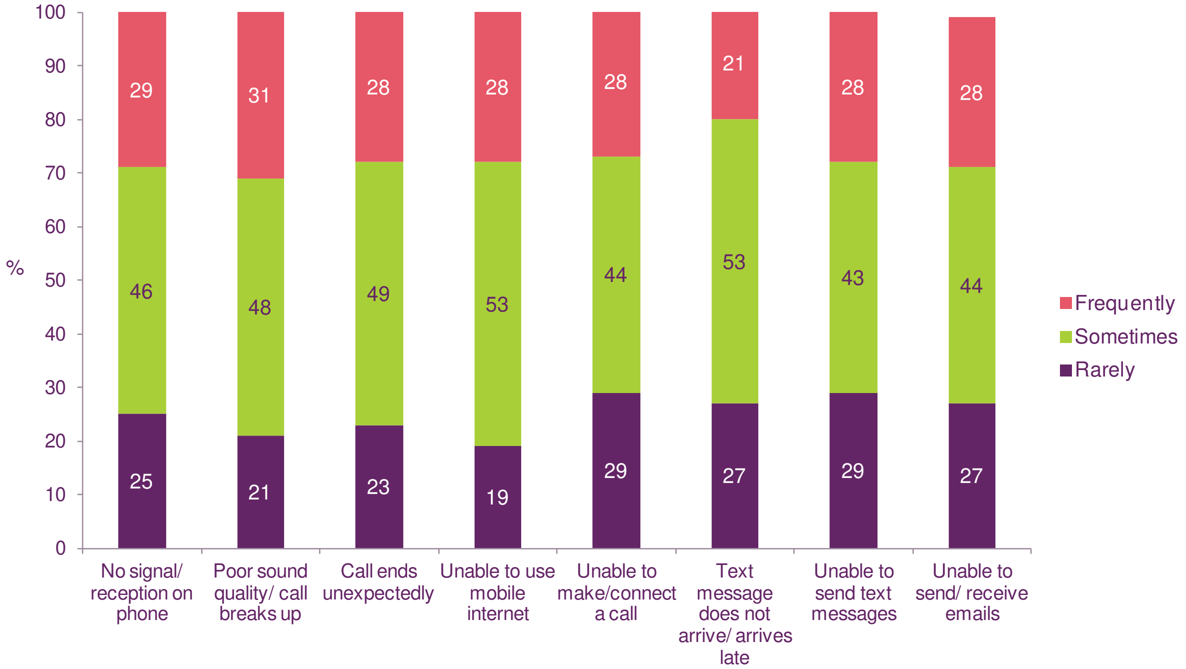
Figure 7: Mobile phone users who have ever experienced problems with reception, by nation



Source: Kantar Media omnibus, (14th – 20th November 2012)
Base: All who use a mobile phone (N=2136/1743/195/95/103)
Q13: Thinking about your mobile reception with ... in the UK, do you ever experience any of the following issues? Some respondents answered ‘don't know’ so the total of those reporting any problem and those reporting no problem, does not add to 100%.

2.20 Figure 8 shows the frequency with which problems are experienced. Having no signal or reception on the phone is experienced most frequently, with 10% of mobile users saying they experience this frequently. The frequency across the other problems we asked about is lower, with between two and five per cent of mobile users frequently experiencing these.

Figure 8: Frequency of mobile reception problems, all mobile users



Source: Kantar Media omnibus, (14th – 20th November 2012)
Base: All who use a mobile phone (N=2136)
Q14: And how often do you experience these issues with mobile reception?

2.21 There are no statistically significant differences in frequency of reported problems between users in urban and rural areas. Among the nations the only difference is that mobile users in Northern Ireland are more likely than those in Scotland to frequently have no signal/reception on the phone (24% vs. 5%)

Conclusions

2.22 These results show us that, as one might expect, being able to use your phone to make and receive calls is very important to consumers. Many mobile users say that they experience no problems at all[15] and the majority of UK mobile users are satisfied with their mobile service overall (81%). However, a significant minority of consumers (most notably in Northern Ireland) experience a range of recurring problems when they try to use their mobile phones (see paragraph 2.20 and figure 7 above).

2.23 The mobile reception issue consumers are most dissatisfied about is their ability to make and receive calls (12% in England, 8% in Scotland, and 11% in Wales are dissatisfied), with those in Northern Ireland most likely to be dissatisfied (18%). No mobile signal or reception (in order to make or receive calls or texts) is also the most common problem consumers say they have experienced (more than twice as many people have ever experienced that problem compared to each of the other problems we asked about – see Figure 7 and paragraph 2.20 above).

2.24 The research clearly indicates that some consumers are not wholly satisfied with the QoE of their mobile service but consider that various aspects of QoE are important to them. This suggests that if appropriate information were available to allow consumers

[15] 43% in England, 48% in Scotland, 39% in Wales and 24% in Northern Ireland stated they had never experienced a problem.

to compare operator's QoE then they would use this information when purchasing mobile services and select packages / providers that better suited their needs.

2.25 We have considered similar matters in the past in the 2010 Quality of Service research report[16]. This research found that consumers particularly valued information on price and network quality of service. Ofcom's accreditation scheme for price comparison websites encourages clear and accurate consumer information on price. As a result of the research report, we also considered the information provided to consumers on fixed line broadband speeds. This initiative has brought about improvements in both the information provided by broadband providers as well as the speeds consumers can expect from their broadband service.

2.26 Adopting a similar approach to the technical performance of mobile networks has the potential to bring about further improvements for those who have a poor experience with mobile performance. Ofcom has a number of existing publications where this information could be published, including (but not limited to) the Communications Market Report, Infrastructure Report and Consumer Experience Report.

2.27 Publishing information on QoE will enable consumers to make better informed purchasing decisions and drive competition between operators. This in turn will result in improved network performance for the benefit of consumers.

[16] http://stakeholders.ofcom.org.uk/consultations/topcomm/qos-report/

Section 3

Technical performance metrics

3.1 In the previous section we set out why publishing information on mobile QoE will benefit consumers. This section sets out our initial thoughts on the type and granularity of information that might of use to consumers, what data we would need to collect to produce the information and we then consider some of the potential sources of this data.

What information do consumers need?

3.2 Consumers use mobile services in different ways and this may influence the importance they place on different aspects of QoE. For example, a sales representative may value the ability to make and receive calls across the UK motorway network, whilst a teenager may be more interested in the performance of text and internet access in her home town.

3.3 Although it is likely that QoE information will need to be presented and tailored to meet the needs of different consumers groups, there are a number of core characteristics to QoE information that we think will be universally applicable. These include:

- **Operator specific information** which allows the performance of rival networks to be compared. There may be merit in publishing information on both MNOs and MVNOs to identify whether consumers receive differing levels of service from operators using the same network.

- **Granular geographic information.** Consumers will wish to know how mobile services perform in the places they wish to use them. Mobile operators already provide on-line coverage checkers which set out information typically down to a 100mx100m grid.

- **The consumer 'use case'**. For any given location, QoE can also vary depending on whether the consumer is indoors or outdoors and whether they are on the move (whether on foot, in a motor vehicle or on a train).

- **Network performance by time of day and day of week.** Our experience from measuring fixed broadband is that performance can degrade at peak usage times. For mobile broadband (and potentially voice calls) similar effects may be present and consumers may wish to know which operator provides the highest network capacity in the areas they wish to use their mobile service.

3.4 For those consumers who particularly value the quality of voice and text services, we consider that there are a number of important QoE metrics:

- Locations in which they are able to reliably make and receive a call under different use cases

- Probability that a call will complete successfully

- Probability that a call will not be blocked or dropped

- Clarity of the call
- Time for a SMS text message to be delivered

3.5 For mobile data services, the following information may be of use to consumers:

- Probability that an internet connection can be established
- Speed, stability and responsiveness of applications and data transfers.
    - For a mobile device, this would typically include activities such as browsing web pages, using online maps, accessing location services, streaming video, using voice over IP services or downloading music.
    - For a laptop with a broadband dongle, the range of activities is likely to be similar to normal fixed broadband.

3.6 For each of the metrics above, given that performance may vary in different parts of the UK, under different use cases and at different times, there may be merit in providing information at a local level, by use case and by time of day.

3.7 As a result of the differing information needs of different consumers groups, we recognise that it may be necessary to aggregate and simplify information when communicating to these different groups. We envisage that this could be achieved by making a more granular 'superset' of information/data publically available which can be aggregated ahead of publication to specific groups, potentially by third parties such as comparison websites.

3.8 We welcome views on whether there is additional or alternative information that would be useful for consumers to that set out above. We also welcome views on the granularity (geographic and time) which would be most appropriate to form the superset of information that could be subsequently tailored for different consumer groups.

3.9 It is clearly important that published information is accurate and up-to-date if the market for mobile service is to operate effectively. Given the rapid rate of change in the market (particularly with the advent of 4G services) regular updates to the information will be required. We currently collect coverage data annually, but we welcome views on how often information should be refreshed to ensure that mobile markets work effectively.

Proxies for QoE metrics

3.10 The QoE metrics that we have set out above have been chosen because we consider that they most closely reflect what consumers' consider important about their mobile services – although we welcome alternative views.

3.11 The data for some of these metrics may be readily available from MNO operation systems at the granularity required – in which case the data can be directly converted to the information provided to consumers. For example, MNOs may already collect data on dropped calls.

3.12 Where suitable data is not available to produce particular QoE metrics, then it may be necessary to use proxies. For example, data on actual network coverage may not be readily available from the MNO operational systems, but can be estimated using planning models (potentially validated with field measurement). This is the approach we currently use when reporting network coverage – predicted signal strengths produced from MNO planning tools are used as a proxy of actual coverage.

3.13 Although proxies may be less accurate than actual data (and hence may be of less use to consumers who wish to know whether they can use a specific service in a specific location and use case) if defined correctly they can be used by consumers to make comparisons between MNOs.

3.14 Under our Infrastructure Reporting duty we are required to report on the state of communications networks in the UK as well as the services they carry. As such, the network data we collect for the Infrastructure Report may also be useful for deriving QoE proxies.

Predicted vs. actual performance

3.15 When collecting data on performance of networks and services, two broad approaches can be considered:

- **Predicted performance**: The network and service performance the operator expects to deliver based on coverage and capacity planning tools. This is the 'designed performance'. This information would be gathered from MNOs.

- **Actual performance**: Data on the actual performance of the network – such as signal strength, dropped calls and speed or latency of mobile broadband experienced by end users. This data could be collected from a third party commissioned by Ofcom and/or directly from MNOs.

3.16 There are advantages and disadvantages for each approach. Actual performance data will typically better represent the consumer experience in that it provides information based on consumers' actual usage and location. However, the disadvantage of actual performance data is that it will only provide data where the tests are carried out (i.e. it will not cover 100% of a geographic area). It will, for example, provide no data in not spots (by definition) and in areas where little data is available (such as highly rural areas) it may not be possible to derive statistically robust comparisons between operators.

3.17 Predicted performance is likely to offer far more granular geographic data (as operator planning tools can operate down to a high level of geographic granularity) but accuracy of predictions will be subject to error margins given the complexity of predicting radio propagation in cluttered environments and inside buildings. In addition planners cannot always predict how heavily a network will be used. As such predicted performance will not always reflect actual performance.

3.18 Predicted and actual performance metrics can be complementary. Predicted performance can offer relatively low cost, highly granular data and the quality of the predictions can then be validated by correlating with actual performance data. For this reason we see there is merit in collecting data on both predicted and actual performance.

Predicted performance

3.19 We already collect data from MNOs on predicted signal strength for 2G & 3G networks. We use this to estimate geographic and premises coverage across the UK.

This data is collected at a granularity of 200mx200m and is the basis of the coverage information we publish in the CMR and Infrastructure Reports.

3.20 In addition to signal strength, there may be other metrics generated by MNOs' planning tools that it would be appropriate for us to gather. For example, for a given location (e.g. a 200mx200m pixel) or cell site footprint the data types shown in figure 7 may provide additional valuable information.

Figure 7: Possible predicted performance metrics available from MNO planning tools

Metric	Benefit in collecting the data
Signal to noise and interference ratio	Potentially a better indicator than signal strength alone in estimating network coverage
Network technology e.g. 2G, 3G, HSDPA, LTE etc. and which 3GPP software revision has been rolled out	Would allow the roll out of different technology types to be tracked. This could be used as a proxy for mobile broadband performance.
The radio spectrum band and number of carriers in use	Provides insight into spectrum utilisation and network capacity
The backhaul arrangements for a given cell site	Provides insights into speed, capacity and potentially latency of mobile broadband
The geographic area, number of premises, vehicles per day and/or predicted number of calls /day	

3.21 We will seek to engage directly with the MNOs to explore which metrics are produced by their planning tools which might be useful in deriving proxies of QoE.

Actual performance

3.22 Whilst predicted performance data is generally only available from MNOs' planning tools, actual performance data can be provided by MNOs or third parties.

3.23 Typically actual performance metrics are collected by third parties (often on behalf of MNOs) by placing test calls and data on the networks in different locations. Often referred to as 'drive testing' this approach seeks to mimic end user behaviour and so provides a good insight to consumer QoE. The main disadvantage of drive testing is the high costs required to cover a representative sample of the UK, particularly if it has to be repeated at regular intervals.

3.24 As an alternative to drive testing, MNOs are likely to have very rich data from their operational systems which could provide very granular information on service quality – effectively analysing the performance data associated with the millions of calls and data sessions made each day on their networks, rather than relying on a small number of drive tests. The advent of "big data" tools has made it possible to process this data cost effectively and it may be possible to produce suitable proxies of QoE.

3.25 We intend to explore with the MNOs what actual performance data they collect, but we also welcome the views of other stakeholders on the types of actual performance data that are available.

Collection approaches

3.26 There are several alternative methodologies for collecting the underlying data that is needed to provide consumer information. Each will have different merits with respect to the granularity of the data collected, the costs of collection and the quality of the data.

3.27 Broadly, we envisage that the data will come from either the MNOs themselves, from third parties or from a hybrid approach.

MNO sourced data

3.28 As outlined above, MNOs may be able to extract a wide range of relevant data from their existing planning tools and operational systems. We recognise that the information available may vary between MNOs and so work would be required to identify a common set of metrics that would allow MNO performance to be compared fairly.

Third party sourced data

3.29 Ofcom has previously commissioned research into mobile network performance from third parties[17] and as described in paragraph 1.19 is considering research into 4G QoE in 2013. Ofcom also undertakes research using third party data to measure fixed broadband[18]. By commissioning a third party contractor to undertake this work we are not reliant on the service providers to extract data from their systems and we have been able to collect data that is not available in their systems. It also ensures that data is collected across operators in a consistent way and is truly independent of the operators.

3.30 There are a number of approaches third parties adopt to collect data. These include:

- **Crowd sourcing:** recruiting a large panel of consumers to collect data, for example by the installation of a measurement application on their smartphone.

- **Drive/walk testing:** where a small number of measurement devices are driven or walked around a pre-defined area by a third party commissioned by Ofcom. Each device would take measurements of the network either at set intervals, or in specific places, in order to measure performance.

- **Fixed probes:** measurement devices can be installed in fixed locations, such as shopping centres or in blocks of flats, to measure performance at regular intervals of time.

3.31 Each approach has its merits. Crowd sourcing can be a cost effective way to gather large quantities of data, and it reflects actual user locations and use. However it does not cover all locations, it is not always clear where the device is located when the test is made (so it could be indoors or outdoors, in a bag or in the users' hand) and sufficient quantities of volunteers are needed for robust results. It also may not be possible to gather all the metrics required. This raises potential challenges in

[17] http://media.ofcom.org.uk/2011/05/26/mobile-broadband-speeds-revealed/
[18] http://stakeholders.ofcom.org.uk/market-data-research/other/telecoms-research/broadband-speeds/broadband-speeds-may2012/

gathering statistically robust results and specific important information. There may also be costs associated with recruiting the crowd sourced volunteers.

3.32 Drive testing and/or walk testing measurements are taken in a far more controlled environment because the location of the tests is chosen and the location of the test device is known. Tests can be repeated in specific locations as required. However, the cost of data collection can be higher, mainly because of travel costs and, potentially, call and data charges. Drive testing may be more appropriate if targeted data collection is required. For example, concentrating tests in areas which are predicted to have poor performance or are otherwise of specific interest and/or in sample areas to allow predicted performance to be validated with actual performance data.

3.33 Fixed probes are likely to give robust, comparable data because tests are completed in the same place at regular intervals. However, only these locations are sampled. Data costs can be high because of the high volume of traffic sent.

3.34 In our mobile broadband research in 2011, we used all three of the measurement approaches discussed above – measurement devices in fixed location, drive testing in a small number of case study areas and an application downloaded to volunteers' smart phones. We wish to explore through this CFI and through our planned work in 2013 the most effective approaches to third-party collection of mobile QoE information.

Other approaches

3.35 **Hybrid approaches**, where third parties collect data from mobile operators' systems, may provide a good balance of cost vs. independence and quality/depth of data. Such an approach could ensure that data were comparable between operators and consistently analysed. Potentially, the third party could also aggregate data before it is provided to Ofcom.

3.36 **Industry led initiative.** Our primary objective is to ensure consumers have access to accurate and comparable information on mobile performance and this does not necessarily require Ofcom to collect and publish all the relevant data (although we do have duties to collect and publish certain data). An industry led initiative could achieve a similar outcome, possibly in conjunction with comparison websites or consumer information bodies. However, the absence of such an initiative to date suggests that the necessary incentives or coordination are not in place.

3.37 We recognise that the collection of any data will incur costs, whether for Ofcom or operators. It is therefore important that we are proportionate when collecting data – balancing the benefits that are derived from providing information to consumers against the costs of collecting it.

3.38 We believe that Ofcom has a role to play in collecting some form of third party data to ensure information is independent and accurate. We would welcome comments from respondents on Ofcom taking this role and also whether respondents consider the role should be more focused on validating data from operators or collecting the data.

3.39 In choosing a solution we need to balance the quality of information provided to consumers against the costs of collection and publication and the time to implement. We are seeking stakeholder views on the merits of the various possible approaches in order that we can make this judgement.

Annex 1

How to make submissions

A1.1 We welcome views from all stakeholders on any of the points raised in this Call for Input.

A1.2 We invite written views and comments on the issues raised in this document, to be made by **5pm on 1 April 2013**.

A1.3 We are particularly keen to get the views of stakeholders representing the needs of consumers in different parts of the UK to ensure we have a clear view of the information that consumers would find useful when purchasing mobile services. If there is sufficient interest from stakeholders, we propose to host a workshop in March to facilitate the development of ideas and options. **To register your interest in a workshop, please contact us by 15 February 2013**.

A1.4 We strongly prefer to receive responses electronically as this helps us to process the responses quickly and efficiently. Responses can be submitted by:

- email (with accompanying attachments as necessary) to MobileQoE@ofcom.org.uk attaching your response in Microsoft Word format, together with a consultation response cover sheet (see last page).
- using the online web form at https://stakeholders.ofcom.org.uk/consultations/mobile-voice-data-experience/howtorespond/form.
- post to the address below, marked with the title of the consultation 'Measuring mobile quality of experience' (and a completed consultation response cover sheet – see last page).

  Ruth John
  Ofcom
  Riverside House
  2A Southwark Bridge Road
  London SE1 9HA

  Tel: 020 7981 3000

A1.5 Note that we do not need a hard copy in addition to an electronic version. We will acknowledge receipt of responses if they are submitted using the online web form but not otherwise.

Confidentiality

A1.6 We believe it is important for everyone interested in an issue to see the views expressed by consultation respondents. We will therefore usually publish all responses on our website, www.ofcom.org.uk, ideally on receipt. If you think your response should be kept confidential, can you please specify what part or whether all of your response should be kept confidential, and specify why. Please also place such parts in a separate annex.

A1.7 If someone asks us to keep part or all of a response confidential, we will treat this request seriously and will try to respect this. But sometimes we will need to publish all responses, including those that are marked as confidential, in order to meet legal obligations.

A1.8 Please also note that copyright and all other intellectual property in responses will be assumed to be licensed to Ofcom to use. Our approach on intellectual property rights is explained further on its website at http://www.ofcom.org.uk/about/accoun/disclaimer/

Cover sheet for response to an Ofcom consultation

BASIC DETAILS

Consultation title:

To (Ofcom contact):

Name of respondent:

Representing (self or organisation/s):

Address (if not received by email):

CONFIDENTIALITY

Please tick below what part of your response you consider is confidential, giving your reasons why

Nothing	☐	Name/contact details/job title	☐
Whole response	☐	Organisation	☐
Part of the response	☐	If there is no separate annex, which parts?	

If you want part of your response, your name or your organisation not to be published, can Ofcom still publish a reference to the contents of your response (including, for any confidential parts, a general summary that does not disclose the specific information or enable you to be identified)?

DECLARATION

I confirm that the correspondence supplied with this cover sheet is a formal consultation response that Ofcom can publish. However, in supplying this response, I understand that Ofcom may need to publish all responses, including those which are marked as confidential, in order to meet legal obligations. If I have sent my response by email, Ofcom can disregard any standard e-mail text about not disclosing email contents and attachments.

Ofcom seeks to publish responses on receipt. If your response is non-confidential (in whole or in part), and you would prefer us to publish your response only once the consultation has ended, please tick here. ☐

Name

Signed (if hard copy)

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄五

PNS: A study of traffic management detection method & tools

Prepared for Ofcom under MC 316

A Study of Traffic Management Detection Methods & Tools

Predictable Network Solutions Limited

www.pnsol.com

June 2015

Contents

List of Figures

Nomenclature

3G Third generation mobile cellular.

ΔQ See Quality Attenuation.

ADSL Asymmetic DSL.

Applet Small program dynamically downloaded from a webpage and executed locally in a constrained environment.

AS Autonomous System.

Asymmetric In the context of UK internet provision, this means that the linkspeed to the end user is higher than the linkspeed from them.

ATM Asynchronous Transfer Mode.

BRAS Broadband Remote Access Server.

CDN Content Distribution Network.

CSP Communication Service Provider.

CT Computerised Tomography.

DDOS Disributed Denial of Service.

Discrimination In this document the definition used is that of choosing between two or more alternatives.

DOCSIS Data Over Cable Service Interface Specification.

DPI Deep Packet Inspection.

DSL Digital Subscriber Line.

FCFS First-Come-First-Served.

FIFO First-In-First-Out.

GGSN Gateway GPRS Support Node.

ICMP Internet Control Message Protocol.

internet (adj) of, or pertaining to, The Intenet.

Internet, the The global aggregation of packet-based networks whose endpoints are reachable using a unique Internet Protocol address.

IP Internet Protocol.

ISP Internet Service Provider.

Java VM Java Virtual Machine.

L2TP Layer Two Tunneling Protocol.

LAN Local Area Network.

Layer 2 The layer in the internet protocol stack responsible for media access control, flow control and error checking.

Layer 3 The layer in the internet protocol stack responsible for packet forwarding including routing through intermediate routers.

LTE Long Term Evolution - fourth generation mobile cellular.

MPLS Multi-Protocol Label Switching.

MT Mobile Terminal.

OS Operating System.

P2P Peer to Peer.

PASTA Poisson Arrivals See Time Averages.

PBSM Packet-Based Statistical Multiplexing.

PDH Plesiochronous Digital Hierarchy.

PDU Protocol Data Unit: the composite of the protocol headers and the service data unit (SDU).

PGW Packet Data Network Gateway.

PRO Predictable Region of Operation.

QoE Quality of Experience.

Quality Attenuation The statistical impariment that a stream of packets experiences along a network path.

RNC Radio Network Controller.

RTT Round Trip Time.

SDH Synchronous Digital Hierarchy.

SDN Software Defined Networking.

SDU Service Data Unit.

SGSN Serving GPRS Support Node.

SGW Service Gateway.

SIN Supplier Information Note.

SLA Service Level Agreement.

Stationarity The degree to which the characteristics of something (for example Quality Attenuation) are constant in time.

TCP Transmission Control Protocol.

TDM Time-Division Multiplexing.

TM Traffic Management.

TOS Type of Service.

TTL Time to live - the number of router hops that a packet can transit before being discarded.

UDP User Datagram Protocol.

VDSL Very-high-bit-rate DSL.

VLAN Virtual LAN - a method for limiting association in a LAN.

VoD Video on Demand.

VoIP Voice over IP.

WFQ Weighted Fair Queuing.

WRED Weighted Random Early Detection.

Executive Summary of the Research Report

As the demand for broadband capacity by a range of end application services increases, a greater focus is being placed on the use of traffic management (TM) to help meet this increasing demand. Given this, Ofcom commissioned work to further understand the availability of techniques and tools that may be used to detect the presence of TM in broadband networks.

Practical TM detection methods have the potential to form a useful part of the regulatory toolkit for helping the telecommunications market deliver online content and services that meet consumer expectations. In principle, they could potentially help in the following ways:

Increasing transparency for consumers: providing consumers with more information on the application of traffic management and its likely effect on the quality of experience (QoE) of accessing different online services;

Increased visibility for the regulator: the ability to verify operator claims on the employed TM practices within their networks; and

Increased benefits for online service providers: Enabling content and application providers to better deliver their services over broadband networks, by providing more information on the potential effects of TMs and on their products and services.

This report provides the outcome from a literature review of the different techniques that could be used to detect the use of TM.

The report provides a comparative analysis of the identified methods and tools, for example, in terms of:

- Their efficacy in detecting and quantifying the presence of TM in a given network;
- The impact on the network and the consumer in terms generated traffic volume, quality of experience, etc; and
- The need for a given tool or methodology to be integrated within, or executed outside, a given ISP's network infrastructure.

Finally, the report also sets out the key attributes that any future effective TM detection method should meet.

To this end, the report first reviews key papers that cover the most recent, most cited and most deployed techniques for detecting differential management of user traffic. These principally aim to provide end-users with tools that give some indication of whether discrimination is being applied to their own broadband connection. Commercial organisations such as content providers appear to have taken relatively little interest in the commercialisation of TM detection.

While their business is dependent on suitable bounds on the network performance along the path from their servers to their end-users, traffic management (differential or otherwise) is only one of many factors affecting this.

Next, the report further considers the operational behaviours and scalability of these detection approaches, and their potential application and impact in an operational context (i.e. by actors other than individual end-users). Relevant technical developments, models of practical TM measures, and details of the UK context are presented in the appendices.

In terms of key attributes that a future TM detection should meet, we suggest the following:

1. Identify who is responsible for the TM, i.e. where along the complex digital delivery chain it is applied;
2. Be reliable, minimising false positives and false negatives; and

3. Be scalable to deliver comprehensive coverage of potential TM locations, without excessive deployment cost or adverse effect on network performance.

The studied TM detection techniques have been mostly developed in North America, where the market structure differs from that of the UK. Where there is a single integrated supplier, as is typical in North America, establishing that discrimination is occurring somewhere on the path to the end-user is broadly sufficient to identify responsibility. However, in the UK, delivery of connectivity and performance to the wider Internet is split across a series of management domains (scopes of control) and administrative domains (scopes of responsibility).

This makes it harder to identify that domain in which differential management is occurring.

The survey of the open literature identified a set of key papers that describe TM detection methods.

These are:

NetPolice which aims to detect content- and routing-based differentiations in backbone (as opposed to access) ISPs. It does this by selecting paths between different access ISPs that share a common backbone ISP, and using ICMP to detect packet loss locations.

NANO which aims to detect whether an ISP causes performance degradation for a service when compared to performance for the same service through other ISPs. It does this by collecting observations of both packet-level performance data and local conditions and by applying stratification and correlation to infer causality.

DiffProbe which aims to detect whether an access ISP is deploying certain differential TM techniques to discriminate against some of its customers' flows. It does this by comparing the delays and packet losses experienced by two flows when the access link is saturated.

Glasnost which aims to determine whether an individual user's traffic is being differentiated on the basis of application. It does this by comparing the successive maximum throughputs experienced by two flows.

ShaperProbe which tries to establish whether a token bucket shaper is being applied to a user's traffic. It does this by sending increasing bursts of maximum-sized packets, looking for a point at which the packet rate measured at the receiver drops off.

Chkdiff which tries to discern whether traffic is being differentiated on the basis of application. Rather than testing for the presence of a particular TM method, this approach simply asks whether any differentiation is observable, using the performance of the whole of the user's traffic as the baseline.

These techniques are largely successful in their own terms, in that they can detect the presence of particular kinds of differential traffic management operating along the traffic path from an individual user to the Internet. Further work would be needed to independently confirm their reliability claims.

However, none of the currently available techniques meet the desired key attributes of a TM detection system. This is because:

1. Some attempt to establish where TM is occurring along the path examined, but only at the IP layer, which will only localise TM performed at user-visible Layer 3 routers; in the UK context there may not be any such between the user and the ISP. This localisation also relies on a highly restricted router resource, which would limit the scale at which such techniques could be deployed.
2. They aim only to detect the presence of differential TM within the broadband connection of a particular end user.
3. Those that are currently in active deployment generate significant volumes of traffic, which may make them unsuitable for large-scale use.

A key constraint of most of the currently available tools is that they focus on detecting a particular application of a particular TM technique. Even in combination they do not cover all of the potential TM approaches that could be applied. Only NANO and Chkdif may be sufficiently general to overcome this problem.

A further difficulty arises because of the need to attain a broader understanding of what the various actors in the UK digital supply chain may or may not be doing from a TM perspective and how these activities interact. This would require a deeper analysis of the results of many measurements, potentially along the lines of network tomography. This requires further research, and so we must conclude that there is no tool or combination of tools currently available that is suitable for practical use.

In our view, further work is required to develop a broader framework for evaluating network performance, within the context of the inevitable trade-offs that must be made within a finite system. This framework should encompass two aspects:

- A way of identifying the network performance requirements for different applications. The process should be unbiased, objective, verifiable and adaptable to new applications as they appear; and
- A way of measuring network performance that can be reliably related to application needs. This measurement system would need to deal with the fragmented nature of the end-to-end inter-connected delivery chain by reliably locating performance impairments. Any such approach would have to avoid unreasonable loads on the network.

Together these could determine whether a particular network service was fit-for-purpose for different applications; some novel approaches outlined in the report have the potential to do this, in particular developing the ideas of network tomography. This uses the 'performance' of packets traversing a given network to infer details about the underlying network, its performance; and potentially the presence and location of TM.

Network tomography requires further work to establish whether it could become a practical tool (other topics for further study are outlined in the recommendations of the report). TM detection could then become a way to fill in any gaps in the overall framework outlined above.

1. Introduction

1.1. Centrality of communications

"The Internet is increasingly central to the lives of citizens, consumers and industry. It is a platform for the free and open exchange of information, views and opinions; it is a major and transformative medium for business and e-commerce, and increasingly a mechanism to deliver public services efficiently. As such it provides access to a growing range of content, applications and services which are available over fixed and wireless networks." [1]

While BEREC defines the Internet as "... the public electronic communications network of networks that use the Internet Protocol for communication with endpoints reachable, directly or indirectly, via a globally unique Internet address", in common usage it is shorthand for an ever-expanding collection of computing devices, communicating using a variety of protocols across networks that themselves increasingly rely on embedded computing functions.

In order to deliver a useful service, both the computing and communication elements of this system must perform within certain parameters, though the complexity of defining what those parameters should be seems daunting. The current delivery of internet services largely separates responsibility for the computing component (typically shared between the end-user and some service provider) from that for the communications (delivered by various 'tiers' of Internet / Communications Service Providers). The end-to-end digital supply chain is complex and heterogeneous, and the demands placed upon it change so rapidly that some sectors of the industry find themselves "running to stand still"; at the same time, network-enabled functions pervade ever deeper into everyday life. If the promise of "the Internet of Things" is fulfilled this process will go much further still.

1.2. Computation, communication and ICT

Fifty years ago the boundary between 'communication' and 'computation' was relatively clear. Communication took place over circuits constructed on a mainly analogue basis, with the analogue/digital conversion occurring at the network edges. Computation occurred in a limited number of very specialised locations, containing mainframes (or, later, minicomputers). Even though those computers consisted of many components that exchanged data (processors, memory, disk drives), these exchanges were not in the same conceptual category as 'communications'. The dominant mode of use was that the edges transferred data (punch card or line-printer images, characters to/from terminals) via communication links to the central location. The computation was centralised; the edges processed and communicated data, the central computer dealt with the information that was contained within that data.

Today, communication involves extensive use of computation, and ICT functions are no longer centralised. The analogue parts of communication have been relegated to a minor role, with even signal construction/extraction and error detection/correction being done digitally. Communication is now intimately tied to computational processes, and computation (of the kind previously only seen in mainframes, etc.) is occurring in myriad locations. The conceptual separation that existed in the mainframe-dominated world has disappeared.

The new dominant model of ICT is that of interacting and collaborating elements that are physically distributed: web services rely on web browsers to render (and interpret scripts within) the content, which is (often dynamically) constructed on remote web servers; video-on-demand relies on rendering in the device to interpret the content served through a CDN or from a server; cloud services, VoIP, Teleconferencing (both voice and video), etc. all rely on outcomes that involve interaction between communication and computation (often not just at the endpoints[1]).

As computation has been distributed, the requirement to 'pass data' has also been distributed - memory and processing may be half a continent apart, disk drives half the world away. This shift has also 'distributed' other aspects from the computational world to the new communications world, in particular the statistically multiplexed use of resources and its associated scheduling issues. The understanding, management and economic consequences of these issues are no longer confined within the mainframe, but pervade the whole ICT delivery chain.

The distinction between computing and communications has become so blurred that one major class of 'communications' service - that of mobile telephony and data - is perhaps better viewed as a the operation of a massive distributed supercomputer. The ability of a mobile network to deliver voice or data is the direct result of a distributed set of connected computational actions; the network elements[2] are all interacting with each other to facilitate the movement of information.

Such movement of 'voice content' and/or 'data content' is far removed from the concept of 'communication' from 50 years ago. It is no longer about the transmission of bits (or bytes) between fixed locations over dedicated circuits, it is about the *translocation of units of information.* In the mobile network case these 'units' may be voice conversation segments or data packets for application use, the translocation being the consequence of interactions between computational processes embedded in network elements.

At the heart of this process is the statistical sharing of both the raw computation and the point-to-point communication capacity.

1.2.1. Circuits and packets

The underlying communications support for ICT has also changed radically in the last 50 years. The dominant communications paradigm is no longer one of bits/bytes flowing along a fixed 'circuit' (be that analogue or TDM) like "beads on a string". Today's networks are packet/frame[3] based: complete information units are split into smaller pieces, copies of which are 'translocated' to the next location. Note that the information does not actually *move*, it simply becomes available at different locations[4]. This translocation is the result of a sequence of interactions between computational processes at the sending and receiving locations. This is repeated many times along the network path until the pieces of data reach the final computational process[5] that will reassemble them and interpret the information.

Each of these 'store-and-forward' steps involves some form of buffering/queueing. Every queue has associated with it two computational processes, one to place information items in the queue (the receiving action, ingress, of a translocation), the other to take items out (the sending action, egress, of a translocation). This occurs at all layers of the network/distributed application, and each of these buffers/queues is a place where statistical multiplexing occurs, and thus where contention for the common resource (communication or computation) takes place.

Statistical multiplexing is at the core of the current ICT evolution. Using it effectively is key to amortising capital and operational costs, as this permits costs to drop as the number of customers increases[6]. This makes it economic for broadband networks to deliver 'always on'

[1] E.g. combining audio streams in a teleconference is another computational process.

[2] I.e handsets, cell towers, regional network controllers, telephone network interconnects, interface points with the general Internet, etc.

[3] Typically using Ethernet and/or IP.

[4] At most network layers original information units are discarded some time after the remote copy is created.

[5] Always accepting that this is not a perfect process and that there are many reasons why it may get 'lost'.

[6] Note that this is not new: the telegraph was a message-based statistically-multiplexed system in which people took the roles now performed by network elements, such as serialisation and deserialisation, routing, and even traffic management.

connectivity[7], and an ensemble of shared servers to provide 'always available' services.

1.2.2. Theoretical foundations of resource sharing

While distributed computing has advanced tremendously over the last several decades in a practical sense[8], there has been comparatively little attention given to its theoretical foundations since the 1960s. Few 'hands-on' practitioners worry about this, on the basis that 'theory is no substitute for experience'. However, given the extent and speed of change in this industry, there is always a danger that continuing to apply previously successful techniques will eventually have unexpected negative consequences. Such hazards cannot be properly assessed without a consistent theoretical framework, and their potential consequences grow as society becomes increasingly dependent on interconnected computational processes.

To understand network 'traffic management' we must first understand the fundamental nature of network traffic, and indeed of networks themselves. This understanding is built upon three well-established theoretical pillars:

1. A theory of computation, started by Turing, that assumes that information is immediately available for each computational step;
2. A theory of communication, developed by Shannon, that assumes that data is directly transmitted from one point to another over a dedicated channel [2];
3. A theory of communicating processes, developed by Milner, Hoare and others, that assumes that communication between processes is always perfect.

While all of these have been enormously successful, and continue to be central to many aspects of ICT, they are not sufficient to deal with the inextricably woven fabric of computation and communication described in §1.2.1 above, that is loosely referred to as 'the Internet'. The first two theoretical pillars are focused on local issues, whereas the key problem today is to deliver good outcomes on a large scale from a highly distributed system. This inevitably requires some degree of compromise, if only to bring deployments to an acceptable cost point. Statistical sharing - the principle that makes 'always on' mass connectivity economically feasible - is also the key cause of variability in delivered service quality. This is because an individual shared resource can only process one thing at a time, so others that arrive have to wait[9]. This is the aspect of communications that is missing from the third pillar.

Distributed computation necessarily involves transferring information generated by one computational process to another, located elsewhere. We call this function 'translocation', and the set of components that performs it is 'the network'. Instantaneous and completely loss-less translocation is physically impossible, thus all translocation experiences some 'impairment' relative to this ideal. Typical audio impairments that can affect a telephone call (such as noise, distortion and echo) are familiar; for the telephone call to be fit for purpose, all of these must be sufficiently small. Analogously, we introduce a new term, called 'quality attenuation' and written 'ΔQ', which is a statistical measure of the impairment of the translocation of a stream of packets when crossing a network. This impairment must be sufficiently bounded for an application to deliver fit-for-purpose outcomes[10]; moreover, the layering of network protocols isolates the application from any other aspect of the packet transport. This is such an important point it is worth repeating: the great achievement of network and protocol design has been to hide completely all the complexities of transmission over different media, routing

[7] Note, however, that it provides only the *semblance* of a circuit, since in commodity broadband there is no dedication of any specific portion (either in space or time) of the common resources to individual customers.

[8] Driven by advances in processing power and transmission capacity combined with remarkable ingenuity in the development of protocols and applications.

[9] Or, in extremis, be discarded.

[10] Just as a telephone call might fail for reasons that are beyond the control of the telephone company, such as excessive background noise or a respondent with hearing difficulties, applications may fail to deliver fit-for-purpose outcomes for reasons that are beyond the control of the network, e.g. lack of local memory, or insufficient computing capacity. Such considerations are out of scope here.

decisions, fragmentation and so forth, and leave the application with only one thing to worry about with respect to the network: the impairment that its packet streams experience, ΔQ.

ΔQ is amenable to rigorous mathematical treatment[11], and so provides a starting point for the missing theoretical foundations of large-scale distributed computation.

For the purposes of this report, a key point is that ΔQ has two sources:

1. Structural aspects of the network, such as distance, topology and point-to-point bit-rate;
2. Statistical aspects of the network, due to the sharing of resources (including the effects of load).

Separating these two components makes the impact of traffic management easier to understand, as it is concerned only with the sharing of resources. As stated above, sharing resources necessarily involves some degree of compromise, which can be expressed as quality impairment. Traffic management controls how the quality impairment is allocated; and since quality impairment is always present and always distributed somehow or other, traffic management is always present.

1.3. Networks: connectivity and performance

A communications network creates two distinct things:

connectivity the ability of one computational process to interact with another even at a remote location;

performance the manner in which it reacts or fulfils its intended purpose, which is the translocation of units of information between the communicating processes.

Any limitation on connectivity (or more technically the formation of the associations) is typically either under the control of the end-user (e.g. using firewalls) or follows from due legal process (e.g. where the Courts require ISPs to bar access to certain sites).

For a distributed application to function at all, appropriate connectivity must be provided; however, for it to function well, appropriate performance (which is characterised by ΔQ) is also essential[12].

Performance, however, has many aspects that act as a limit. Geographical distance defines the minimum delay. Communication technology sets limits on the time to send a packet and the total transmission capacity. Statistical sharing of resources limits the capacity available to any one stream of packets. The design, technology and deployment of a communications network - its structure - sets the parameters for a best-case (minimum delay, minimal loss) performance at a given capacity. This is what the network 'supplies', and this supply is then shared between all the users and uses of the network. Sharing can only reduce the performance and/or the capacity for any individual application/service.

Communications networks are expensive, and so the ubiquity of affordable access is only possible through dynamic sharing of the collective set of communication resources. It is a truism that such dynamically shared networks deliver the best performance only to their very first customers; the gradual decrease in performance for individual users as user numbers increase is a natural consequence of dynamic resource sharing in PBSM.

To give a practical example of what this sharing means, for a single consumer to watch an iPlayer programme successfully, typically there must be 15 to 20 other consumers (on the same ISP) who are not using the network at all in any one instant of the programme's duration[13].

Traffic management (the allocation of quality impairment) is at the heart of this sharing process. It works in one of two ways: it either shares out access to the performance (its

[11] This is discussed in more detail in Appendix A.

[12] This is discussed in more detail in Appendix §A.2.

[13] It doesn't have to be the same 15 to 20 users, it can be a dynamically changing set; note also that it is not just the aggregate capacity that is shared, but the ability to deliver data within time constraints.

delay, its loss and its capacity); or it limits demand on the supply (thus reducing the effects of sharing elsewhere).

1.4. Traffic Management

Clearly a balance needs to be struck between TM techniques applied to improve services to end-users and TM that (either intentionally or otherwise) degrades services unnecessarily. As stated in [3], "The question is not whether traffic management is acceptable in principle, but whether particular approaches to traffic management cause concern."

Statistical sharing of resources inevitably involves a tradeoff: the more heavily a resource is used, the more likely it is to be in use when required. Buffering is needed to allow for arrivals to occur when the resource is busy. This creates contention for two things, the ability to be admitted into the buffer (ingress) and the ability to leave the buffer (egress). Whether the first is achieved determines loss, and the time taken to achieve the second determines delay; together these create the variable component of quality attenuation. Traffic management mechanisms vary in the way they control these two issues. In Appendix B, we discuss the TM techniques that are widely deployed, and their impact on network performance. One key application of TM is to keep services within their 'predictable region of operation' (PRO); this is particularly important for system services (such as routing updates or keep-alives on a L2TP tunnel) whose failure might mean that all the connections between an ISP and its customers are dropped.

It is important to distinguish between TM that is 'differential' (in that it treats some packet flows differently from others) from that which is not, which is far more common (for example rate limiting of a traffic aggregate[14]). Differential TM may be intra-user (treating some flows for a particular user differently to others) or inter-user (treating traffic of some users differently from that of others, for example due to different service packages).

TM may be 'accidental' (the emergent consequences of default resource sharing behaviour) or 'intentional' (configuration of resource sharing mechanisms to achieve some specific outcome). The use of intentional TM to maintain essential services may be uncontroversial, but its application to manage the tension between cost-effectiveness and service quality is not. Because quality attenuation is conserved (as discussed in more detail in §B.3), reducing it for some packet flows inevitably means increasing it for others, to which some users may object. Traffic Management Detection sets out to discover whether such differential treatment is occurring.

1.4.1. Traffic management detection

The purpose of this report is to increase the understanding of the methods and tools available, to understand the art of the possible in the area of TM detection and evaluation. First of all, we must ask: what is the purpose of such detection? It is important to distinguish between testing for the operational effect of an intention and inferring an intention from an observed outcome. The later is logically impossible, because observing a correlation between two events is not sufficient to prove that one causes the other, and, even if an outcome is definitely caused by e.g. some specific configuration, this does not prove a deliberate intention, as the result might be accidental. The former is possible, but must start from an assumption about the intention; TM detection, by its nature, falls into this category. Secondly, we can ask: what criteria should any TMD methods and tools satisfy? At a minimum, we suggest, in addition to 'detecting' TM, any method should:

1. Identify the location of application of TM along the digital delivery chain;
2. Be reliable, minimising false positives and false negatives;

[14]Note that, just because TM is not differential does not guarantee that it will be 'fair' to all packet flows, as discussed in §B.1.1.4.

3. Be scalable to deliver comprehensive coverage of potential TM locations, without excessive deployment cost or adverse effect on network performance[15].

In §2, we review and compare the most up-to-date techniques in the literature for performing TM detection, and discuss the operational context of such detection in §3.

1.4.2. Traffic management in the UK

A consumer of internet access services (whether domestic or commercial) has to have a connection to some infrastructure, which in the UK is quite diverse in both structure and technology. In Appendix C, we explore the particular characteristics of network provision in the UK, and the implications of this for TM and TM detection. An important aspect is that the delivery of connectivity and performance to the wider Internet is split across different entities, some internal and some external. These form a series of management domains (scopes of control) and administrative domains (scopes of responsibility). Boundaries between these domains are points where TM might be applied; some of them are points where TM *must* be applied to keep services within their PRO. These are illustrated for the UK wireline context in Figure 1.1 on the facing page. Note especially the different coloured arrows that distinguish the level of aggregation at which TM might be applied.

It is important to consider what 'positive detection' of traffic management would mean in a UK context. Knowing that there may be traffic management occurring somewhere along the path between an end-user and the Internet does not identify which management / administrative domain it occurs in, which could be:

- before the ISP (even outside the UK);
- within the ISP;
- after the ISP;
- in a local network (depending on router settings).

Thus it is a challenge simply to determine whether the 'cause' is within the UK regulatory context. Even 'locating' the point at which intentional TM seems to be occurring still leaves the question of whose management domain this is in (and whose administrative domain *that* is in), which may not be straightforward to answer.

1.5. Previous BEREC and Ofcom work

1.5.1. BEREC reports

BEREC has published a variety of reports related to this topic. In general their approach is to look at:

1. Performance of internet access as a whole and its degradation;
2. Performance of individual applications and their degradation.

BEREC's 2012 report [4] makes the important point that "A precondition for a competitive and transparent market is that end users are fully aware of the actual terms of the services offered. They therefore need appropriate means or tools to monitor the Internet access services, enabling them to know the quality of their services and also to detect potential degradations." This is a positive and helpful contribution, but it leaves open the question of what parameters should be used to specify the services offered to assure that they are suitable for delivering fit-for-purpose application outcomes. Again this leads to the question of what

[15] By its nature, the intention behind any TM applied is unknowable; only the effects of TM are observable. It may be worth noting that, due to this, the best way to ensure end-users receive treatment in line with expectations may be two-fold: to contract to objective and meaningful performance measurements; and to have means to verify that these contracts are met.

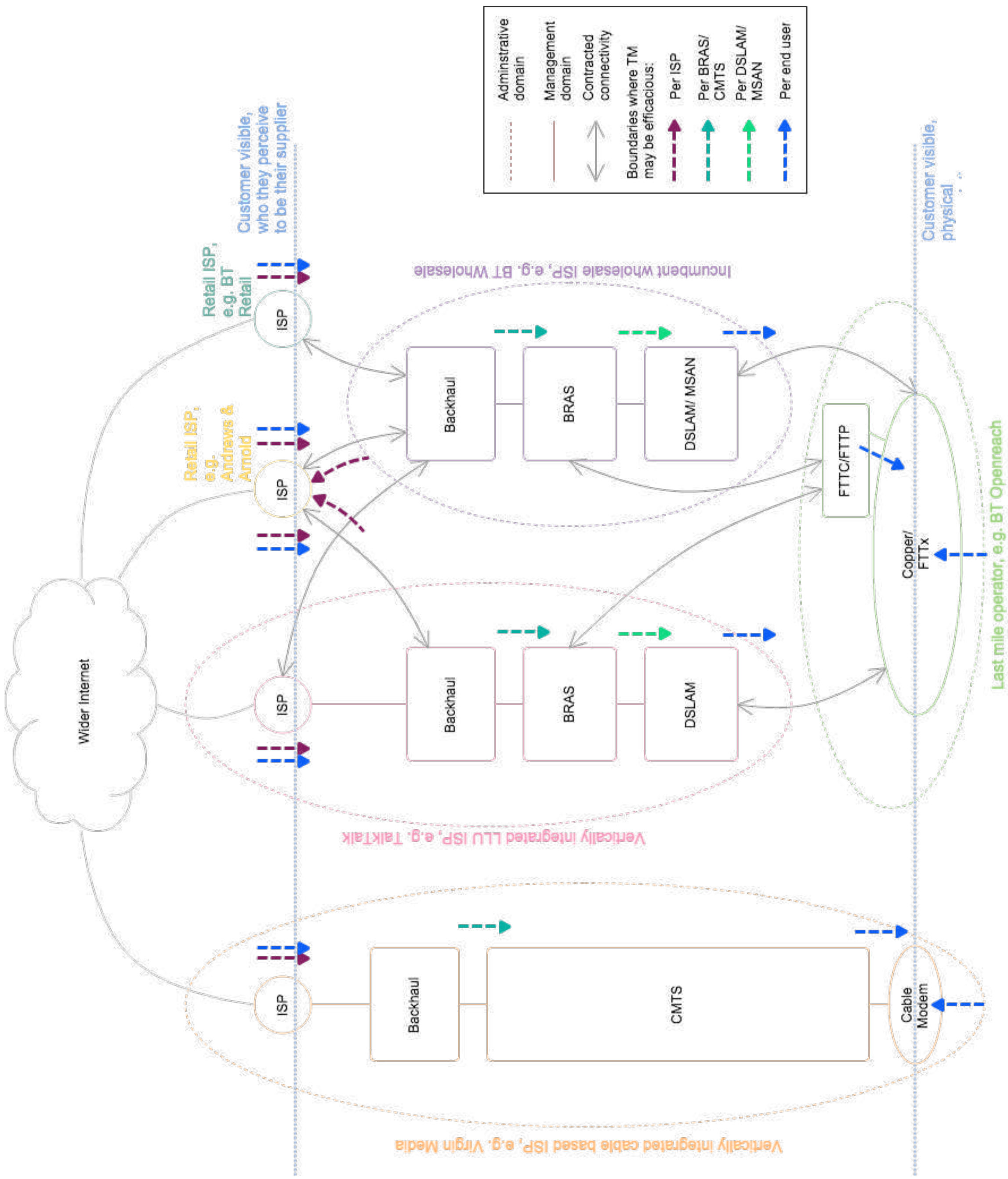


Figure 1.1.: Potential TM points in the UK broadband infrastructure (wireline)

the typical fluctuations of such parameters are during normal operation, so as to distinguish these from the effects of traffic management.

BEREC's most recent report on this topic [5] uses an approach which equates 'equality of treatment' with delivering equality of outcomes. As discussed in §B.2, this assumption does not always hold. It further states (in Section 4.1 of [5]) that delivering good 'scores' against averages of standardised measures will be sufficient to guarantee good outcomes. As discussed in Appendix A, this assumption may also not hold.

However, the BEREC report does identify a number of requirements for quality monitoring systems, but does not explicitly specify that they should be directly relatable to application outcomes. While the report only identifies a small amount of application-specific degradation, it concludes that wide-scale monitoring is desirable. This may be an important recommendation, but may lead to large expenditure and mis-steps without a greater understanding in the industry in general of the relationship between measured performance and fit-for-purpose outcomes.

1.5.2. Notes on previous Ofcom studies

The most recent study on this topic commissioned by Ofcom [6] is very thorough, but misses crucial points:

- There are implicit assumptions that customer QoE is determined primarily by bandwidth and that additional measures such as prioritisation will have predictable effects.
- There is a further assumption that typical measurements of additional parameters, such as average latency, can be reliably related to QoE for 'latency sensitive' applications.

However this 2011 study makes a distinction between 'decision basis' and 'intervention', which is useful, as is the observation that traffic management can vary from user to user depending on their contractual situation and usage history. It also points out that flow identification and marking is generally done at Layer 3, while rate limiting/shaping may be applied at Layer 2; and that traffic management is typically applied in order to deliver better QoE for the majority of users. The comments in section 6 of [6] regarding the difficulty of observing traffic management represent a starting point for this report. However we note that the suggestion in section 8 of [6], that real-time indicators should be provided of whether various services can be supported, can only be realised if they are based on appropriate measures and models (as discussed in §A.2 below) not on proxies such as bandwidth or latency.

1.6. Summary

Communications have changed a great deal in the last half-century, particularly in the shift from dedicated circuits to statistically-shared resources, which has made global connectivity widely affordable. The consequences of this shift are still being worked out, in particular understanding what it would reasonable for users to expect. As more people and services come to depend on this fundamentally rivalrous resource, the issues of experience, application outcome, consistency and differential treatment (intentional or otherwise) are becoming increasingly important. These factors impact the effectiveness of the delivered service for any individual user's needs-of-the-moment, and hence the value that it has for them.

The stakes are increasing and thus so are the pressures to apply intentional traffic management (if only to mitigate the emergent effects of implicit and unintentional traffic management). Having tools to confirm that the delivered operational characteristics are as intended, and to raise appropriate questions when the intention and the observed outcomes are at odds, will be an important part of the regulator's toolset.

2. Traffic management detection and analysis

2.1. Introduction

In statistically-multiplexed networks such as the Internet, it is inevitable that there will be periods in which various resources are overloaded. At such times, some packets will have to be delayed, and some may need to be discarded. Any mechanism that applies a policy as to which packets will receive which treatment[1] can be called 'traffic management'. The main focus of interest, however, is on ISP-configured policies that 'discriminate' against particular traffic flows. This interest is, as far as it is possible to tell, almost entirely academic. While it might be expected that commercial organisations whose business depends on delivering some form of good experience across the Internet would be interested in this topic, on careful consideration this expectation is misguided. These organisations are dependent on suitable bounds on the quality attenuation, on the path from their servers to the end-users[2], which is a function of much more than TM policies applied by an ISP. While some ISPs may enable better performance for the application in question than others, exactly why this is the case is of secondary concern[3].

2.2. Traffic management

Transferring information generated by one computational process to another, located elsewhere, is called 'translocation', and the set of components that performs it is 'the network'. Instantaneous and completely loss-less translocation is physically impossible, thus all translocation experiences some 'impairment' relative to this ideal.

Translocating information as packets that share network resources permits a tremendous degree of flexibility and allows resources to be used more efficiently compared to dedicated circuits. In packet-based networks, multiplexing is a real-time 'game of chance'; because the state of the network when a packet is inserted is unknowable, exactly what will happen to each packet is uncertain. The result of this 'game' is that the onward translocation of each packet to the next element along the path may be delayed, or may not occur at all (the packet may be 'lost'). This is a source of impairment that is statistical in nature.

The odds of this multiplexing game are affected by several factors, of which load is one. In these 'games', when one packet is discarded another is not, and when one is delayed more another is delayed less, i.e. this is a zero-sum game in which quality impairment (loss and delay) is conserved.

As discussed in Appendix B, 'traffic management' is applied to the translocation of information through these networks, and its effect is to alter the odds of the multiplexing game and hence the delivered quality attenuation (ΔQ). This ΔQ is the way in which the network impacts the performance of an application[4].

[1] Even FIFO queuing is a policy, and as discussed in §B.1.1, not one that can be assumed to always deliver good outcomes.

[2] This is discussed in Appendix A.

[3] Although, where this is the case, commercial organisations may want to measure and publicise this to promote their product.

[4] This is discussed in Appendix A.

Most traffic management detection approaches implicitly use application performance to infer aspects of ΔQ, and thereby draw conclusions regarding the nature of the traffic management; a doubly-indirect process.

2.3. Techniques for detecting traffic management

A variety of approaches have been proposed for detecting whether any form of differential traffic management is being applied at some point in the delivery chain (typically by ISPs). The key papers used in this study are [7, 8, 9, 10, 11], which are collectively the most recent, most cited and most deployed techniques, as revealed by a diligent study of scholarly sources (discussed further in Appendix E). These are described in more detail below. Most aim to provide end-users with a tool that gives some indication of whether such intra-user discrimination is being applied to their own connection. A thorough discussion of the constraints this imposes on the testing process can be found in [8], where Dischinger et al. assert that:

1. Because most users are not technically adept, the interface must be simple and intuitive;
2. We cannot require users to install new software or perform administrative tasks;
3. Because many users have little patience, the system must complete its measurements quickly;
4. To incentivise users to use the system in the first place, the system should display per-user results immediately after completing the measurements.

Since information is translocated between components of an application as sequences of packets, any discrimination must be on the basis of attributes of those packet sequences. Most approaches actively inject traffic whose packets differ in one specific respect from reference packets[5] and then seek to measure differences in throughput, loss or delay. These approaches are criticised in [9] on the grounds that ISPs might learn to recognise probing packets generated by such tests and avoid giving them discriminatory treatment[6].

There is a further body of relevant literature, outlined in Appendix E. Few papers appear to have been published in this field in the last two or three years.

2.3.1. NetPolice

This tool was developed at the University of Michigan in 2009, by Ying Zhang and Zhuoqing Morley Mao of the University of Michigan and Ming Zhang of Microsoft Research [10].

2.3.1.1. Aim

This system, called NetPolice, aims to detect content- and routing-based differentiations in backbone (as opposed to access) ISPs. This is mainly to inform large users, such as content providers, rather than individual end-users.

2.3.1.2. Framing the aim

NetPolice focuses on detecting traffic differentiation occurring in backbone ISPs that results in packet loss. Since backbone ISPs connect only to other ISPs, not to end-users, this can only be done by measuring loss between end-hosts connected to *access* ISPs. By selecting paths between different access ISPs that share a common backbone ISP (a technique that is conceptually similar to the network tomography approach discussed in §2.3.7) measurements can be inferred for the common backbone ISP. ISPs are distinguished on the basis of their 'autonomous system' (AS) number.

[5] These reference packets may be passively observed as in [12] or actively generated such as in [8, 10].

[6] It is to be noted that this would only become likely if such methods came to be used widely, which so far none have.

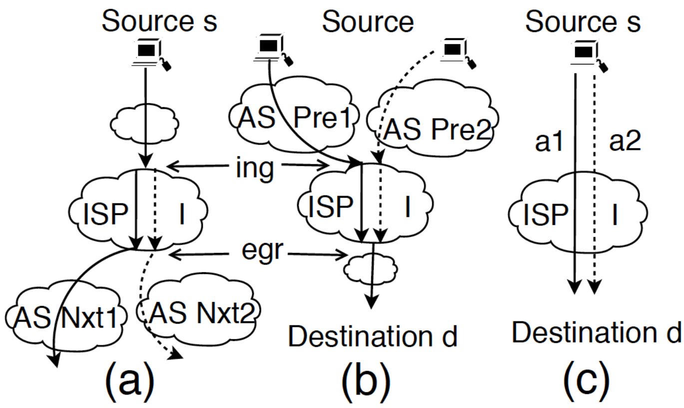


Figure 2.1.: Detecting various types of differentiation with end-host based probing Reproduced from [10]

Key challenges included selecting an appropriate set of probing destinations to get a sufficient coverage of paths through backbone ISPs[7] and ensuring the robustness of detection results to measurement noise. The system was deployed on the PlanetLab platform and used to study 18 large ISPs spanning 3 continents over 10 weeks in 2008.

2.3.1.3. Implementation

NetPolice exchanges traffic between end-hosts, selected so that paths between them have appropriate degrees of difference and commonality, and measures loss rates in order to detect differentiation. To measure the loss rate along a particular subsection of the end-to-end path, NetPolice sends probe packets with pre-computed TTL values that will trigger ICMP 'time exceeded' responses[8], unless the packet is lost. As packet loss may occur in either direction, large probe packets are used to ensure the measured loss is mostly due to forward path loss, on the assumption that large probe packets are more likely to be dropped than small ICMP response packets on the reverse path. Subtracting the measured loss rate of the sub-path to the ingress of a particular AS from that of the egress from it provides the loss rate of the internal path. Figure 2.1 illustrates how NetPolice uses measurements from end systems to identify differentiation in ISP I. In Figure 2.1(a), an end host probes two paths sharing the same ingress and egress within ISP I, but diverging into two distinct next-hop ASes after the egress. By comparing the loss performance of the two paths, NetPolice determines whether ISP I treats traffic differently based on the next-hop ASes. Similarly, Figure 2.1(b) shows how NetPolice detects differentiation based on previous-hop ASes. In Figure 2.1(c), an end-host probes a path that traverses the same ingress and egress of ISP I to the same destination.

To detect content-based differentiation, the tool measures loss rates of paths using different application traffic. Five representative applications were used: HTTP; BitTorrent; SMTP; PPLive; and VoIP. HTTP was used as the baseline to compare performance with other applications, on the assumption that it would receive neither preferential nor prejudicial treatment. The remaining four applications were selected based on a prior expectation that they may be treated differently by backbone ISPs. Packet content from real application traces was used, with all packets padded to the same (large) size, and their sending rate restricted to avoid ICMP rate-limiting constraints[9]. NetPolice detects differentiation by observing the differ-

[7] Choosing the optimal set of hosts to exchange traffic in order to probe a particular sub-path is an instance of the set covering/packing problem, a classic question in combinatorics, computer science and complexity theory. See `https://en.wikipedia.org/wiki/Set_packing`, which also includes some discussion of useful heuristics.

[8] Although an ICMP response may be forwarded on a slow path, this will not affect the loss measurement provided the packet is not dropped.

[9] Intermediate routers limit the rate of ICMP requests they will respond to.

ences in average loss rates measured along the same backbone ISP path using different types of probe traffic.

The issue of network load induced by probing is addressed by means of "collaborative probing". This consists of selecting end-host pairs whose connecting paths traverse the sub-paths of interest. The selection is made so that these sub-paths are probed sufficiently often (by traffic between different pairs of hosts) whilst ensuring that the probing traffic is spread out over different access ISPs.

Differences due to varying network load (rather than 'deliberate' differentiation) were addressed by:

1. taking repeated measurements;
2. assuming even distribution of "random noise"[10];
3. applying multivariate statistical tests to the measurements to compare the distributions of baseline and selected application traffic.

2.3.1.4. TM techniques detected

Only traffic management that induces packet loss can be detected[11]. Since the rate of each probing flow is low, this must be applied to a traffic aggregate (i.e. an aggregated flow of packets from many users sharing some common attribute). Thus rate policing of aggregate traffic based on port number, packet contents and/or source/destination AS is the only mechanism detected.

2.3.1.5. Discussion

In the paper it is assumed that inaccuracy of loss rate measurements is likely to be caused by three main factors:

1. overloaded probers;
2. ICMP rate limiting at routers; and
3. loss on the reverse path.

Little evidence is produced to justify these assumptions other than a partial validation of single-ended loss-rate measurements against a subset of double-ended measurements (i.e. loss rate measured at the remote host), by plotting the corresponding CDFs and showing that they are broadly similar. There is also a correlation of the results with TOS values returned in the ICMP response packets, presumably added by ISP ingress routers.

Since packets are padded to the same (large) size, and their sending rate restricted to avoid ICMP rate limiting constraints, the packet streams are not representative of real application traces.

Note that routers typically limit their ICMP response rate (on some aggregate basis), in order to ensure that other critical router functions remain within their PRO. Thus, it would seem that consistent application of this technique would require a single point of control to coordinate the packet streams in order to avoid exceeding this rate at any router being probed. Also the possibility that routers may have this function disabled altogether must be considered.

This technique is restricted to detecting TM performed by Tier 1 ISPs. Therefore it appears to have limited applicability for ISPs with multiple geographically diverse subnetworks within the same AS.

There is a fundamental difficulty with ensuring that the selection of end hosts is optimal and that all sub-paths will be probed, particularly in the presence of dynamic routing.

[10] The paper's authors' term for the effects of congestion.

[11] In ΔQ terms, what is actually being measured is an approximation to that part of $\Delta Q_{|\mathbf{V}}$ whose packets are never delivered or whose delays are beyond a cut-off, in this case the duration of the test, since it is impossible to distinguish packet loss from very large delay by observation.

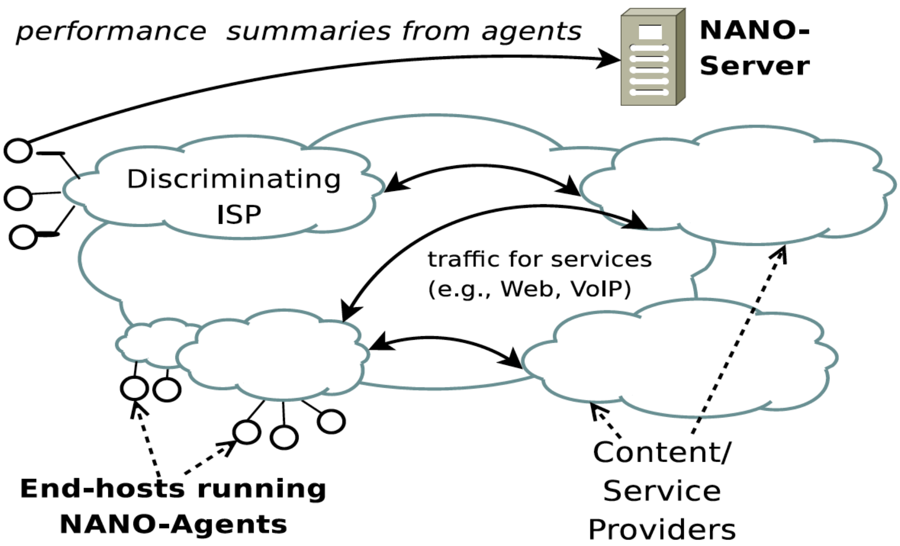


Figure 2.2.: NANO architecture
Reproduced from [9]

2.3.2. NANO

Detecting Network Neutrality Violations with Causal Inference, here referred to by the name of its technique NANO, is a 2009 paper by Mukarram Bin Tariq, Murtaza Motiwala, Nick Feamster and Mostafa Ammar at the Georgia Institute of Technology [9].

Aim

The aim is to detect whether an ISP causes performance degradation for a service when compared to performance for the same service through other ISPs.

Framing the aim

A service is an "atomic unit" of discrimination (e.g. a group of users or a network-based application). 'Discrimination' is an ISP policy to treat traffic for some subset of services differently such that it causes degradation in performance for the service. An ISP is considered to 'cause' degradation in performance for some service if a causal relation can be established between the ISP and the observed degradation. For example, an ISP may discriminate against traffic, such that performance for its service degrades, on the basis of application (e.g. Web search); domain; or type of media (e.g. video or audio). In causal analyses, "X causes Y" means that a change in the value of X (the "treatment variable") should cause a change in value of Y (the "outcome variable"). A "confounding variable" is one that correlates both with the treatment variable in question (i.e. the ISP) and the outcome variable (i.e. the performance).

NANO is a passive method that collects observations of both packet-level performance data and local conditions (e.g. CPU load, OS, connection type). To distinguish discrimination from other causes of degradation (e.g. overload, misconfiguration, failure), NANO establishes a causal relationship between an ISP and observed performance by adjusting for confounding factors that would lead to an erroneous conclusion. To detect discrimination the tool must identify the ISP (as opposed to any other possible factor) as the underlying cause of discrimination.

Implementation

NANO agents deployed at participating clients across the Internet collect packet-level performance data for selected services (to estimate the throughput and latency that the packets experience for a TCP flow) and report this information to centralised servers, as shown in Figure 2.2. Confounding factors are enumerated and evaluated for each measurement. The values of confounding factors (e.g. local CPU load) are stratified[12]. Stratification consists of placing values into 'buckets' (strata) sufficiently narrow that such values can be considered essentially equal, while also being wide enough that a large enough sample of measurements can be accumulated. Measurements are combined with those whose confounding factors fall into the same strata, and statistical techniques drawn from clinical trial analysis are used to suggest causal relationships. Stratification requires enumerating all of the confounding variables, as leaving any one variable unaccounted for makes the results invalid. NANO considers three groups of such confounding variables: client-based, such as the choice of web-browser, operating system, etc.; network-based, such as the location of the client or ISP relative to the location of the servers; and time-based, i.e. time of day.

Discussion

NANO captures specific protocol interactions related to TCP, measuring the interaction of the network with the performance of an application. This is mediated by the behaviour of the sending and receiving TCP stacks. As such, it does not measure delay and loss directly, but rather the combined effects of both the bi-directional data transport and the remote server. From a ΔQ perspective (discussed in more detail in Appendix A), the measurements are of an application outcome (throughput achieved over a TCP connection), which is highly dependent on $\Delta Q_{|\mathrm{G,S}}$, as well as on $\Delta Q_{|\mathrm{V}}$, the component that is affected by TM.

The technique has significant advantages that come with passive data collection such as: protection from preferential treatment for probe traffic; an absence of resource saturation caused by testing; and no impact on user data caps (where applicable), other than server upload (which is not deemed significant). A disadvantage of being entirely passive, however, is that data gathering depends on usage profiles of participating users.

Collecting data on local conditions helps to isolate some confounding factors. While the statistical basis for the work and the use of stratification as a technique within which to do comparative testing is well-established, it has also been criticised, e.g. in [13]. The paper asserts that NANO can isolate discrimination without knowing the ISP's policy, as long as values are known for the confounding factors. It further asserts that these confounding factors are "not difficult to enumerate using domain knowledge", an assertion that may need both further investigation and justification that is not provided in the paper itself. While this technique has had successful test deployments (using a combination of Emulab and PlanetLab), this proof-of-concept run does not seem to provide an adequate basis for the assumptions made with respect to the possible set of confounding factors. There appears to be an implicit assumption that the only difference between one ISP and another is the TM that they perform. At one point the idea of "network peculiarities" is mentioned as something on which performance might depend, but if, for instance, the technology used in one network (e.g. cable) gave a different set of performance criteria to another (e.g. 3G) it is unclear whether or not this would be seen as discrimination[13].

Nano has the advantage of adding only minimal traffic to the network (only that required to report the results to the central server), but it does not seem to provide any way to establish where in the digital supply chain any discrimination is taking place, unless it were possible to observe packets at intermediate points. Combining the sophisticated statistical approach here with some variant of the network tomography ideas discussed in §2.3.7 might produce a

[12] `http://en.wikipedia.org/wiki/Stratified_sampling`
[13] Clarifying this would require laboratory-based study.

powerful and scalable tool, although the computational cost of performing the analysis would need to be investigated.

2.3.3. DiffProbe

DiffProbe was developed by P. Kanuparthy and C. Dovrolis at the Georgia Institute of Technology in 2010 [12].

Aim

The objective of this paper was to detect whether an access ISP is deploying mechanisms such as priority scheduling, variations of WFQ[14], or WRED[15] to discriminate against some of its customers' flows. DiffProbe aims to detect if the ISP is using delay discrimination, loss discrimination, or both.

Framing the aim

The basic idea in DiffProbe is to compare the delays and packet losses experienced by two flows: an Application flow A and a Probing flow P. The tool sends (and then receives) these two flows through the network concurrently, and then compares their statistical delay and loss characteristics. Discrimination is detected when the two flows experience a statistically significant difference in queueing delay and/or loss rate. The A flow can be generated by an actual application or it can be an application packet trace that the tool replays. It represents traffic that the user suspects their ISP may be discriminating against (e.g. BitTorrent or Skype). The P traffic is a synthetic flow that is created by DiffProbe under two constraints: firstly, if there is no discrimination, it should experience the same network performance as the A flow; secondly it should be classified by the ISP differently from the A flow.

Implementation

DiffProbe is implemented as an automated tool, written in C and tested on Linux platforms, comprising two endpoints: the client (CLI, run by the user), and the server (SRV). It operates in two phases: in the first phase, CLI sends timestamped probing streams to SRV, and SRV collects the one-way delay time series[16] of A and P flows; in the second phase, the roles of CLI and SRV are reversed.

DiffProbe generates the A flow using traces from Skype and Vonage[17]. Various aspects of the A flow are randomised (port, payload, packet size and rate) to generate the P flow.

Two techniques are used to minimise the rate of false positives, i.e. to ensure that the two flows see similar network performance when the ISP does *not* perform discrimination. The first of these is to consider only those P packets that have been sent close in time with a corresponding A packet[18]. Secondly, when a P packet is sent shortly after an A packet, it is generated such that it has the same size as that A packet. This ensures that the network transmission delays of the (A, P) packet pairs considered are similar. This is illustrated in Figure 2.3.

[14] WFQ is a form of bandwidth sharing, described in §B.4.3.

[15] WRED is a form of policing and shaping, as discussed in §B.4.4 and §B.4.5, in which packets are discarded with some probability when the queue is in states other than full.

[16] The term 'time series' as used in this paper means the end-to-end delays of a flow, after subtracting the minimum observed measurement from the raw end-to-end delay measurements. The presence of a clock offset does not influence these measurements as the focus is on relative, not absolute delays.

[17] This is presumably based on an expectation that these particular applications may be discriminated against.

[18] This should mean that, even if the P flow includes many more packets than the A flow, with different sizes and inter-arrival intervals, only (A, P) packet pairs that have 'sampled' the network at about the same time are considered.

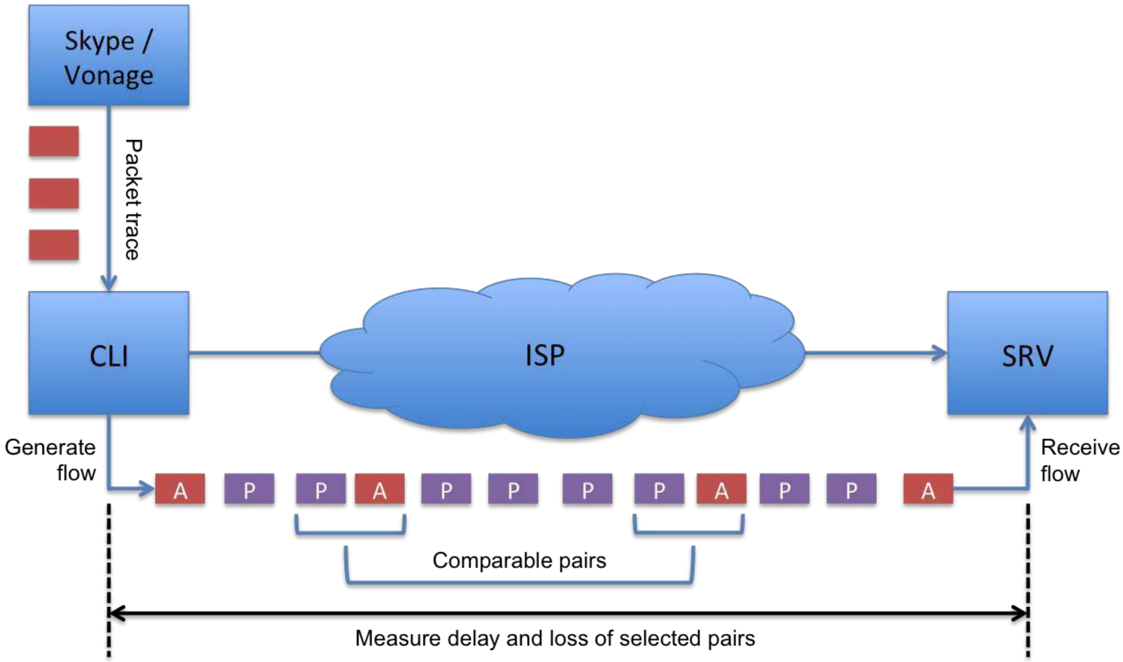


Figure 2.3.: DiffProbe architecture

In order to increase the chances that a queue will form inside the ISP, causing the supposed discriminatory mechanism to be applied, the rate of the P flow is increased to close to the rate of the access link (whose capacity is estimated in a previous phase[19]). If no significant difference[20] is detected between the delays during an interval with a typical load and one with an increased load, the measurement is discarded (on the grounds that no discrimination has been triggered).

Discrimination is detected by comparing the delay distributions of the (A, P) pairs, taking account of the fact that many packets experience a delay that is dominated by propagation and transmission times[21]. If the delay distributions are statistically equivalent, then a null result is returned. Otherwise they are compared to see if one is consistently and significantly larger than the other.

Loss discrimination is also measured, by comparing the proportion of lost packets in the two flows. In order to apply the chosen significance test, the high-load period is extended until at least 10 packets are lost from each of the flows.

TM methods detected

Discrimination due to strict priority queuing is distinguished from that due to WFQ on the basis of the delay distribution of the 'favoured' packets (see Figure 2.4, reproduced from the paper). This approach detects both delay-affecting TM (such as Priority Queuing, discussed in §B.4.2, and bandwidth sharing, discussed in §B.4.3) and loss-affecting TM, such as WRED[15].

Discussion

This paper considers both delay and loss discrimination, but unfortunately treats delay and loss as entirely separate phenomena (whereas they are always linked through the two degrees of

[19]This is done by: sending K packet trains of L packets, each of size S; at the receiver, measuring the dispersion D for each train (the extent to which packets have become separated in their passage across the network); estimating the path capacity as: $C = (L-1)S/D$; finally, taking the median of the K trains [14].

[20]The differential factor for this decision was chosen empirically.

[21]In terms of ΔQ, the process in Footnote 16 can be seen as an estimation of the unidirectional $\Delta Q_{|\mathsf{G}}$. The statistical test used here appears to have been chosen mitigate the effects of $\Delta Q_{|\mathsf{S}}$, which manifests here as a (unwanted) correlation between packet size and delay.

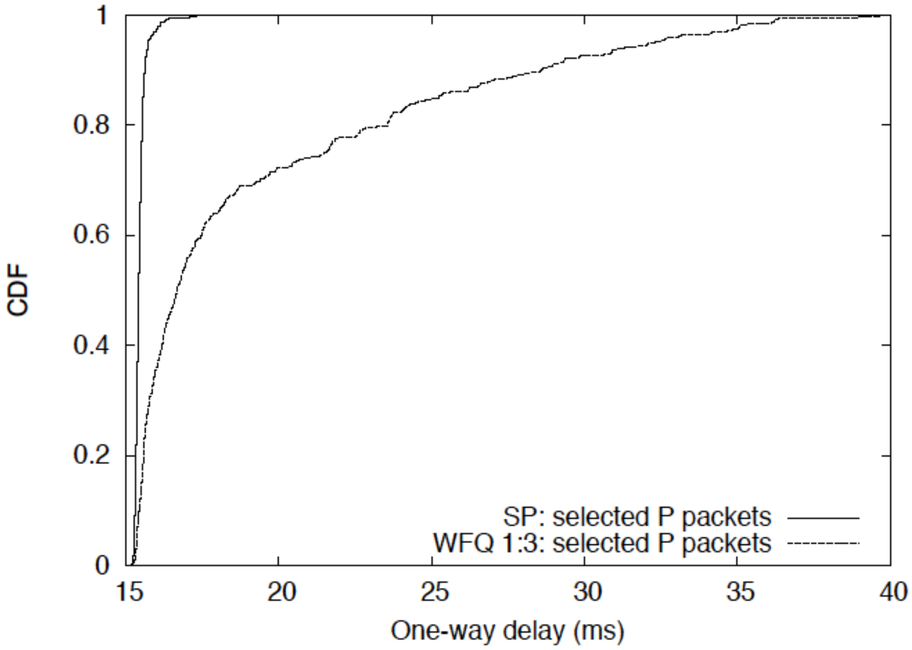


Figure 2.4.: Delay distributions due to strict priority and WFQ scheduling (simulated) Reproduced from [12]

freedom that all queueing systems inherently have). By considering only differential delays[22], $\Delta Q_{|\mathtt{G}}$ is effectively separated from the other components of ΔQ. However, it appears that $\Delta Q_{|\mathtt{S}}$ is not fully considered[23] and the authors do not exploit the fact that $\Delta Q_{|\mathtt{V}}$ can be extracted from the full ΔQ. This leads to the use of a complex statistical test in order to cope with delay distributions having a large cluster of measurements around $\Delta Q_{|\mathtt{G,S}}$.

This approach tries to avoid the (common) overly-strong stationarity assumption (that packets sent at different times will see essentially similar quality attenuation) by selecting packet pairs for comparison. However, this requires care to avoid the edge effect of the loss process due to tail drop[24] (or other buffer exhaustion, see §B.1.1.4). There is no apparent evidence that such care has been taken in this case; in particular the fact that the selected $(A,\ P)$ packet pairs always have the P packet second may introduce bias[25].

There is an assumption in the paper that any differential treatment will only be manifest when a particular network element is reaching resource saturation[26]. To bring this about, the offered load of the P traffic is increased until it reaches the (previously determined) constricting rate. In a typical UK broadband deployment, this method would likely only detect differential treatment on the access link. In the upstream direction this would be in the CPE device (under the nominal control of the end user themselves); and in the downstream direction would typically be under the control of the wholesale management domain[27]. If the *retail* ISP was engaging in such discrimination[28], it would be applied to the traffic aggregate whose load this test would be unlikely to influence to any significant degree.

[22] There appears to be no consideration of clock drift between the client and server during the duration of the test.

[23] By measuring only limiting performance of a fixed size stream of UDP packets, there is an implied assumption that there is a linear relationship between packet size and service time. It is also seems to be assumed that TCP packets will experience identical treatment.

[24] As this is not a continuous process, but a discrete one, it can have a large effect on the relative application outcome.

[25] To investigate this further would require laboratory experiments.

[26] The authors say "we are not interested in such low load conditions because there is no effective discrimination in such cases".

[27] Whose configuration would be independent of the particular ISP serving the end-user.

[28] Some UK retail ISP's Ts&Cs reserve the right to differentially treat certain classes of traffic during "periods of abnormal load", in order to maintain key services within their PRO.

The loss discrimination test requires an arbitrarily long duration since it cannot complete until 10 packets have been lost in each stream.

There seems to be a contradiction between the decision to focus on VoIP applications and the approach for inducing discrimination by loading the network, which is not the normal behaviour of such applications; indeed an ISP could easily classify such traffic as part of a DDOS attack.

It is acknowledged that some appearances of discrimination are due to routing changes and that this needs to be accounted for; such accounting does not seen to have been disclosed in the paper.

There does not appear to be a bulk deployment of this measurement approach, nor does it appear to be in active development. The paper's authors went on to create ShaperProbe (§ 2.3.5 on page 31) which is available on M-Lab, but this only measures throughput and its limitation, not delay and loss characteristics.

This technique seems unable to distinguish TM applied at different points on the path between the client and the server.

2.3.4. Glasnost

Glasnost is the work of M. Dischinger, M. Marcon, S. Guha, K. P. Gummadi, R. Mahajan and S. Saroiu at both the MPI-SWS (Max Planck Institute for Software Systems) and Microsoft Research in 2010 [8].

Aim

The aim of Glasnost is to enable users to detect if they are subject to traffic differentiation. The question that Glasnost tries to answer is whether an individual user's traffic is being differentiated on the basis of application, in order to make any differentiation along their paths transparent to them. This project particularly aims to reach a mass of non-technical users, while providing reliable results to each individual.

Framing the aim

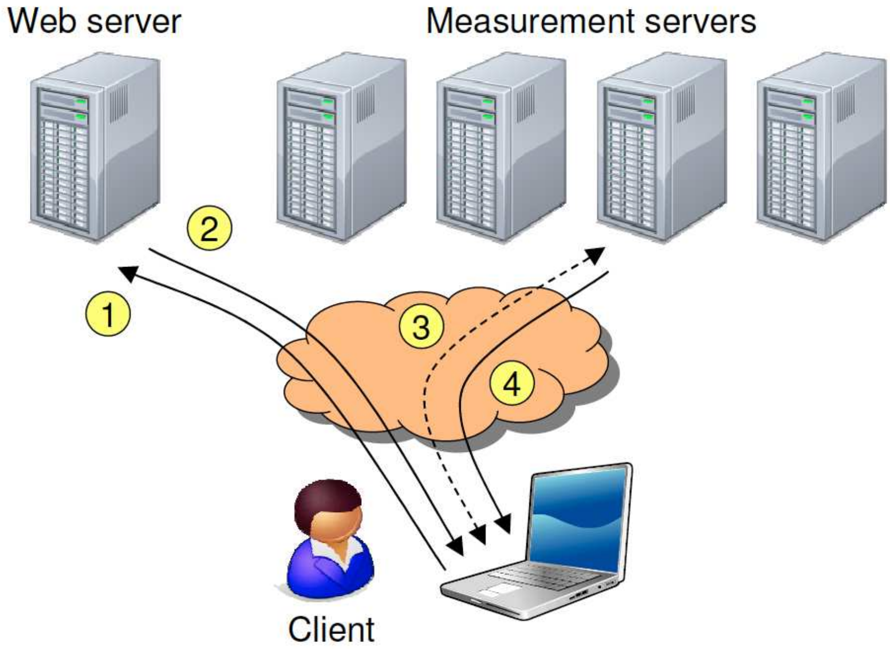
Glasnost detects the presence of differentiation based on its impact on application performance. It does this by determining whether flows exhibit different behaviour by application even when other potential variables are kept constant. The key assumptions are:

1. ISPs distinguish traffic flows on the basis of certain packet characteristics, in particular port number or packet contents;
2. ISPs may treat these distinguished flows to and/or from an individual user differently;
3. Such differential treatment can be detected by its impact on application performance;
4. Confounding factors[29] can be controlled or are sufficiently transient that a sequence of repeated tests will eliminate them, while not being so transient that they have an impact on one flow but not on the other;
5. Users may not have administrative privileges on the computers they use and are unable/unwilling to engage with technical issues.

The approach is to generate a pair of flows that are identical in all respects except one; this one respect is chosen as it is expected to trigger differentiation along the path. This is illustrated in Figure 2.6. Comparing the performance[30] of these flows is the means to determine whether differentiation is indeed present.

[29] Such factors include the user's operating system, especially its networking stack and its configuration, and other traffic, either from the user or other sources.

[30] In principle, various performance measures could be used, but in the current implementation, the only parameter measured is throughput of TCP flows.



(1) The client contacts the Glasnost webpage. (2) The webpage returns the address of a measurement server. (3) The client connects to the measurement server and loads a Java applet. The applet then starts to emulate a sequence of flows. (4) After the test is done, the collected data is analysed and a results page is displayed to the client.

Figure 2.5.: The Glasnost system
Reproduced from [8]

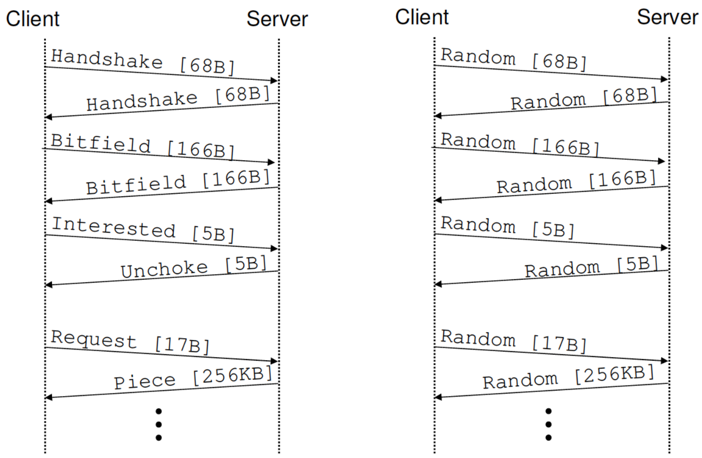
Implementation

The current implementation of Glasnost detects traffic differentiation that is triggered by transport protocol headers (i.e. port numbers) or packet payload. The tool works using a Java applet downloaded from a webpage. This acts as a client that opens a TCP session to communicate with a Glasnost server, as illustrated in Figure 2.5. This client/server service then runs pairs of emulated application flows back-to-back to detect throughput differentiation between them. In each pair the first uses the port number or packet payload that may be being differentiated against; the second uses random data intended to have all the same characteristics except that being tested for (e.g. non-standard port number and random packet contents, as illustrated in Figure 2.6). Upstream and downstream tests are "bundled" to make the tests complete faster and the tests are repeated several times to address the confounding factor of "noise" due to cross-traffic[31]. Experimental investigations on throughput led to a classification of cross-traffic as being one of the following:

- Consistently low;
- Mostly low;
- Highly variable;
- Mostly high.

Measurements that suggest cross-traffic is 'highly variable' or 'mostly high' are discarded.

[31]This means traffic contending in the multiplexing tree to the sink, as discussed in §A.1.



(a) BitTorrent flow (b) Reference flow

A pair of flows used in Glasnost tests. The two flows are identical in all aspects other than their packet payloads, which allows detection of differentiation that targets flows based on their packet contents.

Figure 2.6.: Glasnost flow emulation
Reproduced from [8]

Detectable TM techniques

TM techniques detectable by Glasnost would be those that impact the throughput of a TCP session for certain flows to/from a particular user. Thus techniques such as bandwidth sharing or prioritisation *between* users will not seem to be detectable. Rate-limiting of specific types of traffic should be detectable provided the limit is less than other constraints, such as the rate of the access link. If rate-limiting is being applied to a traffic aggregate (e.g. the total amount of P2P traffic rather than that of any particular user), then it will only be detectable if the aggregate rate exceeds the limit (i.e. it is dependent on the actions of other users of the network). Rate limiting that is applied only when the network is heavily loaded may not be detectable due to the rejection of measurements when cross-traffic is high or highly variable.

Discussion

While this method is capable of detecting differentiation against a single application by a single method, it seems to lack a coherent analysis of potential confounding factors. These are aggregated as "noise", which is dealt with by performing repeated tests[32]. The paper includes a discussion of false results (both positive and negative), quantified by an empirical method. However, claims for the robustness of the results are based on empirical analysis of a relatively small data set, and the assessment appears to be affected by assumptions and axiomatic beliefs (enumerated in Framing the aim above).

Significant emphasis is placed on the advantages of an active measurement approach, and the benefits of using emulated rather than actual applications. However this is likely to be an unfaithful reproduction of real application behaviour, as the timing of the application packet stream is not reproduced. Moreover, using TCP throughput measurements adds variability to the tests, due to the interaction of the Java VM with the specific OS TCP stack; thus two users connected to the same network endpoint could report different results. The paper makes

[32]The paper points out that limitations are imposed by end-user attention span, with the result that the length and number of iterations of the tests was reduced, which may compromise the statistical significance of the results.

strong claims of generality for this approach, while admitting that substantial compromises had to be made for the sake of user-friendliness. For example, in section 5.3 of the paper it is mentioned that new, shorter tests were implemented to increase test completion rates and combat problems caused by user impatience[33]. As part of this the tests for upstream and downstream directions were "bundled". It is unclear what is meant by this, but if it means that both upstream and downstream tests are carried out at the same time or with overlap, self-contention could add a confounding factor, in particular the interaction of TCP 'acks' and bulk elastic data flow behaviour.

While it is claimed that "Glasnost detects the presence of differentiation based on its impact on application performance", in appears the only type of application performance that is measured is achievable TCP throughput. This is relevant if the application in question is BitTorrent, but not if it has real-time characteristics, e.g. an interactive web session or VoIP.

The Glasnost design also tries to create an adaptable system that can be configured for novel management methods. This is laudable and a logical step but, given the potential variety of TM policies that might be applied, detecting all of them from a single end-point may swiftly prove to be infeasible. The construction of the detector itself and its apparent reliance on limited aspects of an application's performance seem to make the system's ability to generally distinguish differentiation questionable.

This technique appears unable to distinguish TM applied at different points on the path between the client and the server.

2.3.5. ShaperProbe

ShaperProbe was developed by P. Kanuparthy and C. Dovrolis at the Georgia Institute of Technology in 2011 [7].

Aim

The question that ShaperProbe tries to answer is whether a token bucket shaper (as described in § B.4.4 on page 73) is being applied to a user's traffic. This is intended to be an active measurement service that can scale to thousands of users per day, addressing challenges of accuracy, usability and non-intrusiveness.

Framing the aim

ShaperProbe tries to address this aim by asking whether a shaper kicks in once a certain (unknown) data transfer rate is reached. It first estimates the link rate, then sends bursts[34] of maximum-sized packets at a series of rising data rates (up to just below the estimated limiting rate). It looks for the point where the packet rate measured at the receiver drops off, by counting arrivals in a given interval (this is illustrated in Figure 2.7). If the delivered rate drops to a lower rate after a period of time, the presence of a token-bucket traffic shaper on the path is declared, and its token generation rate and bucket depth estimated, based on the amount of data sent before the rate dropped and the asymptotic rate.

Measured values are adjusted to smooth the rate-response curve. To minimise intrusiveness, probing is terminated early when either shaping is detected or packets are lost.

Implementation

The technique is to first use short UDP packet trains to get an estimate for the limiting link rate[35]. This is done by sending short trains of back-to-back maximum-sized packets

[33] The number of tests for each combination of port pairs was reduced to one. The remaining tests take 6 minutes.

[34] These bursts have constant spacing between their constituent packets.

[35] This seems to assume that these packet trains are short enough not to be affected by shaping themselves.

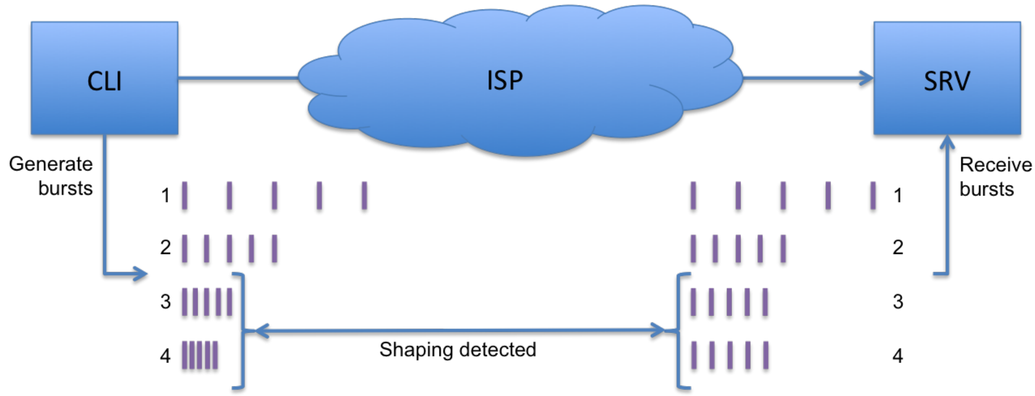


Figure 2.7.: ShaperProbe method

```
DiffProbe release. January 2012.
Shaper Detection Module.

Connected to server 4.71.254.149.

Estimating capacity:
Upstream: 2976 Kbps.
Downstream: 96214 Kbps.

The measurement will last for about 3.0 minutes. Please wait.
Checking for traffic shapers:

Upstream: No shaper detected.
Median received rate: 2912 Kbps.

Downstream: No shaper detected.
Median received rate: 59957 Kbps.

For more information, visit: http://www.cc.gatech.edu/~partha/diffprobe
```

Figure 2.8.: ShaperProbe sample output

and observing their arrival times[36]. The spacing of these packets at the receiver should be constant, given that packet sizes are constant in the offered load. However, the packet arrivals can be affected by experiencing non-empty queues. To deal with this, standard nonparametric rank statistics are applied to derive a “robust estimator” (note that this may differ from the allocated capacity - see Figure 2.8).

The total burst length and the threshold rate ratio for detection were chosen empirically, using a small sample, to maximise the detection rate (this is described in the Technical Report [15]).

The ShaperProbe client is a download-and-click user-space binary (no superuser privileges or installation needed) for 32/64-bit Windows, Linux, and OS X; a plugin is also available for the Vuze BitTorrent client. The non-UI logic is about 6000 lines of open-source code.

An example output from running the tool from a UK cable-connected endpoint is shown in Figure 2.8; note that this appears to seriously overestimate the allocated downstream rate of 60Mb/s (as advertised by the ISP and recorded by SamKnows).

The tool is deployed on M-Lab, which hosts the servers, and the tests reported in the paper were performed on a number of ISPs between 2009 and 2011.

[36] As previously discussed in footnote 19 on page 26.

Detectable TM techniques

Token bucket shapers with a sufficient bucket size should be detected but those which kick in very quickly may not be seen. False positive results could be caused by coupled behaviour, for example a large file download by another user of the same shared last-mile segment (e.g. cable segment), which would result in a drop in the received rate by the tool. Since results are discarded if any loss occurs, policers will not be detected.

Discussion

There is some analysis of the robustness of the results, using case studies where the ISPs had declared their shaping policies, but the vulnerability to 'cross traffic' (i.e. contention along the path between client and server) is unclear.

There are classes of traffic conformance algorithms that would seem to be undetectable using this approach, such as those proposed and used in ATM traffic management [16], and those in use in BRASs in UK networks[37]. Shaping, as detected here, is only likely to be deployed in systems that statistically share last-mile access capacity, as discussed in § B.6 on page 75.

The paper reports a false positive rate of 6.4%, but then claims a rate of less than 5% without apparent further justification.

This technique seems unable to distinguish TM applied at different points on the path between the client and the server.

2.3.6. ChkDiff

Chkdiff is a 2012 work of Riccardo Ravaioli and Guillaume Urvoy-Keller, of l'Université Nice Sophia Antipolis, and Chadi Barakat of INRIA [11].

Aim

The question that Chkdiff tries to answer is whether traffic is being differentiated on the basis of application. It attempts to do this in a way that is not specific to the application or to the discrimination mechanisms in use. Rather than testing for the presence of a particular TM method, this approach simply asks whether any differentiation is observable.

Framing the aim

In order to answer this question, this approach tries to observe user traffic in such a way as to detect whether specific flows have different performance characteristics when compared to the user's traffic as a whole. The key design principles are:

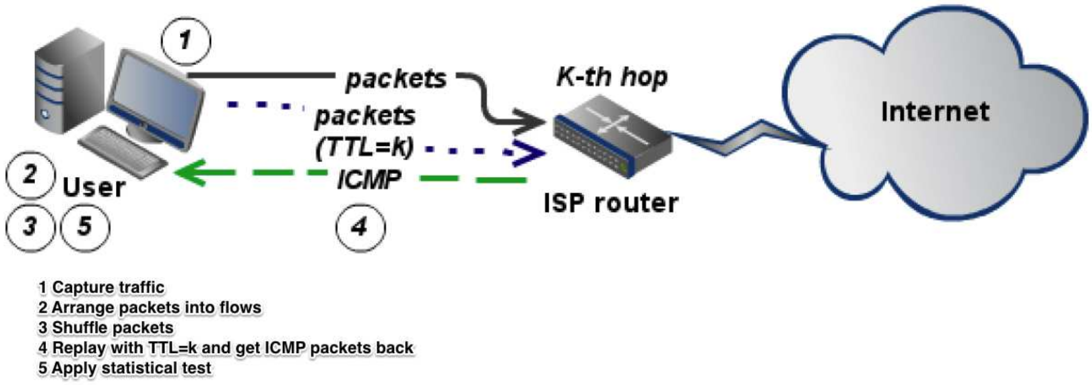
1. Use only user-generated traffic;
2. Leave user traffic unchanged;
3. Use the performance of the whole of the user's traffic as the performance baseline.

Implementation

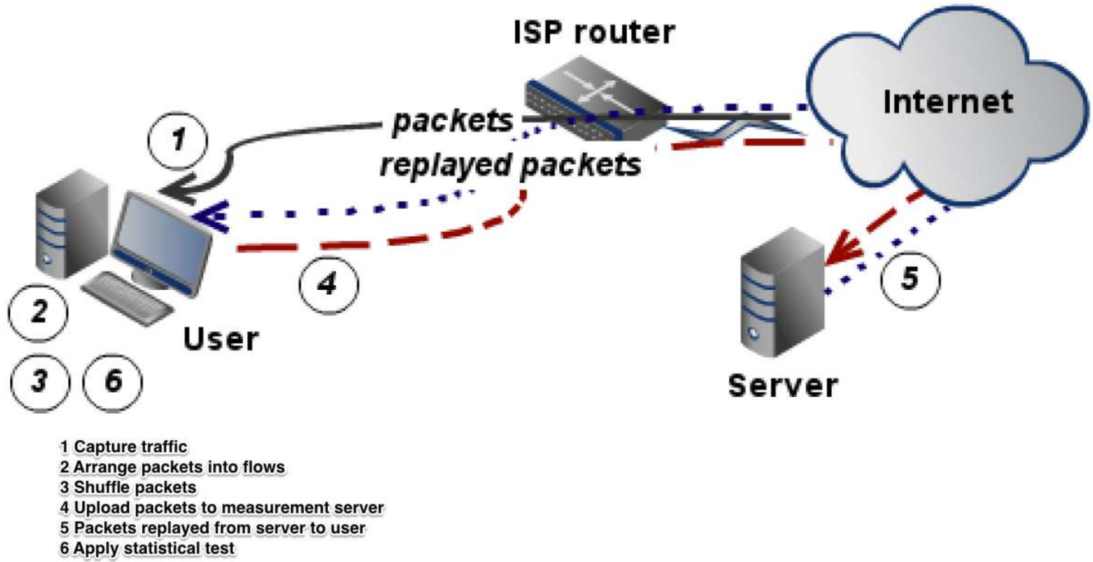
The process is represented in Figure 2.9 (note that the downstream component has not been implemented). The metric used in the upstream direction is the round-trip time (RTT) between the user and a selected router on their access ISP; the number of hops to the router is selected by modifying the TTL field. The process is:

1. Capture user traffic for a fixed time-window of a few minutes;

[37] Fully clarifying the range of applicability and limitations of this technique would require laboratory investigation.



(a) Upstream



(b) Downstream

Figure 2.9.: Chkdiff architecture
Reproduced from [11]

2. Classify the traffic into flows using the packet header information;
3. Generate a test by repeatedly picking packets from different flows at random, weighted by the overall volume of each flow;
4. Focus the measurement by setting the value of the TTL fields of the packets;
5. Apply a statistical test, by fitting delay histograms to a Dirichlet distribution.

User-generated packet traces are replayed with modified TTL fields, and the time to receive the ICMP response is measured[38]. Different flows are mixed by taking Bernoulli samples in order to invoke the PASTA property[39], and the results are compared for different flows on the basis of the distribution of response times (using histograms).

A downstream test is proposed using a similar system, in which arriving packets are captured at the client, and then uploaded to a server for replay. This has not been implemented.

Detectable TM techniques

This very general method would be able to detect delay differentiation between different flows, e.g. due to priority queuing or WFQ applied on a per-application or network host basis. However, it would be unable to detect differentiation on an individual end-user basis, since it relies on the aggregate performance of the user's traffic as a baseline. Thus, any differentiation that affects the user's traffic as a whole (e.g. a token bucket shaper as discussed in § B.4.4 on page 73) would not be able to be detected. Since packet loss is not measured, techniques that affect loss such as WRED could not be detected.

Discussion

By measuring the distribution of round-trip delays, this approach is very close to measuring differential ΔQ, so the aim of "application and differentiation technique agnosticism" is sound. Extending the method to include measuring loss, as proposed, would make their measure correspond more closely to ΔQ, except that it measures round-trip instead of one-way delays. By measuring delays to intermediate points, this approach laudably aims to localise rather than merely detect differentiation. The principal disadvantage of this method appears to be that it relies on the fidelity of the intermediate routers' ICMP response to the packet expiry. Generating ICMP responses is not a priority for routers, and so the response time is highly load-dependent; also the rate limitation on ICMP responses may have an impact on the scalability of the technique.

Applying this technique in the downstream direction would require a server to replay spoofed packets. This has not been implemented.

False positives and negatives do not seem to be well addressed in the paper, but Chkdiff was only in early development when it was written.

Overall this is a promising approach, and it is a pity that it does not seem to have been developed beyond a laboratory prototype.

2.3.7. Network Tomography

Network tomography is a body of work that takes a multi-point observational approach to measuring network performance [17, 18, 19].

Aim

Network tomography uses the 'performance' of packets traversing a network much as radiologic tomography uses the 'performance' of X-rays passing through the body. X-ray intensity is

[38] Note that this is the same technique used by NetPolice [10], discussed in § 2.3.1 on page 20.

[39] This means the results are robust against transient and phase-related effects.

modulated by the tissues passed through; packet performance is modulated by the path traversed. Using multiple ingress and egress points on the periphery of the network means this is seen as analogous to a CT scan of a body, in that distinct internal features become visible by combining multiple measurements. A recent paper by Zhang [20] explores the use of this approach for the detection of differential treatment of traffic.

Framing the aim

The approach is to start with a description of the network's connectivity at a link/path level, expressed as an adjacency matrix $\mathbf{A}$. This is combined with a vector of external observations $\overrightarrow{y}$, to infer a vector $\overrightarrow{x}$ of the link/path properties by solving the following system of equations:

$$\overrightarrow{y} = \mathbf{A} \cdot \overrightarrow{x}$$

In principle, if more than enough observations are available, the system can be solved using only a subset of them. The insight relevant to TM detection is that if different subsets of observations yield *different* results for any particular internal link/path, this could indicate the presence of some differential treatment[40]. By selecting the subsets of observations in different ways, insights might be gained as to the factors that trigger differential treatment. Useful subsets might be aspects of the path and/or association data (addressing, content), packet contents, etc..

Implementation

These papers have been written in the context of mathematical 'thought experiments', and where validation has been performed this has been done as simulations. No deployable tool has yet been produced.

Discussion

There appear to be several underlying assumptions. Firstly, this approach explicitly requires knowledge of the structure of the network at a link/path level, which may be hard to discover. It also seems to assume that the routing and link structure of the network is constant for the set of observations, which may not be the case given the dynamic nature of routing protocols. Secondly, there is an important requirement on the mathematical structure of the performance measure in order to validly solve the equations[41]. This means that the type of values that can be solved for do not seem to correspond to realistic performance measures[42]. In particular, ΔQ (discussed in §A.2.1) is not a simple scalar[43], so the particular solution process proposed in this body of literature could not be directly applied to it.

However, combined with an appropriate performance measure[44], this approach does represent a potential way forward for detecting TM effects. The tomographic approach supports not only detecting whether discrimination is performed on the basis of application or originator, but also the evaluation of differential service between customers. It could provide a scalable means of assessing whether classes of users were actually receiving the service that

[40] Zang et al express this as the system being "unsolvable"; they appear to be making the assumption that a "neutral network" will form a system of equations that are solvable, even if they are massively over-specified.

[41] In order to solve a system of equations, the values have to have a particular set of mathematical properties (such as those that hold for real numbers). Typically they must form a 'field' (see `http://en.wikipedia.org/wiki/Linear_equation_over_a_ring`) in order to form $\mathbf{A}^{-1}$ (the inverse of $\mathbf{A}$) so that $\mathbf{A}^{-1} \cdot \overrightarrow{y} = \mathbf{A}^{-1} \cdot \mathbf{A} \cdot \overrightarrow{x} = \overrightarrow{x}$ can be calculated.

[42] Adding average delays is not meaningful, nor is adding up 'congestion', for example.

[43] Mathematically, ΔQ is akin to a cancellative monoid, `http://en.wikipedia.org/wiki/Cancellative_semigroup`.

[44] Using a solution approach that is mathematically appropriate to such a performance measure.

they expected (for example whether 'premium' customers receive a markedly different service from 'standard' ones). Thus conformance to marketing claims and T&Cs may be able to be independently assessed.

The power of this approach is that it does not focus on a single metric of interest, e.g. throughput, but takes a general observational approach (much like NANO and Chkdiff, with which it might usefully be combined). Also it does not, by its nature, entail stressing the network infrastructure[45]. It could be done in an entirely passive way or make use of only low bandwidth test streams. All of these factors mean it could be deployed on a large scale. However, considerable further research would be required to develop a practical methodology; encouragingly, this is one area in which research seems to be ongoing.

[45] The approach taken by Glasnost and ShaperProbe is to by drive a path to saturation so that any differential treatments come into play and hence become measurable.

3. Traffic Management detection in an operational context

3.1. Introduction

In Chapter 2, various approaches to detecting the presence of differential traffic management were discussed. Most of these approaches are designed for sporadic use by individual end-users. In this chapter, the focus is on the operational behaviours and scalability of these detection approaches and their potential application and impact in an operational context (i.e. by actors other than individual end-users).

3.2. Review of TM detection techniques

It is inherently impossible to detect directly the specific application of differential treatment (other than by inspecting the configuration of network elements). Even when there is such an intention, it may not have any effect, depending on the particular circumstances of load, etc.. Thus the techniques listed in Table 3.1 do not directly detect traffic management, but rather attempt to infer its presence through structured observations. They look for differences in specific aspects of translocation performance, either directly by measuring delay or loss (though none measures both together) or indirectly by measuring the operational performance of TCP bulk transport.

Traffic Management detection literature, as surveyed in §2.3, typically starts from the assumption that discrimination is occurring and that the task is to detect it. Such presumed discrimination falls into one (or both) of two broad categories:

1. Restriction on the freedom of association - the ability to have access to a particular service, to a particular location (e.g. server) or from a particular location (e.g. client)[1]. This restriction can take one of several forms: e.g. port blocking, intercepting protocol behaviour to insert resets, or hijacking domain name resolution. Identification of the association can be done on the basis of the addressing in the packets[2], their ingress/egress ASNs and/or contents (i.e. using DPI);
2. Taking deliberate actions that impact the *performance* of some set of associations[3] identified as above. For example, limiting the transported load of traffic identified as P2P.

The approaches are structured to detect performance differences, typically measured end-to-end. They then aim to infer that these differences are caused by application of discriminatory queueing and scheduling somewhere along the path. This inference hinges on several factors:

- The nature of "discrimination". To discriminate, two steps are needed: firstly, a classification or choice needs to be made to distinguish packets belonging to one flow from those belonging to others; secondly, a difference needs to be applied in the treatment of the packet exchanges making up such flows. How this choice can be made is discussed in § 3.2.1 on the facing page;

[1] Firewalls are an expression of this freedom to associate, in particular the freedom to *not* associate.
[2] An example would be discarding all packets to or from a particular set of addresses when responding to a DDOS attack.
[3] This is done by increasing the ΔQ of the corresponding translocation.

- The underlying assumptions being made in the construction of the detection approach; these are discussed in § 3.2.2;
- The likely efficacy of such approaches in an adversarial context. Some of the aspects of this are explored from a "game" perspective in § 3.3 on the following page. Various forms of discriminatory practice can be envisaged that would not be detected by any of the techniques discussed in §2.3.

3.2.1. Technical aspects of flow differentiation

Packet flow discrimination can be done by classifying packets based on addressing information[4], the pattern of offered load, or a combination thereof. Note that devices have access to more 'address' information than just the IP source and destination contained within the packet itself. This can be explicit[5] or derived[6]: explicitly derived from the packet header[7], or based on an analysis of the SDU[8]. The pattern of offered load can be measured using a token-based scheme[9] or historical information (such as volume used over some previous period).

Only after classification has occurred can a particular queueing/scheduling choice be applied. From that choice, differential behaviour of the end-to-end packet flows can emerge (i.e. differential delivered ΔQ). That, in turn, can lead to differential protocol performance and application outcomes.

3.2.2. Underlying assumptions made in TMD techniques

The general assumption made in most TMD approaches is that TM is the cause of differentiation in service. This is a narrow approach that does not seek to understand the factors influencing the performance of applications and protocols, but rather aims to 'prove' the hypothesis that 'the ISP' is restricting the delivered service to some degree. This is done by trying to disprove the 'null hypothesis' that no differentiation is taking place. Thus TMD techniques typically fall into the general category of statistical hypothesis testing[10]. Such testing depends on being able to conclude that any differences in the resulting outcome can be unambiguously attributed to a constructed distinction between a 'test' and a 'control'. It is important to show that such differences are not due to some other 'confounding' factor that would result in false positive/negative results. In the absence of a comprehensive model of the factors affecting performance, the methodology is to control as many potential confounding factors as possible, and deal with others by means of statistics[11].

There are many possible confounding factors that seem to have not been taken fully into account by any of the approaches. One such factor is the inherent variability in the performance of PBSM, which leads a number of techniques to discard measurements when there is 'noise' due to contention (i.e. for which $\Delta Q_{|\mathbf{V}}$ is too large). However, as discussed in Appendix B, it is precisely in the allocation of $\Delta Q_{|\mathbf{V}}$ that the effects of TM are manifest. Thus many approaches to TMD deliberately ignore the circumstances in which TM is most likely to be active. Another implicit assumption is that occasional tests from self-elected end hosts can

[4]Note that classification on the basis of addressing information is effectively reverse-engineering the end-point association, endeavouring to identify some aspect of the 'parties' involved - such as application, provider and customer.

[5]This can be based on the VLAN, some virtual router function, or the physical port of reception/transmission.

[6]Derived information includes the originating/terminating/next-hop AS number.

[7]One example of this could be port numbers in the transport layer header.

[8]This is typically done by deep-packet inspection. Note that this becomes more difficult when packet contents are encrypted or otherwise modified, e.g. by compression.

[9]This is as described in Appendix B.4.4, where arrivals reduce the token pool that is being filled at a set rate; when the pool empties the stream is treated differently.

[10]`http://en.wikipedia.org/wiki/Statistical_hypothesis_testing`

[11]This can easily lead to assuming that correlation implies causation.

be expected to detect reliably differential traffic management. This would only be the case if such TM were applied uniformly.

A further assumption is that the underlying end-to-end performance (in the absence of any deliberate differentiation) is the same for the 'test' and 'control' experiment streams[12]. The effect of this is minimised when the packets for the two streams are interleaved.

Some techniques assume that ICMP responses from intermediate routers can be relied upon. However, ICMP was not intended to provide accurate performance data, and responses to pings or TTL exhaustion are entirely at the mercy of the processing load of the targeted router and its application of ICMP rate limiting.

In order to create repeatable tests, captured or emulated traces are often used[13], generally of TCP sessions. This implicitly assumes that actual application/protocol behaviour is not important. So, while TMD techniques are attempting to compare application outcomes (in particular protocol performance), some do so only by comparing differential treatment of TCP behaviour, which leads to information fidelity loss[14]. Furthermore, the protocol peer has specific implementation and parameter settings that may differ by application, and there may be other unknown factors such as loading and performance issues (e.g. power saving by the end device).

3.2.3. Comparison of main approaches

We classify the most interesting approaches by the following criteria:

Readiness Level To what extent the technique is available to be exploited;

Active or passive Whether the approach actively injects test packets or passively observes the existing traffic flow; if active, whether it relies on saturating the constraining link of the end-to-end path and an estimate of the traffic volume generated;

Detect based on What measured property of selected flows is used to detect discrimination;

TM types Which TM techniques the approach is designed to detect;

Target TM locations Where in the end-to-end path TM is being looked for;

Measurement duration How long an individual test may take;

Test traffic volume Estimated volume of traffic generated per test; note that this will in many cases depend on the sync rate of the end-user's line[15].

Supply Chain Localisation Ability to localise TM in a heterogenous digital supply chain.

Table 3.1 compares the different approaches on these criteria.

3.3. Likely efficacy of TMD in a UK context

Even where some correlation could be detected, the UK market (see Appendix C) is such that there often would not be a single administrative/management domain to which the discrimination can be attributed, as shown in Figure 1.1. The authors agree with the authors of [20] that detection of the location where traffic management is being deployed is as important as the detection of its existence. A clear issue-isolation process is required for any operational framework.

TM detection techniques have been mostly developed in North America, where the market structure differs from that of the UK. Where there is a single integrated supplier, as is typical

[12] This is to say that $\Delta Q^{A\leftrightarrow Z}$ is stationary over the period of measurement.

[13] With the exception of NANO that collects protocol data; this has the issue that it may leak privacy-related information, such as which servers were contacted.

[14] An example of this, and the consequences of it, can be found in [21].

[15] For example, a 10Mb/s DSL line delivers approximately 1MB/s of user-level data. Thus saturating such a link for one minute will consume 60MB.

Paper	**Readiness Level**	**Active or passive**	**Detect based on**	**TM types**	**Target TM locations**	**Test duration**	**Test traffic volume per test**	**Supply chain localisation**
NetPolice [10]	Deployed on PlanetLab during research	Active	Differential loss by AS number	Rate limiting	Tier 1 ISPs	2 hours	One ICMP packet/s per element tested	ISP exchange points only
NANO [9]	Deployed on PlanetLab and Emulab during research	Passive	TCP throughput and latency by association/ addressing	Various	Local ISP	Unknown	2.5kb/s per end-user for reported results	None
DiffProbe [12]	NS trials - then deprecated	Active Saturating	Differential delay distributions and differential loss by association/ addressing	Queuing and prioritisation	Whole path	15s minimum; many repetitions	Unbounded: 10s link saturation per test	None
Glasnost [8]	Deployed at scale (MLab)	Active Saturating	Differential throughput by association/ addressing	All affecting elastic throughput	Whole path	6 minutes	6 minutes of saturation per test	None
Shaper Probe [7]	Deployed at scale (MLab)	Active Saturating	Throughput variation over time per end-user	Rate limiting	Whole path	2-3 minutes	Variable: up to c. 1GB	None
ChkDiff [11]	Lab trials only	Mixed	Distribution of RTTs to intermediate router by association/ addressing	All delay affecting	All	c. 10 minutes?	Unknown	User-visible Layer 3 routers
Network Tomography	Only tested in simulation	Either	Performance measures over multiple paths by association/ addressing	All (depending on performance metric)	All	unknown	Unquantified but low	Good

Table 3.1.: Taxonomy of Traffic Management Detection Approaches.

in North America, establishing that discrimination is occurring somewhere on the path to the end-user is broadly sufficient to identify who is responsible, but when there are multiple administrative domains involved, as in the UK, the situation is more complex.

3.3.1. Offered-load-based differentiation

Differential service on the basis of offered load has been part of the contractual relationship at network boundaries since the inception of PBSM (e.g. ATM used this as the major basis of service differentiation). Control of the offered load by means of rate limiting is an essential element needed for stable operation of PBSM, and it is present at multiple locations[16]. There is extensive use of such limiting at management/administrative boundaries to manage both bills and costs.

Detection of the most limiting network egress point is feasible, e.g. ShaperProbe, though this technique does make the implicit assumption that network contention effects (which could create false results) are absent.

Detection of the presence of such rate/pattern limiting can be done at the receiving end point with a single-point measurement process[17], and could deliver measurements for each direction separately. As with all single-point measurement processes, there is no spatial isolation. i.e. it is not possible to say where along the path the limiting occurred. In this case, in order to apply a high load, traffic must be sent to a remote host, i.e. along an entire end-to-end path[18]. Without intermediate measurement points (i.e. multi-point measurement) there is no way to isolate which section of the path induces the most stringent limitation.

Several major UK network providers make these limits available, either in their commercial T&Cs (in the terms of "up to") or in their technical interfaces (i.e. ADSL sync rates and BRAS limiters). As each of these measures is an upper bound, which only apply when there are no other data transport quality impairment effects.

3.3.2. Association-based differentiation

Some differentiation may depend on the association, i.e. exactly what the communicating entities are (e.g. an end-host at a particular IP address - the user - communicating with a server in a particular domain, or using a particular protocol). All the TM detection techniques that were found are single-point measures of a composite effect, typically involving multiple administrative/management domains, two directions of flow and some computational element.

Epistemologically the best that such techniques can do is to detect some differential treatment of the traffic flows that will result in a different observed distribution of delay and loss for that composite set of effects. They may do this directly, either by passive observation (as by NANO, §2.3.2), or by active measurement (as by NetPolice, §2.3.1, and DiffProbe, §2.3.3), or indirectly by measuring the effects on the performance outcomes of an application (as by Glasnost, §2.3.4). NetPolice's inability to detect TM applied to individual users would make it of limited use for the detection of differential TM. Its key feature of distinguishing between differentiation applied by backbone ISPs can probably be addressed more systematically by using a variant of network tomography (discussed in §2.3.7).

The majority of approaches endeavour to "prove" that application-based differentiation is occurring on traffic to/from a particular end user. In contrast, network tomography-based approaches would use a more general strategy that may be a better fit for use for the detection of differential TM. Additionally, such approaches would have benefits in terms of scalability and localisation.

[16]Given that every network interface is, in effect, a rate limiter, rate limiting could be said to be everywhere.

[17]This means observing any particular flow at a single point in its journey. There may be multiple measurement locations, but each of them is a single point measure. This means that all the techniques discussed here have no spatial localisation.

[18]Techniques to 'probe' intermediate routers using ICMP responses are inherently rate-limited.

The reviewed techniques may detect the existence of differential traffic treatment, but not pinpoint its location (with the exception of network tomography-type approaches); nor are they reliably able to assure the absence of such treatment due to the sporadic nature of the tests and the effect of confounding factors. Localisation might be addressed by mandating the installation of measurement points at suitable administrative boundaries, rather than relying entirely on measurements performed from the edge of the network.

3.3.3. Cost of the detection process

A common misconception is that additional load 'costs nothing', however wide-scale use of the saturating active methods could place a significant load on the network as a whole. For example, a single test on a 60Mbit/s connection taking several minutes, represents the load of several hundred average broadband users over that period. Although the assumption is that network traffic has no marginal cost, anecdotal evidence suggests that test traffic can be a significant factor driving capacity upgrades [22]. NANO does not have this issue (it is passive) and network tomography approaches could use either passive or low data rate active analysis[19].

3.3.4. TM detection techniques as proxy for user experience impairment

Glasnost and ShaperProbe are the only techniques that appear widely deployed (using M-Lab[20]), and both are focused on bandwidth "impairment". ShaperProbe does this at the uni-directional packet flow level: it is about capping the "up to" speed and does not aim to detect differential treatment based on association, only offered load. Glasnost does this at the bi-directional application outcome level; although the Glasnost paper implies that it can emulate (via synthetic behaviour) multiple applications, examination of the information available via M-Lab[21] shows that this test approach is only suitable for bulk data transfers (transfers that try to saturate the path to the end user) whose time-to-complete is more than 10 seconds. Thus this is not a suitable proxy for many user interactions, which are either short-lived (getting email, interacting with Twitter or Facebook), or have different usage patterns, like video streaming (which may last a longer time). Typical video streaming (e.g. YouTube) is not a bulk data transfer, because it is not endeavouring to saturate the path, but rather aiming to ensure that the play-out buffer does not empty to maintain the continuity of the video delivery. Other types of video streaming such as DASH or iPlayer do use TCP (via HTTP) to download 'chunks' of content. However, in this case maximising the TCP peak transfer rate can have a negative impact on application performance, by downloading a chunk so quickly that the TCP connection closes down before the next chunk is started. Once again, the details of the application behaviour matter.

Scrutiny of the M-Lab data for 2013 does not generate great confidence in the reliability or efficacy of these methods: the data set is actually quite small, and, because tests require active participation by end-users, the sample is inherently biased.

The set of ways in which TM techniques that could be differentially/prejudicially applied is much greater than the set that the available tools could detect. The authors can imagine several ways in which, for example, Glasnost could be 'gamed'[22].

[19] There are distinct advantages to using low data rate active analysis. By exploiting the PASTA principle, as used by ChkDiff, the data rate could be very low - a few bits per second. The active data would not have any particular privacy issues in that it would not contain any information that can be tied back to the user's activity, *except* for the induced delay and loss experienced.

[20] M-Lab hosts are generally located in academic institutions, however, so would not be representative of a typical consumer experience.

[21] `http://broadband.mpi-sws.org/transparency/createtest.html`

[22] The problem of applying a measure whose optimisation actually benefits the end-user is not dissimilar to the problem of creating a CPU benchmark that reflects real application performance; see for example `http://goo.gl/S6sZd7`.

The absence of an established baseline makes it impossible to detect discrimination on a per-user basis (or sub-set of users). Furthermore the absence of detected prejudicial treatment does not imply the received service is going to be fit for any intended purpose, such as video streaming, VoIP conversation or gaming.

4. Conclusions and recommendations

4.1. Conclusions

The success of packet-based statistically-multiplexed networks such as the Internet is dependent on sharing resources dynamically. This dynamic sharing is ubiquitous, occurring at every WiFi access point, mobile base station and switch/router port. Each of these multiplexing points allocates its resources in response to the instantaneous demand placed upon it, which can typically exceed the available supply. The result depends on the sharing mechanism employed, its configuration, and the pattern of the demand (as discussed in some detail in Appendix B). Whether the outcome is 'biased' or 'fair' depends on many factors, including:

- The nature or aspect of the resource being shared (e.g. ingress to versus egress from a buffer);
- The pattern of the demand;
- The configuration of the sharing mechanism; and
- The exact definition of 'fairness' (per packet? per flow? per application? per outcome? per user? etc.).

Insofar as the outcome depends on the configuration of the sharing mechanism, any configuration may be called 'traffic management' (TM). TM may be used to maintain the stability of network services by creating outcomes that are deliberately 'unfair'. For example, it might be 'fair' for a temporary overload to cause equal packet loss and delay across all flows, but where some of those flows are essential to maintain the operation of the network such 'fairness' is undesirable. TM may also be used to select one form of 'fairness' over another, for example, to ensure that all users receive a similar level of service, even when some are applying much higher levels of demand than others.

The emergent effects of many multiplexing points joined in a network are complex; consequently so is the relationship between desired outcomes and actual behaviour[1]. What ultimately matters to any application is the probability distribution of loss and delay in the delivery of its packets; this may be influenced by TM but not completely controlled by it. It is this delivered distribution[2] that determines user satisfaction; how this is achieved is of little concern to either end-users or their content and service suppliers - except when it is unsatisfactory. Poor performance may have many causes, including the overall network architecture and topology, capacity planning and in-life management. 'Traffic Management' is only part of the equation.

Presumably for this reason, traffic management detection (TMD) has been pursued almost entirely from an academic perspective[3]. Given the complexity of the relationship between desired outcomes and actual behaviour, inferring an intention from observed outcomes is effectively impossible. Rather than trying to address this general problem, most TMD starts from assumed intentions mediated by assumed particular TM techniques and then attempts to deduce whether or not certain observations are consistent with such assumptions. However, even positive results do not prove a deliberate intent to introduce bias; given the overall

[1]Further, laboratory-based study would be required to elucidate this relationship further. It may be possible to quantify 'typical' behaviour, so that unusual circumstances meriting investigation, for example by TMD, can be detected.

[2]Which we refer to as 'quality attenuation' and designate 'ΔQ'.

[3]Initial interest from M-Lab (supported by Google) has diminished in the last few years.

complexity of relating intentions to outcomes, demonstrating a differential outcome does not demonstrate an intent to produce that outcome.

Most research completed in this area (explored in Chapter 2) has been undertaken from the perspective of allocating responsibility for both quality of experience and use of traffic management in single, vertically-integrated suppliers. These approaches might not be suitable in the UK due to its heterogenous broadband delivery structure, detailed in Appendix C; even if it could be shown that some users or applications were being differentially treated, there is (in most cases) no single administrative entity that can be shown to be responsible. Some approaches attempt to localise the TM by using responses from intermediate routers; apart from the potential inaccuracy of this method, any attempt at large-scale deployment risks hitting the limits imposed on such responses[4].

Table 4.1 summarises table 3.1 with respect to the criteria set out in §1.4.1, using the legend that '✔' means a requirement is met; a '✕' means that it is not met; a '—' means that it is partially met; and a '?' means that there is insufficient evidence to reach a reliable conclusion. Reliability of the methods is essentially unknown because, while most of the papers make estimates of their technique's reliability, there has been no independent and uniform confirmation of these claims.

Technique	**Localisation**	**Reliability**	**Scalability**
NetPolice	—	?	—
NANO	✕	?	✔
Diffprobe	✕	?	✕
Glasnost	✕	?	✕
ShaperProbe	✕	?	✕
ChkDiff	—	?	✔
Network Tomography	✔	?	✔

Table 4.1.: Comparison of techniques with criteria

None of the TMD methods studied satisfy all the key attributes that would make them suitable for effective practical use. In particular, those that are currently in active deployment generate significant volumes of traffic, which would risk damaging the QoE of other users if applied widely, and incur costs to the service providers of carrying this traffic; thus they may be unsuitable for large-scale use. The reliability of these tools would require further study, using a uniform test environment in which their performance could be objectively compared.

It is easy to envisage TM policies that would not be detectable by any of the methods analysed, and in any case, TMD techniques that test for specific configurations of specific TM mechanisms risk being rendered rapidly obsolete by new TM approaches and more sophisticated service provider policies[5]. The introduction of SDN, as discussed in [23], makes it likely that TM polices may be reconfigured on a timescale much shorter than any of the available tools can obtain statistically reliable results. It is not clear where the effort would come from to update TMD techniques or to develop new ones, particularly since the focus of academic interest appears to have moved elsewhere. Finally, these tools are limited in that they aim only to detect the presence of differential (intra-user) traffic management, as the detection of non-differential traffic management (inter-user or aggregate) was not their goal.

These tools are not sufficient to enable effective detection and location of TM application along a fragmented digital delivery chain such as that in the UK. Our conclusion is thus that no tool or combination of tools currently available is suitable for effective practical use.

[4]Indeed, service providers might well conclude that their routers were under attack and thus decide to disable such responses altogether.

[5]Only NANO and Chkdif may be sufficiently general to overcome this problem.

4.2. Recommendations

TMD sits within a wider context of ensuring that internet service provision satisfies suitable criteria of fitness-for-purpose, transparency and fairness. Confirming such properties is challenging because of the inherently statistical nature of packet-based networks, and is further complicated by the heterogeneity of the digital supply chain. The absence of differential traffic management does not, by itself, guarantee fairness, nor does fairness guarantee fitness-for-purpose. TMD is thus, at best, one component of an overall solution for measuring network service provision. However, it could be used to help establish transparency; for example, if TM policies to be used on end-user traffic were published, their implementation could be independently verified.

Another difficulty in measuring fairness and fitness-for-purpose of network service provision is the application-dependent relationship between network performance and application outcomes (discussed in Appendix A). This means that particular differences in performance may or may not matter to end-users, depending on the applications they are using. The choice of application also determines which aspects of the delivered performance are significant[6]. TMD thus risks highlighting aspects of service provision that are largely irrelevant, while overlooking others that could have a significant impact, depending on the applications in use. This is a subject for further study.

TMD needs to be considered in relation to a broader framework for evaluating network performance. This framework should encompass two aspects. The first would be application-specific demands, captured in a way that is unbiased, objective, verifiable and adaptable to new applications as they appear. This could be used to ascertain the demand profile of key network applications, which would give operators more visibility of what performance they should support, and OTT suppliers encouragement to produce "better" applications (imposing a lower demand on the network). The second would be a system of measurement for service delivery that could be unequivocally related to application needs. This would be necessary if one wished to know if a particular network service was fit-for-purpose with respect to an particular application. This measurement system would need to deal with the heterogeneous nature of the supply chain by reliably locating performance impairments whilst avoiding unreasonable loads on the network. Due to significant boundaries along the end-to-end path, responsibility could only be ascribed to commercial entities if these needs were met. A development of the tomographic approaches discussed in §2.3.7, combined with a generic network performance measure such as ΔQ (outlined in Appendix A) has the potential to do this. TMD could then become a way to fill in any gaps in this overall framework[7].

Collection and publication of data within such a framework could have a transformative effect on the broadband market in the UK and beyond. Ofcom's publication of performance tables has already significantly benefited the market situation. Further benefit may be gained by enhancing this with richer data relating to application needs and complete network performance (beyond bandwidth measures). Users could then be empowered to choose applications that were appropriate for their network service[8]. Conversely users could choose network services that were fit for the applications they want to use[9]; if there were any interest in selecting network services that additionally did or did not apply specific forms of TM, then TMD would have a role.

More work is needed to better mange the relationship between supply, demand and delivered quality. This should address the systemic issue of the lack of feedback on demand, either

[6] VoIP is more sensitive to delay while VoD is typically more sensitive to loss, for example.

[7] How much benefit there would be in checking conformance to criteria that have no significant impact on end-user application performance is debatable.

[8] For example, a user whose service was known to have significant variation in latency could choose the online gaming platform that was least sensitive to this.

[9] For example, a user interested in a streaming video service might prefer a service with sufficient throughput and stable translocation characteristics over one with much higher throughput but occasional variations that might cause playout glitches.

to consumers (encouraging them to time shift demand, making better use of spare capacity) or to application producers (to make applications more efficient). Consistency of supply can be addressed with an appropriate measurement framework, as discussed above. Finally, we recommend investigating how a "quality floor[10]" could be maintained, perhaps requiring short-timescale incentives[11] such as some form of Pigovian tax[12].

[10] I.e. a bound on the end-to-end quality attenuation.

[11] This is needed because the timescales on which customers can switch are far too long compared with the timescales on which bad-actors could exploit them.

[12] http://en.wikipedia.org/wiki/Pigovian_tax

Bibliography

[1] Ofcom Commercial Team. Consultancy framework mini competition: A study of traffic management detection methods and tools mc no: Mc/316. Restricted Tender, February 2014.

[2] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. Number ISBN 0-252-72548-4. Univ of Illinois Press, 1949.

[3] Ofcom. *Ofcom's approach to net neutrality*, 2011.

[4] Guidelines for Quality of Service in the scope of Net Neutrality. Technical Report BoR (12) 32, BEREC, May 2012.

[5] Monitoring quality of internet access services in the context of net neutrality. Technical Report BoR (14) 24, BEREC, March 2014.

[6] Jeremy Klein, Jonathan Freeman, Rob Morland, and Stuart Revell. Traffic management and quality of experience. Technical report, Ofcom/Technologia, April 2011.

[7] Partha Kanuparthy and Constantine Dovrolis. Shaperprobe: End-to-end detection of isp traffic shaping using active methods. pages 473–482, 2011. URL: `http://www.measurementlab.net/measurement-lab-tools#tool5`, `doi:10.1145/2068816.2068860`.

[8] Marcel Dischinger, Massimiliano Marcon, Saikat Guha, P Krishna Gummadi, Ratul Mahajan, and Stefan Saroiu. Glasnost: Enabling end users to detect traffic differentiation. In *NSDI*, pages 405–418, 2010.

[9] Mukarram Bin Tariq, Murtaza Motiwala, Nick Feamster, and Mostafa Ammar. Detecting network neutrality violations with causal inference [online]. 2009. URL: `http://noise-lab.net/projects/old-projects/nano/`.

[10] Ying Zhang, Zhuoqing Morley Mao, and Ming Zhang. Detecting traffic differentiation in backbone isps with netpolice. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 103–115. ACM, 2009.

[11] Riccardo Ravaioli, Chadi Barakat, and Guillaume Urvoy-Keller. Chkdiff: Checking traffic differentiation at internet access. In *Proceedings of the 2012 ACM Conference on CoNEXT Student Workshop*, CoNEXT Student '12, pages 57–58, New York, NY, USA, 2012. ACM. URL: `http://doi.acm.org/10.1145/2413247.2413282`, `doi:10.1145/2413247.2413282`.

[12] Partha Kanuparthy and Constantine Dovrolis. Diffprobe: Detecting isp service discrimination. In *IEEE Conference on Computer Communications (INFOCOM)*, San Diego, CA, USA, 2010.

[13] Kevin Arceneaux, Alan S. Gerber, and Donald P. Green. A cautionary note on the use of matching to estimate causal effects: An empirical example comparing matching estimates to an experimental benchmark. *Sociological Methods & Research*, 39(2):256–282, 2010.

[14] C. Dovrolis, D. Moore, and P. Ramanathan. Packet Dispersion Techniques and Capacity Estimation. *IEEE/ACM Transactions on Networking*, 12(6):963–977, Dec 2004.

[15] Partha Kanuparthy and Constantine Dovrolis. End-to-end detection of isp traffic shaping using active and passive methods. Technical report, Technical Report, Georgia Tech, 2011. http://www. cc. gatech. edu/˜ partha/shaperprobe-TR. pdf, 2011.

[16] Natalie Giroux and Sudhakar Ganti. *Quality of Service in ATM Networks*. Prentice Hall PTR, 1999.

[17] Rui Castro, Mark Coates, Gang Liang, Robert Nowak, and Bin Yu. Network tomography: recent developments. *Statistical science*, pages 499–517, 2004. URL: `http://projecteuclid.org/euclid.ss/1110999312`, `doi:doi:10.1214/088342304000000422`.

[18] Earl Lawrence, George Michailidis, Vijay Nair, and Bowei Xi. Network tomography: A review and recent developments. *Ann Arbor*, 1001:48109–1107, 2006.

[19] Yiyi Huang, Nick Feamster, and Renata Teixeira. Practical issues with using network tomography for fault diagnosis. *ACM SIGCOMM Computer Communication Review*, 38(5):53–58, 2008.

[20] Zhiyong Zhang, Ovidiu Sebastian Mara, and Katerina Argyraki. Network neutrality inference. In *Proceedings of the ACM SIGCOMM Conference*, 2014. URL: `http://infoscience.epfl.ch/record/186414/files/neutralityInference_1.pdf`.

[21] Systems Research Lab. Apology: Broadband network management [online]. URL: `http://systems.cs.colorado.edu/mediawiki/index.php/Broadband_Network_Management` [cited 2014/05/05].

[22] Anonymous. Private communication. commercially confidential, 2008.

[23] Fujitsu. Carrier software defined networking (sdn). Technical report, OfCom, March 2014.

[24] Razvan Beuran. *Mesure de la qualité dans les réseaux informatiques.* PhD thesis, Bucharest, Polytechnic Inst. and St. Etienne U., 2004.

[25] Chris J Vowden and Laura Lafave. Analysis of composed M/D/1/K networks. In *UKPEW'01: proceedings of 17th annual UK performance engineering workshop*, 2001.

[26] Aleksandar Kuzmanovic and Edward W Knightly. Low-rate tcp-targeted denial of service attacks: the shrew vs. the mice and elephants. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 75–86. ACM, 2003.

[27] Keith Winstein and Hari Balakrishnan. Tcp ex machina: Computer-generated congestion control. *SIGCOMM Comput. Commun. Rev.*, 43(4):123–134, August 2013. URL: `http://doi.acm.org/10.1145/2534169.2486020`, `doi:10.1145/2534169.2486020`.

[28] Leonard Kleinrock. A conservation law for a wide class of queueing disciplines. *Naval Research Logistics Quarterly*, 12(2):181–192, 1965.

[29] Frank Kelly. Notes on effective bandwidth. *Stochastic networks: theory and applications*, pages 141–168, 1996.

[30] A Arulambalam, Xiaoqiang Chen, and N. Ansari. Allocating fair rates for available bit rate service in atm networks. *Communications Magazine, IEEE*, 34(11):92–100, Nov 1996. `doi:10.1109/35.544198`.

[31] J.W. Roberts. A survey on statistical bandwidth sharing. *Computer Networks*, 45(3):319 – 332, 2004. In Memory of Olga Casals. URL: `http://www.sciencedirect.com/science/article/pii/S1389128604000544`, `doi:http://dx.doi.org/10.1016/j.comnet.2004.03.010`.

[32] Cisco Tech Notes. Comparing traffic policing and traffic shaping for bandwidth limiting. *Document ID*, 19645.

[33] William Lehr, Steven Bauer, Mikko Heikkinen, and David Clark. Assessing broadband reliability: Measurement and policy challenges. In *Research Conference on Communications, Information and Internet Policy, Arlington, VA*, 2011.

[34] Steven Bauer, David Clark, and William Lehr. Powerboost. In *Proceedings of the 2nd ACM SIGCOMM workshop on Home networks*, pages 7–12. ACM, 2011.

[35] Marcel Dischinger, Andreas Haeberlen, Krishna P Gummadi, and Stefan Saroiu. Characterizing residential broadband networks. In *Internet Measurement Comference*, pages 43–56, 2007.

[36] Myles Hollander and Douglas Wolfe. *A.(1973). Nonparametric Statistical Methods.* John Wiley and Sons, New York, 1979.

[37] Karthik Lakshminarayanan and Venkata N Padmanabhan. Some findings on the network performance of broadband hosts. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pages 45–50. ACM, 2003.

[38] Guohan Lu, Yan Chen, Stefan Birrer, Fabián E Bustamante, Chi Yin Cheung, and Xing Li. End-to-end inference of router packet forwarding priority. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pages 1784–1792. IEEE, 2007.

[39] Ratul Mahajan, Ming Zhang, Lindsey Poole, and Vivek S Pai. Uncovering performance differences among backbone isps with netdiff. In *NSDI*, pages 205–218, 2008.

[40] Mukarram Bin Tariq, Murtaza Motiwala, and Nick Feamster. Nano: Network access neutrality observatory. 2008.

[41] George Varghese. *Network Algorithmics: an interdisciplinary approach to designing fast networked devices.* Morgan Kaufmann, 2005.

[42] Udi Weinsberg, Augustin Soule, and Laurent Massoulie. Inferring traffic shaping and policy parameters using end host measurements. In *INFOCOM, 2011 Proceedings IEEE*, pages 151–155. IEEE, 2011.

[43] Marcel Dischinger, Alan Mislove, Andreas Haeberlen, and Krishna P Gummadi. Detecting bittorrent blocking. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pages 3–8. ACM, 2008.

[44] EFF "Test Your ISP" Project. URL: `https://www.eff.org/testyourisp`.

[45] Nikolaos Laoutaris and Pablo Rodriguez. Good things come to those who (can) wait. In *Proc. of ACM HotNets*. Citeseer, 2008.

[46] Vuze: Bad ISPs [online]. URL: `http://wiki.vuze.com/w/Bad_ISPs` [cited 2014/05/05].

[47] M-Lab [online]. URL: `http://www.measurementlab.net` [cited 2014/05/05].

[48] The ICSI Netalyzr [online]. URL: `http://netalyzr.icsi.berkeley.edu/` [cited 2014/05/05].

[49] John Markoff. 'neutrality' is new challenge for internet pioneer [online]. September 2006. URL: `http://www.nytimes.com/2006/09/27/technology/circuits/27neut.html?_r=1&oref=slogin` [cited 2014/05/02].

[50] Brad Stone. Comcast: We're delaying, not blocking, BitTorrent traffic [online]. October 2007. URL: `http://bits.blogs.nytimes.com/2007/10/22/comcast-were-delaying-not-blocking-bittorrent-traffic/?_php=true&_type=blogs&_r=0` [cited 2014/05/02].

[51] The Associated Press. F.T.C. Urges Caution on Net Neutrality [online]. June 2007. URL: `http://www.nytimes.com/2007/06/28/technology/28net.html`.

[52] The Associated Press. F.C.C. Chairman Favors Penalty on Comcast [online]. July 2008. URL: `http://www.nytimes.com/2008/07/11/technology/11fcc.html` [cited 2014/05/02].

[53] Vern Paxson, Andrew K Adams, and Matt Mathis. Experiences with nimi. In *Applications and the Internet (SAINT) Workshops, 2002. Proceedings. 2002 Symposium on*, pages 108–118. IEEE, 2002.

[54] Planet Lab [online]. URL: `http://www.planet-lab.org/` [cited 2014/05/05].

[55] Neil Spring, David Wetherall, and Tom Anderson. Scriptroute: a public internet measurement facility. In *Proceedings of the 4th conference on USENIX Symposium on Internet Technologies and Systems-Volume 4*, pages 17–17. USENIX Association, 2003.

[56] Velocix (Alcatel-Lucent) [online]. URL: `http://www.velocix.com/` [cited 2014/05/05].

[57] Vuze network status monitor. Technical report. URL: `http://plugins.vuze.com/plugin_details.php?plugin=aznetmon` [cited 2014/05/05].

[58] Ying Zhang, Z Morley Mao, and Ming Zhang. Ascertaining the reality of network neutrality violation in backbone isps. In *Proc. of ACM HotNets-VII Workshop*, 2008.

[59] David Andersen, Hari Balakrishnan, Frans Kaashoek, and Robert Morris. Resilient overlay networks. Master's thesis, 2001.

[60] Robert Beverly, Steven Bauer, and Arthur Berger. The internet is not a big truck: toward quantifying network neutrality. In *Passive and Active Network Measurement*, pages 135–144. Springer, 2007.

[61] Canadian radio-television and telecommunications commission 2008-11-20 - #: 8646-c12-200815400 - public notice 2008-19 - review of the internet traffic management practices of internet service providers [online]. November 2008. URL: `http://crtc.gc.ca/PartVII/eng/2008/8646/c12_200815400.htm`.

[62] Yu-Chung Cheng, Urs Hölzle, Neal Cardwell, Stefan Savage, and Geoffrey M Voelker. Monkey see, monkey do: A tool for tcp tracing and replaying. In *USENIX Annual Technical Conference, General Track*, pages 87–98. Boston, MA, USA, 2004.

[63] COMCAST. Attachment b: Comcast corporation description of planned network management practices to be deployed following the termination of current practices [online]. 2008. URL: `http://downloads.comcast.net/docs/Attachment_B_Future_Practices.pdf`.

[64] Weidong Cui, Marcus Peinado, Karl Chen, Helen J Wang, and Luis Irun-Briz. Tupni: Automatic reverse engineering of input formats. In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 391–402. ACM, 2008.

[65] The DIMES Project [online]. URL: `http://www.netdimes.org/`.

[66] Nicholas P Jewell. *Statistics for epidemiology*. CRC Press, 2004.

[67] Keynote homepage [online]. URL: `http://www.keynote.com/` [cited 2014/05/05].

[68] Diane Lambert and Chuanhai Liu. Adaptive thresholds: Monitoring streams of network counts. *Journal of the American Statistical Association*, 101(473):78–88, 2006.

[69] Harsha V Madhyastha, Tomas Isdal, Michael Piatek, Colin Dixon, Thomas Anderson, Arvind Krishnamurthy, and Arun Venkataramani. iPlane: An information plane for distributed services. In *Proceedings of the 7th symposium on Operating systems design and implementation*, pages 367–380. USENIX Association, 2006.

[70] Matt Mathis, John Heffner, Peter ONeil, and Pete Siemsen. Pathdiag: automated tcp diagnosis. In *Passive and Active Network Measurement*, pages 152–161. Springer, 2008.

[71] Nate Anderson. Cox ready to throttle P2P, non "time sensitive" traffic [online]. January 2009. URL: `http://arstechnica.com/tech-policy/2009/01/cox-opens-up-throttle-for-p2p-non-time-sensitive-traffic/` [cited 29/04/2014].

[72] Judea Pearl. *Causality: models, reasoning and inference*, volume 29. Cambridge Univ Press, 2000.

[73] Charles Reis, Steven D Gribble, Tadayoshi Kohno, and Nicholas C Weaver. Detecting in-flight page changes with web tripwires. In *NSDI*, volume 8, pages 31–44, 2008.

[74] Joel Sommers, Paul Barford, Nick Duffield, and Amos Ron. Accurate and efficient sla compliance monitoring. *ACM SIGCOMM Computer Communication Review*, 37(4):109–120, 2007.

[75] Mukarram Tariq, Amgad Zeitoun, Vytautas Valancius, Nick Feamster, and Mostafa Ammar. Answering what-if deployment and configuration questions with wise. In *ACM SIGCOMM Computer Communication Review*, volume 38, pages 99–110. ACM, 2008.

[76] Larry Wasserman. *All of statistics: a concise course in statistical inference*. Springer, 2004.

[77] Andy C Bavier, Mic Bowman, Brent N Chun, David E Culler, Scott Karlin, Steve Muir, Larry L Peterson, Timothy Roscoe, Tammo Spalink, and Mike Wawrzoniak. Operating

systems support for planetary-scale network services. In *NSDI*, volume 4, pages 19–19, 2004.

[78] TelecomTV One. Its back to 'pipes' and 'free rides': Internet neutrality under attack (again) [online]. June 2009. URL: `http://www.telecomtv.com/comspace_newsDetail.aspx?n=45072&id=e9381817-0593-417a-8639-c4c53e2a2a10` [cited 2014 04 29].

[79] BT heavily throttling BBC, all video [online]. June 2009. URL: `http://fastnetnews.com/dslprime/42-d/1758-bt-heavily-throttling-bbc-all-video` [cited 29/04/2014].

[80] Internet 2 Performance tools [online]. URL: `http://www.internet2.edu/products-services/performance-monitoring/performance-tools/` [cited 29/04/2014].

[81] Ian Clarke. A distributed decentralised information storage and retrieval system. Master's thesis, University of Edinburgh, 1999.

[82] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: Simplified data processing on large clusters, osdi04: Sixth symposium on operating system design and implementation, san francisco, ca, december, 2004. *S. Dill, R. Kumar, K. McCurley, S. Rajagopalan, D. Sivakumar, ad A. Tomkins, Self-similarity in the Web, Proc VLDB*, 2004.

[83] ED FELTEN. Three flavors of net neutrality [online]. December 2008. URL: `https://freedom-to-tinker.com/blog/felten/three-flavors-net-neutrality/` [cited 29/04/2014].

[84] cPacket Networks Inc. Complete Packet Inspection on a Chip [online]. URL: `http://www.cpacket.com/` [cited 2014/05/05].

[85] Paul Francis, Sugih Jamin, Vern Paxson, Lixia Zhang, Daniel F Gryniewicz, and Yixin Jin. An architecture for a global internet host distance estimation service. In *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 1, pages 210–217. IEEE, 1999.

[86] Lixin Gao. On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on Networking (ToN)*, 9(6):733–745, 2001.

[87] Vikrant S Kaulgud. Ip quality of service: Theory and best practices, 2004.

[88] Stavros G Kolliopoulos and Neal E Young. Approximation algorithms for covering/packing integer programs. *Journal of Computer and System Sciences*, 71(4):495–505, 2005.

[89] Arbor Networks [online]. URL: `http://www.arbornetworks.com/` [cited 2014/05/05].

[90] Ratul Mahajan, Neil Spring, David Wetherall, and Tom Anderson. Inferring link weights using end-to-end measurements. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pages 231–236. ACM, 2002.

[91] Ratul Mahajan, Neil Spring, David Wetherall, and Thomas Anderson. User-level internet path diagnosis. In *ACM SIGOPS Operating Systems Review*, volume 37, pages 106–119. ACM, 2003.

[92] Andrew W Moore and Denis Zuev. Internet traffic classification using bayesian analysis techniques. In *ACM SIGMETRICS Performance Evaluation Review*, volume 33, pages 50–60. ACM, 2005.

[93] Vern Paxson, Jamshid Mahdavi, Andrew Adams, and Matt Mathis. An architecture for large scale internet measurement. *Communications Magazine, IEEE*, 36(8):48–54, 1998.

[94] Larry Peterson, Tom Anderson, David Culler, and Timothy Roscoe. A blueprint for introducing disruptive technology into the internet. *ACM SIGCOMM Computer Communication Review*, 33(1):59–64, 2003.

[95] Jerome H Saltzer, David P Reed, and David D Clark. End-to-end arguments in system design. *ACM Transactions on Computer Systems (TOCS)*, 2(4):277–288, 1984.

[96] Joel Sommers and Paul Barford. An active measurement system for shared environments. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 303–314. ACM, 2007.

[97] Neil Spring, Ratul Mahajan, David Wetherall, and Thomas Anderson. Measuring isp topologies with rocketfuel. *Networking, IEEE/ACM Transactions on*, 12(1):2–16, 2004.

[98] Liz Gannes. At&t continues to adjust tos to limit 3g video [online]. April 2009. URL: `http://newteevee.com/2009/04/29/att-continues-to-adjust-tos-to-limit-3g-video.` [cited 2014/05/05].

[99] Neil Spring, David Wetherall, and Thomas Anderson. Reverse engineering the internet. *ACM SIGCOMM Computer Communication Review*, 34(1):3–8, 2004.

[100] Maurice Kendall, Alan Stuart, J Keith Ord, and A OHagan. Kendalls advanced theory of statistics, volume 1: Distribution theory. *Arnold, sixth edition edition*, 1994.

[101] M Kendall, A Stuart, KJ Ord, and S Arnold. Kendalls advanced theory of statistics: Volume 2a–classical inference and and the linear model (kendalls library of statistics). *A Hodder Arnold Publication,*, 1999.

[102] John W Tukey. Bias and confidence in not-quite large samples. In *Annals of Mathematical Statistics*, volume 29, pages 614–614. Institute Mathematical Statistics, 1958.

[103] Charles V Wright, Fabian Monrose, and Gerald M Masson. On inferring application protocol behaviors in encrypted network traffic. *The Journal of Machine Learning Research*, 7:2745–2769, 2006.

[104] Aditya Akella, Srinivasan Seshan, and Anees Shaikh. An empirical evaluation of wide-area internet bottlenecks. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pages 101–114. ACM, 2003.

[105] Brice Augustin, Timur Friedman, and Renata Teixeira. Measuring load-balanced paths in the internet. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 149–160. ACM, 2007.

[106] Brice Augustin, Xavier Cuvellier, Benjamin Orgogozo, Fabien Viger, Timur Friedman, Matthieu Latapy, Clémence Magnien, and Renata Teixeira. Avoiding traceroute anomalies with paris traceroute. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 153–158. ACM, 2006.

[107] Ioannis C Avramopoulos and Jennifer Rexford. Stealth probing: Efficient data-plane security for ip routing. In *USENIX Annual Technical Conference, General Track*, pages 267–272, 2006.

[108] Cisco. Configuring port to application mapping [online]. URL: `http://www.cisco.com/en/US/products/sw/iosswrel/ps1835/products_configuration_guide_chapter09186a00800ca7c8.html` [cited 2014/05/05].

[109] Marta Carbone and Luigi Rizzo. Dummynet revisited. *ACM SIGCOMM Computer Communication Review*, 40(2):12–20, 2010.

[110] Augustin Soule, Kavé Salamatia, Nina Taft, Richard Emilion, and Konstantina Papagiannaki. Flow classification by histograms: or how to go on safari in the internet. *ACM SIGMETRICS Performance Evaluation Review*, 32(1):49–60, 2004.

A. ICT and network performance

A.1. Translocation

Distributed computation necessarily involves transferring information generated by one computational process to another, located elsewhere. We call this function ‘translocation’, and the set of components that performs it is ‘the network’. Instantaneous and completely loss-less translocation is physically impossible; thus all translocation experiences some ‘impairment’ relative to this ideal.

Translocating information as packets that share network resources permits a tremendous degree of flexibility in how computational processes interact, and allows resources to be used more efficiently compared to dedicated circuits[1]. In packet-based networks, multiplexing is a real-time ‘game of chance’; because the state of the network when a packet is inserted is unknowable, exactly what will happen to each packet becomes uncertain. At each multiplexing point, the ‘game of chance’ is played out between packets of the multiplexed flows. The result of this game is that the onward translocation of each packet to the next element along the path may be delayed, or may not occur at all (the packet may be ‘lost’). This is a source of impairment that is statistical in nature.

The odds of this multiplexing ‘game’ are affected by several factors, of which load is one. In these ‘games’, when one packet is discarded, another is not. Similarly, when one is delayed more, another is delayed less - i.e. this is a zero-sum game in which quality impairment (loss and delay) is conserved.

A.1.1. Mutual interference in network traffic

There is a common misconception that the complexity of networks comes from their interconnectivity - the fact that they can form an arbitrary ‘graph’[2]. However, given the use of routing protocols that select particular paths through this connectivity graph, the particular path of network elements traversed by the packets in a given flow[3] is essentially fixed. The translocation characteristics of the flow are affected only by the other flows that share a common network element on that path, so the complexity of the problem is bounded. The process of sharing resources between flows that follow a common path is called multiplexing. For any particular end-to-end flow, the network is effectively a tree of multiplexers, as illustrated in Figure A.1.

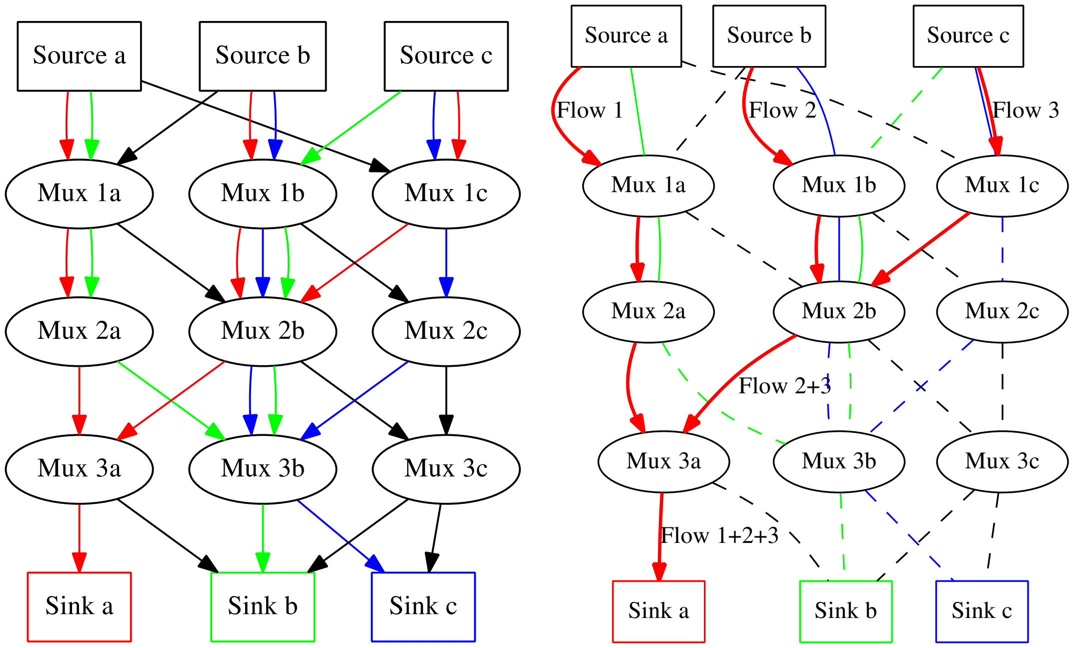
In Figure A.1a, the different coloured lines indicate potential valid routes. Black lines are potential routes that have been ‘pruned’ by the operation of routing algorithms. The lines coloured in red, green and blue represent traffic flowing from sources to sinks, passing through multiplexers (‘Mux’). In practice, any network endpoint functions as both a source and sink, but, for understanding network traffic, it is essential to separate these two roles.

If we now focus on the traffic flowing towards any one sink, for example that flowing to Sink a

[1] This is similar to the familiar benefits of sharing individual computing elements between a number of processes. However, processor sharing is better understood than network resource sharing. This is partly because packets share many and varied network elements, and partly because the number of packets exchanged between processes tends to far exceed the number of processes in a computing node. Thus the sharing of network resources is complex, and predicting its consequences seemingly intractable.

[2] `http://en.wikipedia.org/wiki/Graph_(mathematics)`

[3] Where a flow is the sequence of packets between a particular source and sink.



(a) Complete Network (b) Focussing on one multiplexing tree

Figure A.1.: The network is a tree of multiplexors

(represented by the red lines in Figure A.1b), these flows share resources[4] over portions of the path with other flows (represented by the solid green and blue lines). Note that it is only the common sub-paths that are sources of inter-stream impairment; the rest of the traffic in the network has no influence, as it is running over disjoint paths that do not share resources with the red flows (represented by dotted lines in the figure). Thus, when evaluating the impairment due to competition for resources (the statistical multiplexing) within any network, it is sufficient to consider the tree of multiplexors rooted at each sink.

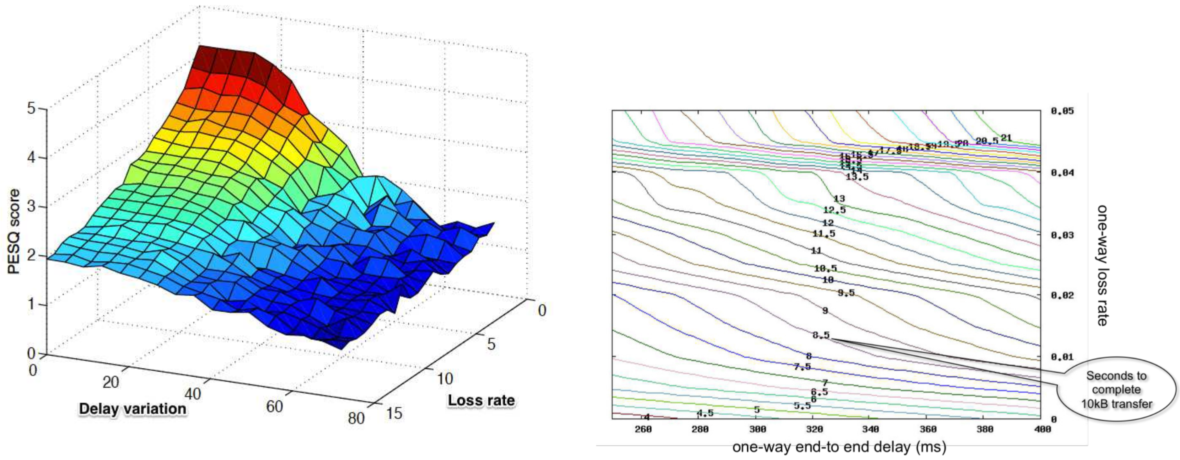
A.2. Network influence on application outcomes: ΔQ

Typical impairments that can affect an analogue telephone call (such as noise, distortion and echo) are familiar; for the telephone call to be fit-for-purpose, all of these must be sufficiently small. Analogously, we introduce a new term, called 'quality attenuation' and written 'ΔQ', which is a measure of the impairment of the translocation of a stream of packets when crossing a network. This impairment must be sufficiently bounded for an application to deliver fit-for-purpose outcomes[5]. For example, Figure A.2a (reproduced from [24]) shows the impact of delay variation and loss rate (both of which are aspects of ΔQ) on the audio quality of a G.711 VoIP call. Figure A.2b shows the impact of delay and loss rate on the 95th percentile time to complete a 10kB HTTP transfer, such as a small web page.

ΔQ captures the effects of the network's structure, together with the the impairment due to statistical multiplexing (as discussed in §A.2.2 below). Thus ΔQ is an inherently statistical measure that can be thought of as the probability distribution of what might happen to a packet transmitted at a particular moment from source A to destination B, or the statistical properties of a stream of such packets.

[4]For example, the finite capacity to transmit data from each Mux to the next, and the finite capacity to buffer data for transmission at each egress point from a Mux.

[5]Just as a telephone call might fail for reasons that are beyond the control of the telephone company (such as excessive background noise or a broken handset), applications may fail to deliver fit-for-purpose outcomes for reasons that are beyond the control of the network (e.g. lack of local memory or insufficient computing capacity). Such considerations are out of scope here.



(a) Impact of ΔQ on VoIP performance (b) Impact of ΔQ on HTTP performance

Figure A.2.: Impact of ΔQ on application performance

A.2.1. Application performance depends only on ΔQ

Applications depend on information to complete computations. To provide appropriately timely outcomes, delivery of this information needs to be done in a timely and correctly sequenced manner. If information takes too long to arrive (and/or too much of it is missing[6]) then the computations cannot proceed, and the application fails to deliver the requested service or to deliver an acceptable performance of that service.

Different components of a distributed application (e.g. a client and a server) exchange information as streams of packets. If those packets were all delivered instantaneously (i.e. if there were no impairment in the translocation), and the computational components performed correctly, the application would work. However, as discussed above, sending packets over distances using shared resources *inevitably* means there will be some delay and occasionally packets may be lost - this is ΔQ. Whether the application still delivers fit-for-purpose outcomes depends entirely on the extent of the quality impairment (the magnitude of ΔQ), and the application's sensitivity to it. The layering of network protocols isolates the application from any other aspect of the packet transport. This is such an important point it is worth repeating: the great achievement of network and protocol design has been to completely hide all the complexities of transmission over different media, routing decisions, fragmentation and so forth, and leave the application with only one thing to worry about with respect to the network: ΔQ.

'Bandwidth required' is a characteristic of the application load. If many of the packets the application offers are discarded, users would typically say that the 'available bandwidth' is too low; however, from the perspective of the application, the immediate problem is that ΔQ is too large. Indeed such packet loss might well occur for reasons other than the capacity limitation of the transmission links. If it is delay (rather than loss) that is too large, this may not be because of constraints of capacity, but rather of schedulability[7] - i.e. issues of instantaneous, rather than average, loading[8].

A.2.2. How ΔQ accrues across the network

Network structure (including the types, lengths and speeds of network links) affects ΔQ. To illustrate this, consider Figure A.3, which focuses on the path from $\text{Source}_b \rightarrow \text{Sink}_a$ from

[6]It may be thought that data 'corruption' could also occur, but the underlying data transport mechanisms have checksums that cause any such corruption to be treated as loss. Even though a data packet may be lost, the protocols recover (typically through retransmission, where needed), transforming such loss into delay.

[7]Where schedulability is the ability to sequence the instantaneous demand to meet requirements.

[8]Loss can also be caused by schedulability constraints, especially where applications produce large bursts of packets.

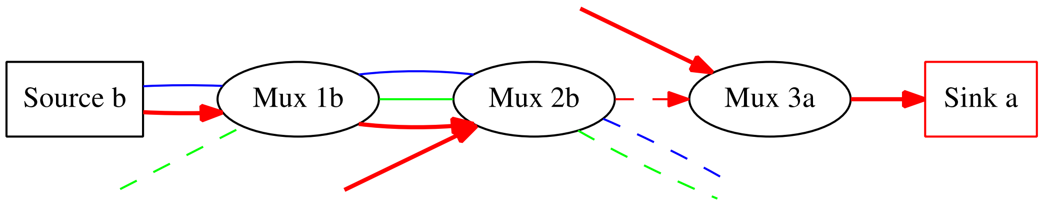


Figure A.3.: An end-to-end path through a network (from A.1b)

Figure A.1b. The overall end-to-end ΔQ, is the 'sum' of the ΔQ associated with each path[9], i.e.:

$$\Delta Q^{\text{Source}_b \to \text{Sink}_a} = \Delta Q^{\text{Source}_b \to \text{Mux}_{1b}} \oplus \Delta Q^{\text{Mux}_{1b} \to \text{Mux}_{2b}} \oplus \cdots \oplus \Delta Q^{\text{Mux}_{3a} \to \text{Sink}_a}$$

The overall ΔQ of flows following this path is dependent on several aspects, which can be split into two broad categories: *structural* and *variable*. Structural ΔQ captures properties such as the geographical distribution of the network elements (denoted $\Delta Q_{|\mathbf{G}}$) and the extent to which bigger packets take longer to be transmitted[10] (denoted $\Delta Q_{|\mathbf{S}}$).

Figure A.4 illustrates the process of extracting ΔQ and its components from raw point-to-point delay data. If one measures delays for packets with a range of sizes and then plots these delays by packet size, a structure emerges. Structural components of ΔQ can be extracted, the remainder is the variable component.

Like the overall ΔQ, the individual elements can also be combined:

$$\begin{aligned}
\Delta Q_{|\mathbf{G}}^{\text{Source}_a \to \text{Sink}_b} &= \Delta Q_{|\mathbf{G}}^{\text{Source}_a \to \text{Mux}_{1b}} \oplus \Delta Q_{|\mathbf{G}}^{\text{Mux}_{1b} \to \text{Mux}_{2b}} \oplus \cdots \oplus \Delta Q_{|\mathbf{G}}^{\text{Mux}_{3a} \to \text{Sink}_a} \\
\Delta Q_{|\mathbf{S}}^{\text{Source}_a \to \text{Sink}_b} &= \Delta Q_{|\mathbf{S}}^{\text{Source}_a \to \text{Mux}_{1b}} \oplus \Delta Q_{|\mathbf{S}}^{\text{Mux}_{1b} \to \text{Mux}_{2b}} \oplus \cdots \oplus \Delta Q_{|\mathbf{S}}^{\text{Mux}_{3a} \to \text{Sink}_a} \\
\Delta Q_{|\mathbf{G},\mathbf{S}}^{\text{Source}_a \to \text{Sink}_b} &= \Delta Q_{|\mathbf{G},\mathbf{S}}^{\text{Source}_a \to \text{Mux}_{1b}} \oplus \Delta Q_{|\mathbf{G},\mathbf{S}}^{\text{Mux}_{1b} \to \text{Mux}_{2b}} \oplus \cdots \oplus \Delta Q_{|\mathbf{G},\mathbf{S}}^{\text{Mux}_{3a} \to \text{Sink}_a}
\end{aligned}$$

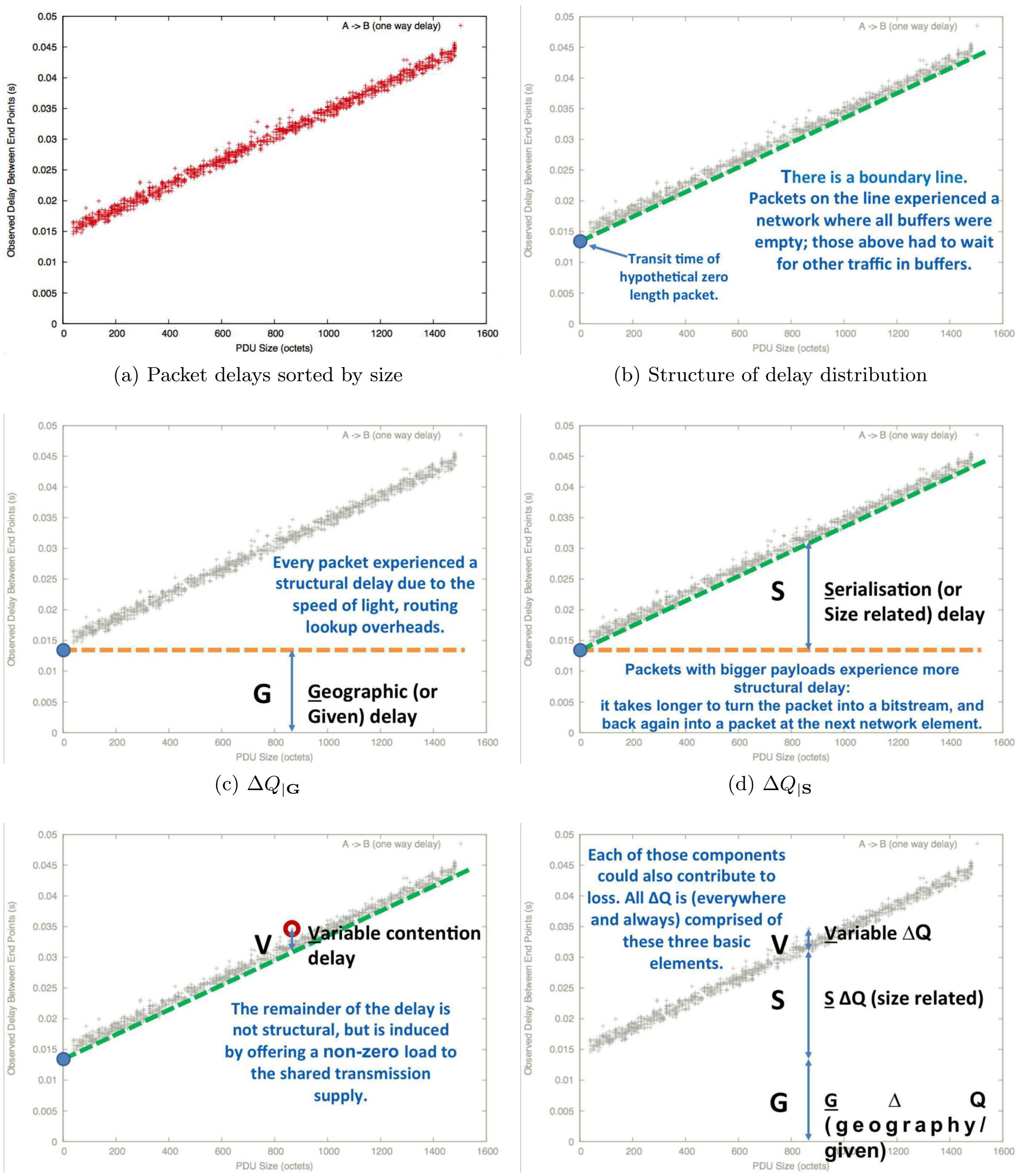
In addition to the $\Delta Q_{|\mathbf{G},\mathbf{S}}$ (structural ΔQ) along a path, there is a variable component, denoted $\Delta Q_{|\text{V}}$. This component captures the effects of multiplexing resources (such as link capacity in wired networks, or local spectrum capacity in wireless access). In Figure A.3, multiplexing will occur at each of the nodes (Source_b, Mux_{1b}, Mux_{2b} and $\text{Mux}_{3\text{a}}$). This is where $\Delta Q_{|\mathbf{V}}$ accrues, as a function of the load offered there, i.e. the set of packets requiring to be forwarded at a particular moment. Note that $\Delta Q_{|\mathbf{V}}$ is related to the *total* offered load[11] and is a direct and unavoidable consequence of packet-based statistical multiplexing. In exchange for the efficiency gained by not dedicating resources to individual data flows (as circuit-based networking does), we must accept the possibility that more packets will arrive than can immediately be forwarded, so some must wait (or be lost). ΔQ is conserved (as discussed above). So, whatever mechanism is used to affect the $\Delta Q_{|\mathbf{V}}$ of any flow at any point (say the blue flow as it egresses $\text{Mux}_{1\text{b}}$ in Figure A.3), the *best* that can be achieved is that the overall $\Delta Q_{|\mathbf{V}}$ (without regard to any particular flow) is not increased. The constraint that the sum of the $\Delta Q_{|\mathbf{V}}$ for individual streams cannot be less than that for the aggregate flow is expressed in the equation:

$$\sum_{c \in \{\text{red,green,blue}\}} \Delta Q_{|\mathbf{V}}^{\text{Mux}_{1\text{b}} \xrightarrow{c} \text{Mux}_{2\text{b}}} \geqq \Delta Q_{|\mathbf{V}}^{\text{Mux}_{1\text{b}} \to \text{Mux}_{2\text{b}}}$$

[9] We treat 'ΔQ' as a plural noun.

[10] This is more than the 'speed' of the network link, it incorporates the influence of transmission technology on the time taken to service packets of varying length.

[11] Where the total offered load is the combined load of all flows passing through the shared node.



(a) Packet delays sorted by size

(b) Structure of delay distribution

(c) $\Delta Q_{|\mathbf{G}}$

(d) $\Delta Q_{|\mathbf{S}}$

(e) $\Delta Q_{|\mathbf{V}}$

(f) Decomposition of ΔQ

Figure A.4.: ΔQ and its components

The way in which the $\Delta Q_{|\mathbf{V}}$ is distributed between different flows at a particular multiplexing point is the result of the queuing and scheduling mechanisms operating there. However, any such mechanisms are inherently subject to the above conservation constraint (this is a generalisation of the work in [25]). Thus, the overall $\Delta Q_{|\mathbf{V}}$ that the red traffic experiences is:

$$\Delta Q_{|\mathbf{V}}^{\mathrm{Source}_a \xrightarrow{\mathrm{red}} \mathrm{Sink}_b} = \Delta Q_{|\mathbf{V}}^{\mathrm{Source}_a \xrightarrow{\mathrm{red}} \mathrm{Mux}_{1b}} \oplus \Delta Q_{|\mathbf{V}}^{\mathrm{Mux_{1b}} \xrightarrow{\mathrm{red}} \mathrm{Mux_{2b}}} \oplus \cdots \oplus \Delta Q_{|\mathbf{V}}^{\mathrm{Mux_{3a}} \xrightarrow{\mathrm{red}} \mathrm{Sink_a}}$$

For a given end-user communicating with a given endpoint, the main network factor that influences the variation in their experience is the $\Delta Q_{|\mathbf{V}}$ (in both directions) of the translocation along the path connecting them.

Each user experience of a particular application is affected by the presence of other resource-sharing traffic. This traffic acts as ‘pollution’ that, from the user’s point of view, potentially degrades their application’s performance. TM is one approach to addressing this problem, but it is again subject to the conservation law above - any ‘pollution’ can be ‘traded’ but never eliminated.

Trading occurs whenever resources are shared, whether this is explicitly acknowledged or not. In networks, such trading occurs at every network element and at every network port (i.e. every multiplexing point). If no action is taken, these trades are determined implicitly by the various mechanisms operating in each element, and are of an unstructured and disordered nature. They do not intrinsically provide fairness nor do they explicitly support the policy or aims of the network operators or designers. Managing this may appear to be an overwhelmingly complex problem[12], but mathematically-based approaches (such as the one outlined here) can contain that complexity and clarify the constraints on what is achievable. These can be used to inform a higher-level discussion of desirable outcomes, and can also enable the identification of any related hazards to the delivery of fit-for-purpose outcomes.

A.3. Summary

In this appendix, we have introduced the notion of translocation - the end-to-end transport of information units between computational processes. We have outlined the notion of ΔQ, a statistical measure that captures the performance of such translocation, in a way that is independent of the underlying network technology[13]. As a measure, ΔQ:

- accrues along the end-to-end path of each data flow;
- expresses the impact of the structural aspects of the network on translocation;
- can be directly related to the delivered QoE of applications;
- is conserved, in that having been ‘created’ it can not be ‘destroyed’ - although some aspects can be differentially shared;
- depends on load, thus incorporating the way in which ‘bandwidth’ is typically used to express requirements;
- captures the variability of translocation due to the statistical sharing of resources at multiplexing points.

We have shown how the apparent complexity of analysing interactions between multiple packet flows can mitigated by focusing on the tree of multiplexors rooted at a particular sink. By combining this with the composability of ΔQ, the analysis of network performance interactions becomes tractable.

[12] From an ontological point of view, these systems are completely predictable (that is they would produce the same results given precisely the same starting conditions and inputs over time). The overall outcome can be highly dependent on seemingly minor aspects of the inputs; thus it is in their epistemology that the complexity lies.

[13] This holds whether the underlying network is wired, wireless, copper, fibre, 2G/3G/4G, satellite, etc..

B. Traffic Management methods and their impact on ΔQ

As discussed in § 1.2.1 on page 12, multiplexing in ICT systems is the statistical sharing of common resources, such as point-to-point transmission capacity. Buffering is needed to allow for arrivals to occur when the resource is busy. This creates contention for two things: the ability to be admitted into the buffer (ingress), and the ability to leave the buffer (egress). Whether the first is achieved determines loss, and the time taken to achieve the second determines delay; together these represent the mechanisms that create $\Delta Q_{|\mathrm{V}}$, the variable component of quality attenuation[1]. At every multiplexing point in a network a 'game' is being played out between different streams of packets. The term 'Traffic Management' is usually associated with the configurations of multiplexing points, as these determine the 'odds' of this game.

B.1. Packet-based multiplexing and $\Delta Q_{|\mathrm{V}}$

In packet-based networks, each packet has a header that contains the information necessary to direct it towards its destination on a hop-by-hop basis (this is the function of routing). Each point along this hop-by-hop path acts as a multiplexor, processing complete packets[2]. As packets can arrive when the ongoing transmission path is busy, buffering is needed[3].

While the competition for network resources is typically viewed in terms of 'bandwidth', it is more useful to regard multiplexing as two competitions between packets; one to get into the buffer (ingress); and another to get out of it (egress). Queueing and scheduling techniques differ solely in their ingress and egress actions with respect to this buffer[4]. Viewing the operation in this way, it is clear that:

1. The failure to be admitted to the buffer, as part of the ingress behaviour, is a source of packet loss[5];
2. The instantaneous occupancy of the buffer represents *the total accrued delay*; this total delay is independent of the order in which the packets are eventually serviced[6];
3. The order in which packets are chosen, the egress behaviour, determines the delay that the individual packets experience.

In point 2 above, there is an assumption that the egress behaviour is work-conserving - i.e. packets will be sent whenever the buffer is non-empty. Most queueing and scheduling

[1]While there are other ways in which the overall ΔQ can accrue, for example due to electrical noise in transmission and associated recovery, these are not the dominant factors for most broadband connections.

[2]I.e. when a packet is sent, a complete packet is sent; when a packet is discarded, a complete packet is discarded. While packet fragmentation is possible, for the purposes of this report it is an advanced topic.

[3]For the sake of completeness, we note that this is where TDM-based transmission fundamentally differs from PBSM. TDM's design eliminates the need for buffering at intermediate routing points. Between entering and leaving a pure TDM network, packets will experience 'perfect' $\Delta Q_{|\mathrm{V}}$, zero delay and no loss from multiplexing.

[4]We note that equipment may allocate separate buffer capacity to different purposes. This is an operational refinement that does not affect the total buffering being used. It is the total buffer use that we will consider here.

[5]While there are techniques in which existing packets can be 'pushed-out' by other arriving packets, they do not represent a fundamental change to the nature of the problem and so we will not consider them in this report.

[6]More accurately, the instantaneous occupancy represents an absolute lower bound on the overall delay.

techniques work this way, with the exception of rate limiting (§B.4.4), whose specific aim is to control the egress rate from the buffer[7].

When examining the effects of a queueing and scheduling mechanism, there are two complementary viewpoints. The first is a component-centric view, considering the total $\Delta Q_{|V}$ being created by the component's operation; the second is a translocation-centric view, which focuses on the $\Delta Q_{|V}$ that the packets for an individual application (or end-user) experience. Application outcomes are not generally determined by the fate of any one particular packet, so the $\Delta Q_{|V}$ of interest is the *probability distribution* of the individual packet experiences. This includes the two extremes of $\varnothing$ (perfect transmission without delay) and $\perp$ (loss). The fine-grain behaviour of network protocols is sensitive to the pattern of the end-to-end $\Delta Q_{|V}$. Taking TCP/IP as a case in point, timeouts are calculated on recent round-trip times[8], and the pattern of loss drives congestion avoidance.

B.1.1. FIFO

The most common queueing and scheduling approach, and hence the most common 'traffic management' technique, is a FIFO (first-in first-out) queue[9].

B.1.1.1. Ingress behaviour

On arrival, a packet is admitted to the buffer if there a free slot. Packets arriving (from all sources) whilst the buffer is fully occupied are discarded (this is referred to as 'tail-drop'). It should be noted that the destination system receives no direct indication of this loss, but must infer it from the non-arrival of an expected packet[10].

B.1.1.2. Egress behaviour

Packets are chosen from the buffer in the order that they were admitted[11]. The delay that each packet will experience is made of two components. The first is the time taken for the transmission link to become idle, i.e. to finish processing the packet currently being sent, if any. The second is the time for the packet in question to be chosen for transmission (the queueing time).

B.1.1.3. Discussion

When a packet arrives at a FIFO where both the buffer is empty and the transmission resource is idle, it will be forwarded immediately[12] without being discarded[13]. In this case there is no contention for the common resource, and the experienced $\Delta Q_{|V}$ is $\varnothing$ - 'perfection', no delay or loss.

When a packet arrives at a FIFO whose buffer is full (which implies the transmission resource is non-idle) it will be discarded and never arrive at its intended destination[14]. This corresponds

[7] The use of buffers to de-jitter streams, such as in VoIP, has a similar non-work-conserving property.

[8] These RTTs are, in turn, dependent on the bi-directional $\Delta Q_{|V}^{A\leftrightarrow Z}$.

[9] This is also known as FCFS (first-come first-served).

[10] This is the role of sequence numbers and timeouts in protocols.

[11] This is typically done by choosing the packet at the head of a queue. The queue in question is formed by placing each admitted packet at the back of the queue as they arrive during ingress processing.

[12] We are assuming that the FIFO is work-conserving, unless stated otherwise.

[13] From the point of view of an external observer, the leading edge of the packet will commence transmission at the time the trailing edge arrives ($\Delta Q_{|S}$ would come into play if, for example, the time was measured between arrival and departure of leading edges). Any difference between the end of the packet arriving and the packet being transmitted (such as time required to look up routing tables) would be a contributor to the $\Delta Q_{|G}$.

[14] This may seem a spurious distinction, however the difference is important. The non-arrival at the receiver within a time period of interest is an externally observable phenomenon, whereas the packet discard is an

to a $\Delta Q_{|V}$ of $\perp$ (mathematically called 'bottom'). When a packet arrives at a FIFO whose buffer is not full but whose transmission link is not idle, it will experience a delay determined by the current state of the buffer. This delay is dependent on both the length of the queue on arrival and the residual service time for the packet being transmitted.

As discards occur when the buffer is full, it is interesting to ask the following questions:

1. Given that a buffer is full, how long can it remain full;
2. How many packets can arrive while the buffer is full?

The buffer remains full until the packet in transmission has been completely sent. The time taken to send this packet is dependent on the size of the packet (bounded by the technology and its maximum packet size) and the transmission rate of the egress link. For example, a 2Mbps ADSL connection[15] takes 6.1ms to transmit a 1500 byte IP packet. In the same amount of time a 1Gbps Ethernet connection can transmit 495 such packets[16].

The number of packets that can arrive in any period of time is dependent on the aggregate ingress rate to the device. When the multiplexing point is at the egress of a switch, the maximum ingress rate would be the sum of the individual ingress link rates. Thus, if the individual rates are the same[17], the maximum number of packets that can arrive while the buffer is full is given by the number of ports on the switch.

Under the assumption of 'random' traffic arrivals, at low loading there is a high probability that a packet arriving will experience a $\Delta Q_{|V}$ of $\varnothing$, and a very low probability of experiencing $\perp$. Hence traversing this particular hop is highly likely to increase only the overall end-to-end ΔQ by its contribution to $\Delta Q_{|G,S}$. For this to hold, 'randomness', i.e. the independence of packet arrivals, is essential. Even in networks with very low average loads[18] correlated loading patterns can generate significant $\Delta Q_{|V}$. These correlation issues are discussed in §B.2.

When the ingress rate approaches or equals the egress rate and the load is uncorrelated, FIFO has the interesting property that all possible states[19] of the buffer become equally likely[20]. For example, at 100% offered load, a FIFO with 100 buffers[21] would deliver a link utilisation of 99%, a loss rate of 1%, and a uniform distribution of all the possible delays between 0 and 99 packet service times.

B.1.1.4. Fairness with respect to ΔQ

In data networks, and ICT in general, resource usage is often 'rivalrous'[22]. The instantaneous state of a buffer can be seen as recording the recent history of that rivalry[23].

FIFO is often viewed as a 'pure' mechanism that treats traffic 'fairly'. This sense of fairness may have arisen from a particular mathematical formulation of FIFO queueing[24]. In practice, the distribution of $\Delta Q_{|V}$ between competing translocation streams can be substantially biased by their individual arrival patterns[25]. The authors have had experience of large network

internal event and thus not necessarily observable. One can be measured by an external third party, the other cannot.

[15] That is, one that would sync at around 2,208 kbps and transmit up to 5,208 ATM cells per second.

[16] A 1Gbps ethernet connection can carry 81,274 maximum ethernet frames per second - `http://goo.gl/xPY5g2`

[17] Here, the "individual rates" include those of the egress and all ingresses.

[18] This could be measured by, for example, link utilisation over five minute periods.

[19] These states would include $\varnothing$, $\perp$, and all values of delay in between.

[20] Thus, the system is at maximum entropy.

[21] That is, one with 1 buffer for the packet in service and 99 queueing slots.

[22] This means that use by one party prevents use by another. `http://en.wikipedia.org/wiki/Rivalry_(economics)`

[23] Noting that the 'memory' of that history is wiped clean every time the buffer becomes empty.

[24] There is a circumstance in which the arriving streams will experience the 'same' $\Delta Q_{|V}$, i.e. they will experience the same distribution of delay and the same rate of loss. This occurs when the service pattern is Markovian and all traffic sources are Poisson processes - i.e. the overall aggregate arrivals are Markovian.

[25] This is particularly true for the distribution of loss, a phenomenon that has been exploited in the design of low rate denial-of-service attacks [26].

providers encountering issues stemming from this when increasing capacity in core parts of their systems[26].

The $\Delta Q_{|\mathrm{V}}$ of a single network element is not the only contributory factor to the overall end-to-end ΔQ, even where this is the only network element at which there is contention. The other aspects of ΔQ, the difference in $\Delta Q_{|\mathrm{G}}$ and $\Delta Q_{|\mathrm{S}}$ between two end-points, can substantially influence the delivered outcome of the same application at different locations[27]. Thus the key question regarding fairness is: *with respect to what metric?* Fair distribution of $\Delta Q_{|\mathrm{V}}$ at a single contention point does not assure overall fairness in outcome, and may even hinder such a goal.

B.2. Load correlation, elastic protocols and Predictable Region of Operation (PRO)

In this report we view traffic management as the choice and configuration of queueing and scheduling within network elements, combined with their order and location[28].

As $\Delta Q_{|\mathrm{V}}$ is conserved, traffic management can differentially share it (see §B.3.1) and/or change in which network element it occurs (this is discussed in §B.3.2). Even in the case of the finite FIFO discussed above, there is a choice of how many buffers to configure; this biases the trade between delay and loss (as mentioned in §B.3.1).

Multiplexed resources are ones that match demand and supply over some timescale. In this case, the demand is the arrival (ingress) pattern and the supply is the departure (egress) pattern from the buffer for onward transmission[29]. The instantaneous occupancy of the buffer is influenced by both the loading factor (the ratio of arrival rate to departure rate) and any correlation in the arrival pattern[30]. For a given loading factor, the correlation between arrivals will have a substantial effect. In the Internet, a significant cause of such correlation is the operation of protocols that are 'elastic' (i.e. they endeavour to adapt their offered load to the apparent capacity constraints on the end-to-end path). TCP/IP is the most widespread example. Different choices of protocol behaviour at end-points have an influence on the delivered quality attenuation[31] [27].

Correlated load causes $\Delta Q_{|\mathrm{V}}$ to vary, as shown occurring between two ISPs within the UK in Figure B.1. The issue for end-to-end service delivery is that excessive $\Delta Q_{|\mathrm{V}}$ can cause a network service to leave its *Predictable Region of Operation* (PRO). This arms the hazard that it will not perform 'correctly'. The consequences of the hazard maturing are service dependent. For a video-on-demand service, it might mean a video artefact on the screen or a 'buffering pause'. For an integral system service (such as routing updates or keep-alives on a L2TP tunnel), the consequence might be that all the connections between an ISP and its

[26] When capacity was increased, longer and more dense back-to-back packet sequences formed. These sequences then generated burst loss in a downstream FIFO, with the overall effect of reducing the delivered QoE for some applications.

[27] For example, the rate at which TCP/IP increases its window is a function of the overall round trip time ($\Delta Q^{\mathrm{A}\leftrightarrow\mathrm{Z}}_{|\mathrm{G,S,V}}$). This TCP/IP performance property has a great impact on the 'time-to-first-frame', which is an important QoE metric in video delivery.

[28] We are focusing on those factors that affect a data translocation service between defined boundaries.

[29] We are going to assume that the onward transmission is not, itself, a dynamically multiplexed resource (as would be the case if the transmission was being carried as an MPLS circuit or some other statistically multiplexed transmission such as LTE). This does not affect the general argument, and that situation still remains amenable to analysis, but full explanation is beyond the scope of this report.

[30] A general discussion of the causes and effects of correlation is beyond the scope of this report. Interested readers can find more material in works on Large Deviations Theory (`http://en.wikipedia.org/wiki/Large_deviations_theory`) and texts on teletraffic engineering (`http://en.wikipedia.org/wiki/Teletraffic_engineering`). Correlations do not occur at the network translocation level only; correlation of load also occurs in the demand for service.

[31] Many approaches have been taken within the framework of TCP, where the key concern is the "avoidance of congestion", not the delivery of consistent performance. See `http://en.wikipedia.org/wiki/TCP_congestion-avoidance_algorithm`.

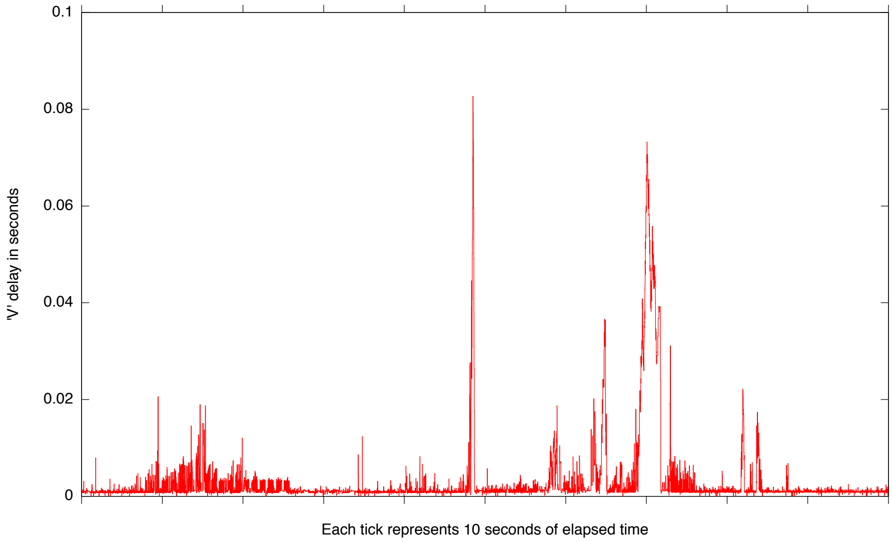


Figure B.1.: Example of one way delay between two points connected to UK internet

The figure shows a measure of the combined $\Delta Q_{|\mathrm{V}}$ over time between a network element within $\mathrm{ISP_a}$'s core network and a network element within $\mathrm{ISP_b}$'s core network, across a UK internet exchange. The data rate applied was less than 3Mbps. There were no reported errors or performance issues along the path over the measurement period.

customers are dropped. This potential for operational 'catastrophe' is a key driver for traffic management[32].

This risk of catastrophe is a consequence of the coupling of system stability with operational activity. It results from combining control plane and data plane traffic, a practice fundamental to the internet design philosophy. This $\Delta Q_{|\mathrm{V}}$-related issue, and the associated performance hazards, is inherent in the current use of PBSM. The fundamental distinction is between data bearers for which $\Delta Q_{|\mathrm{V}}$ is $\varnothing$ (e.g. PDH, SDH[33]) and those for which it is not (e.g. MPLS, Carrier Ethernet[34]).

Where (and hence within which management domains) quality attenuation accrues has changed over time due to the commercial evolution of large-scale broadband. This means that inter-user effects have become possible (as described in §A.1.1) and the PBSM supply chain can now influence how any resulting $\Delta Q_{|\mathrm{V}}$ is distributed. As traditional telcos have taken on the delivery of broadband using PBSM, some control over the distribution of $\Delta Q_{|\mathrm{V}}$ has left the telcos' customers' (i.e. ISPs) hands[35]. This has two consequences:

1. The customer sees a $\Delta Q_{|\mathrm{V}}$ that is no longer in direct relationship with their own pattern of load. In particular, a level of control over the PRO of their applications of interest has been removed;
2. The PBSM network operator has taken on the inter-end-user $\Delta Q_{|\mathrm{V}}$ hazard, typically with little or no associated contractual risk. In particular, the hazard of $\Delta Q_{|\mathrm{V}}$ causing the end-user's application to leave its PRO is outside their contractual scope[36].

[32]This is likely to become more important due to SDN and other developments, as discussed in a recent Ofcom report[23]. In section 4.6.3 (p49) Ofcom touches on issues of emergent fairness in traffic management.

[33]In fact, this could be any resource where there is strong isolation between users - namely each user's traffic patterns and usage don't affect the $\Delta Q_{|\mathrm{V}}$ for other users of the same resource. Examples of this are: different light wavelengths within the same fibre, unshared point-to-point wireless, and the use of SDH/TDM from end-to-end.

[34]This is true even when such resources are allocated to peak.

[35]This situation is in contrast to the days of dial-up modems, when all of the contention for resources occurred in the end-users' premises or within the ISP's own network.

[36]SLAs are typically about long term (e.g. monthly) averages and ΔQ is about instantaneous properties. A

The current nature of the management and administrative domains in the UK, and their traffic management influences, is discussed in Appendix C.

B.3. Trading space available for traffic management

It is self-evident that if a packet is delayed whilst traversing a network it cannot be 'un-delayed'. Similarly, if a packet is discarded (lost) it cannot be 'un-lost'[37]. ΔQ is 'conserved'[38], i.e. it can only increase. It cannot be 'undone'; at best it can be differentially shared. The individual components ($\Delta Q_{|\mathrm{V}}$, $\Delta Q_{|\mathrm{G}}$ and $\Delta Q_{|\mathrm{S}}$) are also conserved in the same way. When considering TM, we focus on the $\Delta Q_{|\mathrm{V}}$ component.

At any point in time, the contents of a network element's buffer would take a particular time to empty. This would be independent of the order in which the packets were serviced (i.e. the delay is conserved). The fact that the overall delay in a queuing system is independent of the choice of scheduling algorithm has been well known since the mid-1960's [28]. It is of interest to note that this analysis assumed an infinite buffer - in such a case delays would then be unbounded. With a finite buffer, the overall delay is always bounded; however this bounding of delay is at the cost of sometimes discarding packets[39].

B.3.1. Overall delay and loss trading

The fact that quality attenuation is conserved has profound consequences for PBSM systems, influencing not only their design and deployment but also their underlying cost structures [29]. Traffic management can be used to 'trade' within the overall conservation constraints. This trading process can be viewed from two different perspectives: one focusing on the accrual of $\Delta Q_{|\mathrm{V}}$ at a component; one focusing on the effects on the overall translocation for a specific flow.

B.3.1.1. Component-centric view

Given that a finite buffer must discard some packets whenever its instantaneous load is too high, increasing its size will decrease the rate of loss (at the cost of increasing the maximum total delay). Similarly, if the experienced delay is deemed too high (for a given arrival pattern), reducing the number of buffers[40] will reduce the overall delay, with increased loss[41]. In data networks such trades may occur many times along an end-to-end path, at every multiplexing point (in particular, every switch and router), so the configuration of these network elements influences the resultant $\Delta Q_{|\mathrm{V}}^{A\leftrightarrow Z}$.

A way of reducing the overall $\Delta Q_{|\mathrm{V}}$ at a network element is to lower its loading factor (the ratio of the arrival rate to the departure rate). This can be done either by reducing the offered load or by increasing the service capacity. The latter is the common industry practice of "use more bandwidth" or "apply generous dimensioning". This can be cost-effective; however its effectiveness is predicated on certain assumptions:

1. That arrivals are independent and 'random'. This assumption is fragile for the reasons discussed in §B.2. The operation of elastic protocols means that increasing capacity does not generate as much performance headroom as might be expected.
2. That the increased capacity improves the statistical multiplexing gain, i.e. increases the number of active load sources required to saturate the constrained resource. The

Telco meeting an SLA does not mean that an application of interest will remain within its PRO.

[37] The information in a packet can be resent, but this generates a new packet.

[38] ΔQ is thus similar to the concept of entropy in thermodynamics.

[39] As loss is also quality attenuation, the overall ΔQ is still conserved.

[40] Alternatively, packets already queued may be discarded.

[41] Such a trading space is a common property of all statistically shared resources.

market-driven trend to increase capacity in the last mile (narrowband → broadband → superfast) has reduced the number of active end-points required to saturate network resources along the path. The corollary is that the ability of one user to affect the QoE of neighbouring[42] users has increased.

These factors have lead to a reduction in the effectiveness of capacity increases to maintain customer experience. In the absence of any economic incentive to temper the volume and pattern of demand over the short timescales on which QoE is most affected, an increasing focus on traffic management has emerged as an alternative solution.

B.3.1.2. Translocation-centric view

The telecommunications supply chain tends to take a component-centric view, e.g. upgrade planning tends to be done on the basis of how busy or 'hot' individual network elements are[43]. However, the overall $\Delta Q_{|\mathrm{V}}$ at a multiplexing point is determined by a combination of the total buffering, the ingress pattern and the egress rate. This is a more complex relationship than can be captured by, for instance, a 5-minute average of utilisation; in general, there is no lower bound of such utilisation that will guarantee a bound on $\Delta Q_{|\mathrm{V}}$[44].

It is possible to 'trade' $\Delta Q_{|\mathrm{V}}$, that is, the $\Delta Q_{|\mathrm{V}}$ of a given translocation through a network element can be made different from the rest. This can be done:

- by modifying the ingress behaviour (to the buffer). That is giving the particular flow (or class of flows) preferential access to some or all of the buffers, which has the effect of reducing the loss rate experienced;
- by modifying the egress behaviour (from the buffer). That is preferentially servicing packets from the chosen flow (or class of flows), which has the effect of reducing the the delay experienced.

These ingress and egress treatments are driven by some notion of precedence, which itself can be based on:

- the association (information derived from the source or destination address or similar, e.g. protocol type);
- recent usage patterns (e.g. offered load rate);
- some notion of 'share' (which could be some weighting, like servicing several packets for a particular flow for every one for another flow).

It should be remembered that whether this differential treatment can deliver an upper bound on the quality attenuation ($\lceil \Delta Q_{|\mathrm{V}} \rceil$) of a given flow will depend on the pattern of its offered load as well as properties of the total load[45]. Also, such differential treatment has consequences for the other flows passing through this multiplexing point, since the overall $\Delta Q_{|\mathrm{V}}$ is conserved.

B.3.2. Location-based trading

As has been discussed above, the $\Delta Q_{|\mathrm{V}}$ that occurs at a component depends on the traffic pattern, so changing that pattern can reduce the overall $\Delta Q_{|\mathrm{V}}$ that accrues at this point in the network. This occurs during traffic shaping and rate policing, which induce additional $\Delta Q_{|\mathrm{V}}$

[42] This is in the sense of §A.1.1.

[43] This 'temperature' is typically some measure of average utilisation, such as a moving average of the 5 or 15-minute load. Note that this is a pure heuristic, since averaging averages does not have any coherent mathematical interpretation.

[44] The inference does work the opposite way around: when $\Delta Q_{|\mathrm{V}}$ is frequently exceeding some threshold, often this implies high utilisation.

[45] For example $\lceil \Delta Q_{|\mathrm{V}} \rceil$ can be shown to depend only on the number of streams when applying the policy of "allocation to peak" (where the individual offered loads are strongly policed - either by the physical characteristics of the interface/circuit or otherwise - and their peak, including any encapsulation overheads, cannot exceed the service capacity of the egress). In all other cases delivering a $\lceil \Delta Q_{|\mathrm{V}} \rceil$ depends on schedulability constraints being able to be met.

at one point to change the arrival pattern at a subsequent point. Thus rate limiting/traffic shaping 'moves' where the $\Delta Q_{|\mathrm{V}}$ (for that particular translocation stream) accrues.

From the point of view of application outcomes, such $\Delta Q_{|\mathrm{V}}$ trading does not necessarily have a detrimental effect. The contour lines of 'equal outcome' in Figure A.2 in §A.2.1 show that there is scope for trading ΔQ, while maintaining application outcome and hence user experience.

Interfaces implicitly act as traffic shapers. Thus, the change from narrowband to broadband to superfast can be seen as the slackening of rate limiters. This effectively moves $\Delta Q_{|\mathrm{V}}$ between locations, in particular between management domains.

B.4. Other queueing and scheduling approaches

As we have seen, there are a set of inherent performance properties that naturally arise out of the operation of PBSM networks. The simplest implementation of a broadband network (comprising first-in first-out, tail-drop queues served by fixed-rate dedicated circuits) still engages in 'traffic management', in that it shares out the $\Delta Q_{|\mathrm{V}}$ that inevitably occurs.

The particular $\Delta Q_{|\mathrm{V}}$ that streams experience at a multiplexing point is the result of the 'game' that is being played out there, for the ingress and egress of the buffer. FIFO represents one set of rules for that game, but there are others. The game is driven by the arrival patterns of the streams as they pass through the multiplexing point. Although in many games there are notions of 'winner' and 'loser', the measure of success for the statistical multiplexing game is more complex. Indeed, the notion of success, and the value of delivering performance bounds, is an area with which the industry is only beginning to engage.

Although success may be difficult to quantify, the notion of failure is more amenable to analysis. Application performance over broadband is a technically sophisticated topic, but at its highest level the objective is delivering the outcome required in a suitably bounded time. The technical aspects of this can be summed up as delivering a bound on ΔQ so that the application remains within its *predictable region of operation*.

The internet design philosophy is one in which control traffic (such as routing updates) and data traffic traverse the same paths using common infrastructure. Thus some of the services that need to be kept within their PRO are essential to the effective operation of the Internet as a co-operating system. Typically, such services are maintaining associations (routing information, tunnel/encapsulation keep-alive exchanges) or detecting their failure (to manage redundancy and resilience). Failure to meet the translocation constraints for these services arms an operational hazard that may have wide-ranging effects[46]. Trying to avoid such hazards often drives the deployment of different traffic management approaches. This is an attempt to maintain suitable translocation quality for 'key' applications (the notion of what is 'key' being driven by other concerns).

Inevitably, in a relatively new and technical subject, thinking is often driven by analogy with other areas or experiences. The term 'traffic' naturally evokes other applications of that word, but the nature of network packet traffic means that management and mitigation strategies from other sectors may not apply. Significantly, it is not possible to have control information flow faster than the packets themselves, which restricts the applicability of control theory. This has implications for the potential efficacy of control loops, in particular congestion management.

[46] For example, when congestion on a path delays router updates too much, routers may conclude that the path is no longer available and so update their tables, shifting traffic onto another path that then becomes congested, and so on. This contributes to 'router flap'.

B.4.1. Prerequisites for deployment of differential treatment

In order to apply TM, e.g. to maintain key services within their PRO, it is essential to be able to distinguish different components of the traffic. This requires some form of classification. This classification is typically performed on association information (addressing information or packet marking), though it can also be based on recent offered load or on the packet contents (through use of DPI). It should be noted that a particular marking does not imply that a particular treatment will occur, as the mapping between marking and treatment is determined by the per-device configuration.

B.4.2. Priority queueing

Priority queueing operates by differentially servicing the egress from the buffer. Packet flows are assigned to particular treatment classes on the basis of some classification (as described above).

Ingress

On arrival, a packet is admitted to the buffer if there a free slot, as with the tail-drop behaviour in the FIFO case in §B.1.1. There are two common variants of buffer management:

1. where the buffering is shared amongst all the queues;
2. where there is an allocation of buffering to each treatment class.

Thus the loss element of $\Delta Q_{|\mathrm{V}}$ can be influenced either by all traffic or by just a subset of that traffic. This choice determines the exact nature of the coupling that occurs between the streams, within the constraint of there being two degrees of freedom in all finite queueing systems[47].

Egress

The egress treatment (and hence the delay component of $\Delta Q_{|\mathrm{V}}$) is determined by the relative precedence of treatment classes. Packets are serviced from the highest precedence treatment class first. Packets are serviced from lower precedence classes only when all higher precedence classes are empty. Within a treatment class, packets are serviced in order of arrival[48].

Discussion

The highest precedence traffic experiences lower mean delay and a lower delay variance than other traffic. It also gets the strongest isolation from other streams, with its delay being affected only by other traffic in the same class[49]. Traffic from other precedence classes can potentially experience large perturbations in delay, depending on both the volume and arrival pattern of traffic in higher precedence classes. Where there is per-treatment-class buffer allocation, the collective arrival pattern of all higher precedence treatment classes can cause the buffers for lower precedence classes to fill. This has the effect of differentially allocating loss to the lower precedence classes. If the buffer is shared, the loss rate is the same for all treatment classes[50]. This is illustrated in Figure B.2.

If the highest precedence treatment class arrival rate is not limited (either explicitly in the device, or implicitly by other design constraints or traffic management approaches), then the lower precedence treatment classes can experience an effective denial-of-service.

[47] These two degrees of freedom are loss and delay.

[48] This is typically implemented by placing arriving packets at the end of the corresponding treatment class queue, and by servicing non-empty queues in precedence order.

[49] Traffic in lower precedence classes can affect higher precedence classes only if it is already being serviced. When this is the case, higher precedence traffic is delayed by any residual packet service time.

[50] This assumes the conditions mentioned in §B.1.1.4 are met. That is, that arrivals are Markovian etc.

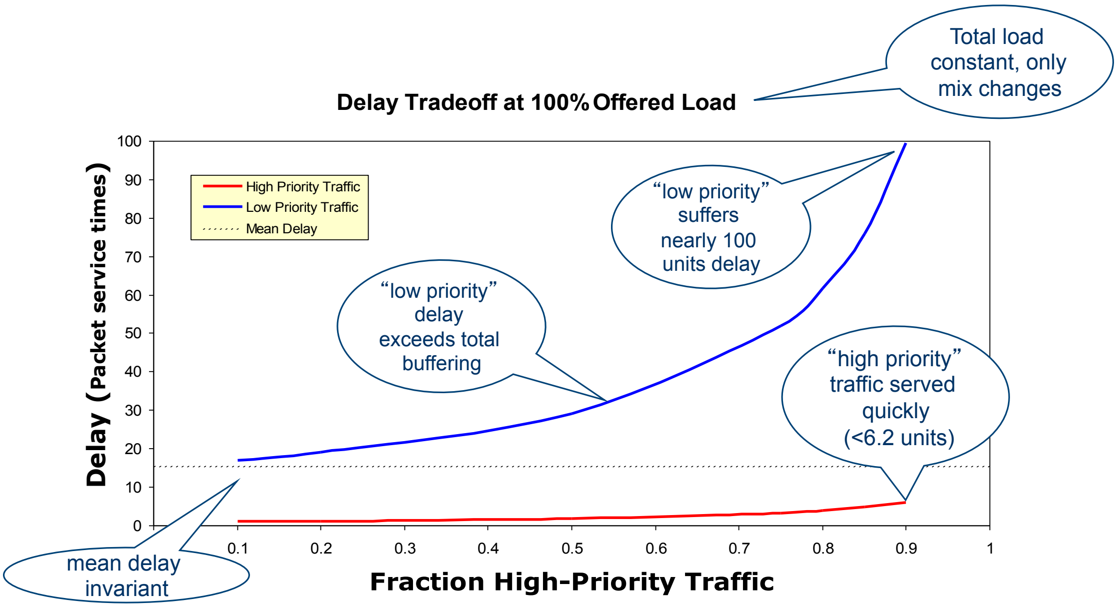


Figure B.2.: Differential delay in a two-precedence-class system (with shared buffer)

B.4.3. Bandwidth sharing

Bandwidth sharing endeavours to share egress capacity so as to deliver some minimum service capacity (‘bandwidth’) to a set of streams over a long period. It does not aim to deliver a particular bound on, or ordering between, the $\Delta Q_{|\mathrm{V}}$ of any of these streams[51]. Differential treatment is expressed as the shares (which may be equal) that the (dynamic) set of competing streams receive (the comments on classification on the previous page apply). The design and discussion of queueing and scheduling of this type is underpinned by the fluid-flow approximation[52] [30]. There are several approaches to implementing this approximation (such as round-robin, hierarchical token bucket shaping, etc.) [31], and substantial variation in the way that the resulting sharing can be configured within a particular implementation. The aim here is not to discuss these differences, but rather to describe the common effects of bandwidth sharing on the distribution of the $\Delta Q_{|\mathrm{V}}$ inherently created in the multiplexing process.

Ingress

On arrival, a packet is admitted to the buffer if there is a free slot, as with the tail-drop behaviour in the FIFO case in §B.1.1. The total buffering can be shared, or can be reserved/limited on a per-stream basis.

Egress

In bandwidth sharing, the way that one stream’s $\Delta Q_{|\mathrm{V}}$ is affected by the other streams depends on both their offered load and the number of streams that are active. We will consider examples (see Cases 1 and 2 below) of these two distinct couplings before discussing their composite $\Delta Q_{|\mathrm{V}}$ effects and the consequences on example deployments.

A useful mental model is to consider bandwidth sharing as dividing the traffic into separate queues amongst which the egress service capacity is distributed in some fashion (i.e. as a collection of FIFOs with continuously varying egress service, see Figure B.3).

[51] Delivering more ‘bandwidth’ to a stream does not imply that its $\Delta Q_{|\mathrm{V}}$ will be better; it depends on the balance of supply and demand.

[52] The fluid-flow approximation treats packets as entities whose service can be ‘spread out’ over time. This fails to capture aspects of the discrete service time of packets.

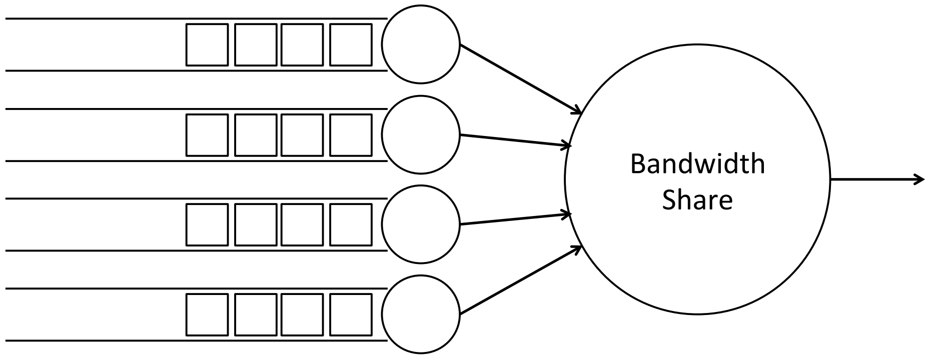


Figure B.3.: Bandwidth sharing viewed as a collection of FIFOs

Case 1: offered load variation and effects on $\Delta Q_{|\mathrm{V}}$ stationarity Consider the situation where there are two such streams[53], A and B, each being assured 1/2 of the egress capacity[54]. Let us also assume that each has access to any spare capacity that is not being used by the other (this is a commonly used configuration).

Imagine the situation where stream B is idle, and stream A is using 60% of the overall capacity (which is 120% of its assured capacity). In this circumstance, queue A behaves as a FIFO under a 60% load, exhibiting all the properties discussed in §B.1.1. Now imagine that B becomes active and offers a 45% load (90% of its assured capacity). The system is now being offered a total load of 105%. Stream A is receiving 55% of the egress capacity to carry 60% of the load, and so it is operating at just over 109% loading. The result is that queue A fills and starts losing in excess of 8% of its packets[55], until the offered load is reduced. Thus, the $\Delta Q_{|\mathrm{V}}$ experienced by stream A is coupled to both its offered load *and* the offered load of B. As a result, the overall ΔQ for stream A can vary quickly.

As was discussed in §B.2 there is no explicit feedback in PBSM. The network relies on the originators of $A's$ traffic reducing their load. The end-to-end principle means that the time to reduce the load at this single point is dependent on how quickly these remote systems can adjust their behaviour. This is related to the round-trip time these systems are experiencing, which includes the $\Delta Q_{|\mathrm{V}}$ induced by the change in the offered load through queue B.

Thus bandwidth sharing induces variability in the distribution of $\Delta Q_{|\mathrm{V}}$, which manifests as non-stationarity[56] for the constituent translocations passing through it.

Case 2: active streams and the effect on $\Delta Q_{|\mathrm{V}}$ stationarity Consider the situation where there are a number of streams[57], each receiving the same share of the egress capacity (when all are busy). In this situation, the $\Delta Q_{|\mathrm{V}}$ of any one stream is affected not only by the collective offered load of the others, but also by how many of them are active at any point in time.

Imagine that stream A's load is low compared to its assigned share. Whenever it has just received service, it must wait to be serviced again until each of the other busy streams has received service. Thus the time between successive services for stream A is proportional to the number of other busy streams. So its $\Delta Q_{|\mathrm{V}}$, in particular the delay distribution, is now

[53] When using the mental model mentioned above, it is important to recognise the difference between a 'stream' and a 'queue'. A stream, in this context, is a packet-flow of information (from a particular source and/or to a particular destination). A queue is the logical arrangement of packets, from a stream, within a network element. In this example, the two queues also separate the treatment of the two flows, however in general several flows may share one queue.

[54] When both queues are non-empty they are serviced so as to get a 50:50 share.

[55] The exact proportion will depend on A's arrival pattern and the distribution of B's busy period.

[56] Non-stationarity is the variation of $\Delta Q_{|\mathrm{V}}$ over timescales of seconds or less.

[57] Access network elements will often have several hundred configured streams (at least one per end customer). In core networks (where MPLS or Carrier-Ethernet services are sold) such streams can be used in significant numbers to offer bandwidth-based service levels.

dependent not only on the overall load, but on how many other streams are busy. This can vary in the short term (introducing non-stationarity), as already configured streams become busy. It can also vary over longer time scales, as the configured number of streams grows. This increases the potential non-stationarity.

Discussion

Here we consider some concrete examples of: how bandwidth sharing can be deployed; the magnitude of the effect on $\Delta Q_{|\mathrm{V}}$; and the possible consequences of the resulting variation in $\Delta Q_{|\mathrm{V}}$.

A commercial response to the costs of delivering high-speed mobile broadband has been to create RAN sharing agreements. One of the ways in which costs are saved is by sharing backhaul connectivity. This sharing is typically contracted on a bandwidth basis. Picking up on Case 1 above, consider a 400Mbps Carrier Ethernet link to a shared MNO base station. This link is shared by two MNO partners, each having an assured 200Mbps, with access to the full 400Mbps in bursts. Under some general assumptions[58], in the initial scenario (stream A offering 60% load, while stream B's load is negligible) the $\Delta Q_{|\mathrm{V}}$ that A experiences is effectively $\varnothing$[59]. After B's offered load increases to 45% (which drives the system to 105% capacity) the $\Delta Q_{|\mathrm{V}}$ for A will rise, to a loss of over 8% with mean delays ranging between $2ms$ and $50ms$[60]. Although elastic network protocols (e.g. TCP) will adapt, this takes several end-to-end round trip times[61].

This $\Delta Q_{|\mathrm{V}}$, if evenly distributed amongst the sub-streams in A, would represent an operational hazard to both control protocols (where loss and delay stability are important) and voice streams (for which it is consuming a substantial portion of the end-to-end delay variation budget[62]). This is an example of how variation of offered load generates variable $\Delta Q_{|\mathrm{V}}$.

Referring to Case 2 above, as an example of how the number of active streams affects $\Delta Q_{|\mathrm{V}}$, consider a multiplexing point in the downstream path to broadband end-users (e.g. a Head End or BRAS from Figure C.1 on page 79). As the number of active users increases, so does the time between service of their individual streams. Not only does this change in the distribution of $\Delta Q_{|\mathrm{V}}$ potentially influence the user experience (particularly affecting services like gaming and VoIP), if it persists it can also cause issues with timing protocols[63].

Unlike in Case 1 (where stream A could return to a stable $\Delta Q_{|\mathrm{V}}$ by reducing its load to a suitable level, however long that might take), achieving a stable $\Delta Q_{|\mathrm{V}}$ is now dependent on the the number of other active streams. Thus, no level of reduction in the volume of traffic for stream A will achieve the desired result. This is where some differential distribution of

[58] These assumptions are: the fluid flow approximation; the Markovian arrival/service assumption; that each queue is treated as $M/M/1/K$; that the packets in question are full QinQ Ethernet packets of 1546 octets, including both the frame overheads and the inter-packet gap; and that the change in the egress capacity changes the Markovian service rate of the queue under consideration.

[59] There is a slight dependence on the number of buffers, but the loss would be $\ll 10^{-7}$pkts/sec and the mean delay would be around 28µs.

[60] The loss and delay depend on the buffering available, for example: at 50 buffers the mean delay would be about 75% of the buffering capacity or $2.17 \times 10^{-3}s$; at 100 buffers, 88% ($4.95 \times 10^{-3}s$); at 500 buffers 97.6% ($27.4 \times 10^{-3}s$); and at 1000 buffers 98.8% ($55.5 \times 10^{-3}s$). Note that the effective service rate has dropped. In the first scenario the *effective* packet service time (for the queuing calculations) was 30.9×10^{-6}, in the second scenario it becomes $56.2 \times 10^{-6}s$.

[61] The RTT could vary from $100ms$ to $500ms$ or more, depending on the location of the sources of the variable data load. Note that the congestion feedback signal (the discard of packets) does not occur until the queue is full, i.e. there is already - for the 1000 packet buffer case - $50ms$ of additional delay to the round trip.

[62] ITU Y.1541 suggests that the one-way voice delay budget should be bounded at $\leq 150ms$ mouth-to-ear, with $50ms$ due to delay variation. ETSI TS 103 210 V1.2.1 suggests that access networks should induce $< 35ms$ of "jitter".

[63] Protocols such as NTP and IEEE1588 underpin the deployment of small cell technology over commodity broadband infrastructure. The response of such a cell to the sort of variation described is to conclude that the local clock has suffered a precision failure. This, in turn, can result in the device using an incorrect frequency at the radio interface or having to reset.

$\Delta Q_{|\mathrm{V}}$ may be required for keeping key services (be they user or system orientated) within their PRO.

A common sharing model is to offer a guaranteed lower bound, together with a potentially achievable peak bandwidth for a particular stream (a 'committed/peak' model). In broadband the peak may be set: implicitly by the sync rate of the line (as in xDSL); explicitly (as in Cable); or by a combination of both (as in 3GPP). The lower bound is implicitly determined by the number of end-users connected to the last hop multiplexor. Between these two limits there is an even distribution of the capacity amongst the active end-users. Such scheduling arrangements can act as a bandwidth leveller; as the number of end-users (and their offered load) increases, those end-users that have the highest peaks (whether explicit or implicit) will start to experience a reduction in their rates *before* those with lower peak rates. This leads to the situation where those with the highest peak data rates actually experience the largest variation in delivered rate and $\Delta Q_{|\mathrm{V}}$ non-stationarity (which many may consider counterintuitive).

B.4.4. Rate shaping

As has been discussed above, a key factor determining the amount of $\Delta Q_{|\mathrm{V}}$ occurring at a contention point is the arrival pattern of the offered load. Rate shaping [32] is a mechanism used to mitigate the 'worst' patterns of such offered load. It is, in effect, a tail drop FIFO that is used to adjust the inter-packet gap and discard packets when the ingress rate exceeds the egress rate for a sufficient period of time.

Ingress

On arrival, a packet is admitted to the buffer if there a free slot, as with the tail-drop behaviour in the FIFO case in §B.1.1. Since the packet is dropped if the buffer is full, rate shaping also implicitly limits the maximum sustained rate.

Egress

Packets are sent in the order they are received. The average rate at which packets can depart is fixed. When the instantaneous arrival rate exceeds this fixed rate[64], the packets experience $\Delta Q_{|\mathrm{V}}$.

Discussion

Resource sharing in PBSM is rivalrous, in that consumption by one user affects the service delivered to others. Rate shaping, by modifying the arrival pattern of the offered load, can help place some limits on these effects.

The limiting aspect can be used to enforce the demand side of peak-rate-based contracts. Such rate shaping can also be applied to specific sub-streams of user data. When such sub-streams are being offered a 'higher quality' translocation, it can be used to ensure that the offered load conforms to agreed limits.

B.4.5. Rate policing

Another approach to demand-side management is rate policing [32]. This approach does not introduce any $\Delta Q_{|\mathrm{S}}$ or any delay part of $\Delta Q_{|\mathrm{V}}$. Its function is to drop (or re-mark) packets that exceed a pre-configured rate, so it contributes to the loss component of $\Delta Q_{|\mathrm{V}}$.

[64] There are several implementation choices which trade the ΔQ of the rate limiter between $\Delta Q_{|\mathrm{S}}$ and $\Delta Q_{|\mathrm{V}}$, e.g. Token Bucket shaping. They can also permit bursts (where the packet rate exceeds the rate limit for a short time), which have the effect of moving the point at which $\Delta Q_{|\mathrm{V}}$ accrues further downstream.

Ingress

The packet is discarded, or re-marked, if the time since the arrival of previous packets is too short. The time would be deemed to be too short if the short-term average packet arrival rate exceeds some preconfigured limit[65]. There is no buffer, as Rate Policing is never performed at a transmission egress point.

Egress

Packets that are admitted are passed on immediately in the order of arrival.

Discussion

Rate policing and rate shaping (discussed in B.4.4 above) take a different approach from the methods previously discussed to sharing out the overall $\Delta Q_{|\mathrm{V}}$. FIFO, priority queuing and bandwidth sharing are concerned with sharing $\Delta Q_{|\mathrm{V}}$ at a particular multiplexing point, whereas rate policing and rate shaping are concerned with how it is shared along a path. By modifying the arrival pattern (introducing additional $\Delta Q_{|\mathrm{V}}$ locally) they are changing the $\Delta Q_{|\mathrm{V}}$ that would occur downstream.

Rate policing (using re-marking) is currently deployed in UK broadband delivery (as outlined in BT Supplier Information Note 506, discussed in Appendix D).With the general increase in the variety of services using PBSM, it is likely that the use of rate policing and rate shaping techniques will increase. If some services were assured, this assurance would come at a cost of worse ΔQ for others, and/or increased costs on the provider to maintain the overall service level. This would mean that there would be an advantage for other traffic to masquerade as an assured service. This is a hazard already armed by sharing the last mile between different services (for example VoD and CDN distribution).

B.5. Factors influencing the further deployment of traffic management

It is in the very nature of statistical multiplexing that $\Delta Q_{|\mathrm{V}}$ will increase with the number of subscribers[66]. From a ΔQ perspective, the trend in broadband provisioning has been a race between improving technology and increasing demands. Improving technology has created faster links (thereby decreasing $\Delta Q_{|\mathrm{S}}$) and increased switching capacity (thus potentially reducing $\Delta Q_{|\mathrm{V}}$). Rising demand, in overall and in individual peak quantity (due, in particular to higher access link speeds), increases $\Delta Q_{|\mathrm{V}}$ and its non-stationarity. This increasing non-stationarity is not only a challenge for 'critical services', whose PRO is essential for network stability, but may become an issue for some end-user services, depending on their sensitivity.

If there is a demand for more consistent delivery of end-user services whose PRO cannot be maintained by the current 'best effort' approach, the service paradigm may have to change. Attempting to deliver all traffic within the ΔQ bound of the most sensitive applications (allowing "the needs of the few to drive the costs of the many") may prove to be commercially unsustainable. This could be a strong driver for an increase in the use of differential traffic management.

[65] There are several approaches to implementing this; Token Bucket is one that is commonly available

[66] This phenomenon was heavily exploited by users during the earlier dial-up internet days when the 'smart thing' was to hop from one ISP to the next as they launched.

B.6. Static and dynamic allocation of translocation resources

As has been seen, a key element of traffic management is the process of deciding whether, and in what order, to deliver service (be that ingress or egress). The outcome of these decisions is the distribution of $\Delta Q_{|\mathrm{V}}$ across competing streams.

Although we have described the broad range of traffic management approaches that are available, there is another important factor that also needs to covered. This is the difference between scheduling at a single point (that has been discussed so far) and the distributed scheduling that is required over a shared medium. In broadband deployments, such distributed scheduling is found mainly in the last mile. It is a key characteristic of: 3GPP-based systems (2G/3G/4G-LTE); 802.11 WiFi (and similar); the upstream (from the customer premise) of DOCSIS cable and Satellite systems[67]; and some last-mile fibre systems, depending on the deployed variant and its configuration. Such systems deliver connectivity when the physical medium is shared, and are able dynamically to allocate resources in order to deliver high peak rates to individual users.

Where multiple 'talkers' share a common medium, this creates a *distributed contention domain*. In such a distributed contention domain, permitting any arbitrary end-point to talk at will gives neither a predictable service nor efficient use of the shared medium[68]. The typical approach to improve the PRO of such systems is to use an arbiter, an entity that receives requests for service and grants access to a portion of the shared medium's capacity[69]. This has consequences for both the $\Delta Q_{|\mathrm{V}}$ and the inherent efficiency of the aggregate system[70].

Consider how a conversation would start. Initially, a talker has to arrange for some of the shared resource[71] to be allocated to it[72] (thereby removing that capacity from the common pool). Not only does this take time, it also has to be achieved by use of un-arbitrated capacity[73]. This allocation process has an associated $\Delta Q_{|\mathrm{V}}$, that depends on the instantaneous demand on the un-arbitrated capacity from all the other talkers[74].

As a talker increases its demand, there is a time lag in being granted more capacity (if it is available) which contributes to $\Delta Q_{|\mathrm{V}}$. When a talker reduces its demand, there is a corresponding time lag before that resource can be allocated to another talker, causing some inefficiency in the use of the shared medium[75].

B.7. Consistent performance and a heterogeneous delivery chain

The Internet comprises a heterogeneous delivery chain. It is a set of autonomous entities (telcos of various tiers, carriers, ISPs, etc.) that can be seen as collectively constructing connectivity, with emergent translocation performance. The equipment under the control

[67] In the downstream direction these systems have visibility of the instantaneous offered load at their head-end, and therefore can apply the scheduling techniques discussed above.

[68] For example, random talking (Aloha) access restricts the PRO to < 18% of capacity (under a Markovian arrival assumption). This rises to about 36% when access to the resource is slotted (as is used for grant requests in DOCSIS and 3GPP). See `http://en.wikipedia.org/wiki/ALOHAnet` .

[69] In 3GPP this would be the RNC or the eNodeB, in 802.11 the access point, and in DOCSIS the CMTS.

[70] The arbiter scheduling the upstream capacity typically has the same emergent behaviour as bandwidth sharing systems (described above). In particular, with a large number of active talkers, the interval between grants for service for any particular talker can become so large (and so variable) that higher-level protocols (such as TCP/IP) are pushed outside their PRO.

[71] This could be a time slot in DOCSIS or a portion of the code space in 3GPP.

[72] By doing this, it is effectively setting its future $\Delta Q_{|\mathrm{G,S}}$ for this hop for the period of the allocation.

[73] This typically operates as described in footnote 68.

[74] If the number of talkers becomes too large, the demand on the un-arbitrated grant/request capacity can exceed its PRO (as described in footnote 68), resulting in a failure to issue grants and a consequent denial of service.

[75] The speed with which resources are allocated to and deallocated from individual talkers is thus a key factor determining the tradeoff between delivered ΔQ and the efficient use of the shared medium.

of a single entity (which represents an administrative domain) may be operated by several management domains. Thus, traffic being translocated end-to-end may well cross a plethora of management, administrative and even regulatory jurisdictions.

To keep a particular application within its PRO, the 'sum' of the ΔQ across all the multiplexing points traversed has to be sufficiently bounded and $\Delta Q_{|\mathrm{V}}$ sufficiently stationary. Since ΔQ is conserved, if the $\Delta Q_{|\mathrm{V}}$ accrued through even a few multiplexing points is large, the delivered delay and loss may cause the application to exceed its PRO. From the end-user's perspective, the application will have 'failed'.

When constructing connectivity the entities do not, typically, contract to any instantaneous performance guarantees[76]. They may aspire to a given level of of availability, or even some assurance of reachability, but performance is almost always provided on a purely 'Best Efforts' basis[77].

Detecting a lack of connectivity is a relatively clear-cut process, as there will be an identifiable location that is either reachable or not. The additive nature of performance impairment means that identifying the underlying location of a performance issue (let alone its root cause) is more difficult. The ΔQ that accrues along a section of the end-to-end path is indistinguishable from that which accrues along another - it is only their combined effect that can be observed at an end-point.

For any entity in the end-to-end delivery chain, maintaining consistent performance as load increases may be difficult to justify commercially. This is especially true given the absence of contractual performance guarantees, and the current inability of interested parties to pinpoint where ΔQ accrues. There are particular problems in access networks, which are commercially predicated on amortising capital costs over a large number of customers. This creates the natural tendency for the multiplexing points in the network to be run "hotter" over time, resulting in increasing $\Delta Q_{|\mathrm{V}}$.

Along any end-to-end path several multiplexing points may be manifesting this trend of increasing $\Delta Q_{|\mathrm{V}}$. If the end-user experiences an 'application failure' (due to translocation being outside the PRO), it could be simply due to the aggregated effects of these commercial trends, rather than due to any one entity's traffic management policies.

[76] Guarantees that might be offered would be in terms of averages, usually over long periods of time (e.g. a month). Such measures cannot be composed along a path, and entities do not take on the risks of their suppliers or onward connections, so these assurances are not 'transitive'.

[77] "Best Efforts" in internet network terms - `http://en.wikipedia.org/wiki/Best-effort_delivery` is at complete variance with the use of the term in UK commercial practice `http://dictionary.cambridge.org/dictionary/business-english/best-efforts` .

C. The Internet in the UK from a traffic management perspective

> "Internet: A global computer network providing a variety of information and communication facilities, consisting of interconnected networks using standardised communication protocols." - OED

Evidently, in the current commonly-understood definition of the Internet, computation and communication capabilities are mingled together. The term "internet" was originally used to refer solely to information exchange capability; now "internet" is used to refer to the facilities based upon this capability.

Here we are going to focus on how consumers are provided access to this global network from the premises onwards[1]. This description is intended to highlight the major boundaries within, and distinctions between, different widespread forms of access provision. Despite being quite detailed, this is not a complete statement of the UK market, as there are additional delivery models (e.g. specific fibre-to-the-premise installations or point-to-point/multipoint wireless) that will not be covered here. There is also a significant simplification of the international picture, which is included for reference.

C.1. UK network administrative and management boundaries

A consumer of internet access services (whether domestic or commercial) has to have a connection to one of a variety of infrastructures. The UK market has a considerable diversity of structure and technology in terms of ISPs' service provision. We will cover this in more detail later in this section, but for now let us highlight an aspect common to all, namely that the management of a user's connection to the wider Internet is split across different entities, some internal and some external.

The UK ISP market is based around the construction and reselling of connectivity monopolies. For example, BT OpenReach may provide the physical communication path between a premise and the first active component (this being a natural monopoly). That monopoly is 'sold', on a per end-point basis, to either an LLU unbundler or BT Wholesale, who then have sole use of it. They take that physical circuit and 'activate' it (placing active components at each end), which creates point-to-point connectivity (monopolistically). That connectivity is then (in the case of BT Wholesale) sold on to the ISP[2], who has a monopoly over all the traffic sent and received by the end-user[3].

The organisations that supply broadband tend to be large and so are split into multiple internal, semi-autonomous, management domains (silos). Each of these management domains is assigned key performance indicators that it then works to optimise[4]. This pattern is common to all of the services provided in the UK, the differing factor being whether (and how)

[1] Local network conditions (LANs, router configurations, wireless interference) are not going to be considered.

[2] Other services can be multiplexed into that connectivity, but only within the same exclusivity constraints. These interactions of other services with the 'normal' ISP function represent a potential source of performance impairment, as is clear from the BT Supplier Information Notes (SINs: discussed in Appendix D). Specific quantitative investigation of this topic is outside the scope of this report.

[3] This is why end-users typically assume that the ISP has sole responsibility for the performance of their services.

[4] Unfortunately, it is a general feature of the engineering of complex systems that the combination of local optimisations rarely delivers a globally optimal result.

these management domains are split into different administrative domains. In Figure C.1 we see that vertically integrated ISPs, for example Virgin Media's cable division, have different management domains that all fall under the administrative domain of one company. With ISPs like BT Retail, though the consumer deals with BT Retail their connection is managed at various points by the separate companies (and thus administrative domains): BT Retail, BT Wholesale, and BT OpenReach.

Each of the boundaries between these different companies' administrative domains is subject to contracts for the supply of services. It should be noted that whether this series of bilateral agreements delivers anything that is enforceable end-to-end is an open question. So, in the example above, BT Retail has no direct contractual relationship with BT OpenReach, even though it is the latter who is responsible for the actual physical connection to the customer[5].

Figure C.1 illustrates the general administrative and management domain structure of the UK broadband market. What consumers consider to be 'their ISP' really represents an administrative domain (encompassing a collection of management domains), which typically provides their service using other administrative domains' services.

Different customers of the same ISP (for example at different locations) can have their service delivered through entirely different collections of administrative and management arrangements. This is an ecosystem that contains a large diversity, not only in the paths that can be taken through the management and administrative domains, but also in the management, configuration and operational polices and practices thereof. Despite this diversity, any given end consumer is presented with only a small number of connection options, typically ADSL and possibly VDSL and/or cable. Changing ISP does not change the physical medium for the final access tail, except in the case of a cable or non-wireline provider.

From a performance analysis perspective, the difference in how the final access tail operates has a significant bearing. Broadly it can be dedicated connectivity (e.g. ADSL) or a contended medium (e.g. cable/3G/LTE), see §B.6. In both cases, the peak potential capacity is asymmetric. A brief comparison of the performance affecting properties of these different access tails can be found in Table C.1.

Figure C.2 provides a simple depiction of how UK network users connect to the wider Internet. The diagram follows upwards from Figures C.1 and C.3. For clarification, the 'Tier 1 Service Providers' are international higher-level CSPs who provide connectivity to other networks in other locations (the multitude of connections from them has not been shown).

C.1.1. Non-Wireline ISP provision

Although the following discussion is going to focus on wireline-based ISPs, we must also consider the wireless side of the UK telecoms industry, illustrated by Figure C.3 on page 82. The techniques and approach in this report can be applied to these networks, but the detailed analysis is out of scope for this study.

Satellite-based ISPs inevitably have a large $\Delta Q_{|\mathsf{G}}$ associated with the path to/from the geosynchronous satellite[6], which affects the efficiency of the upstream resource allocation (the grant of a time/frequency slot to transmit), as discussed in §B.6.

Mobile cellular radio networks have a complex set of resource management issues. These networks have multiple resource constraints and multiplexing points[7], all of which are po-

[5] There is no reason to believe that, because a set of management domains are part of the same administrative domain, they will work 'better' together. Logically, these silos operate to maximise their incentives, which are set by their overarching administrative domain. Administrative domains set objectives (be they fiscal or of another type) and create contractual arrangements with other administrative domains. Management domains take those objectives (as input aspirations), and construct operational steps to meet them (architecting interconnects, planning/provisioning capacity, configuring equipment, etc.). None of this fosters an end-to-end view of service provision.

[6] MEO/LEO satellites have a lower $\Delta Q_{|\mathsf{G}}$, but this varies with time, depending on the orbital position.

[7] These include the GGSN, SGSN, RNC, SGW, PGW, and the connectivity between them, as well as the (e)NodeB air interface.

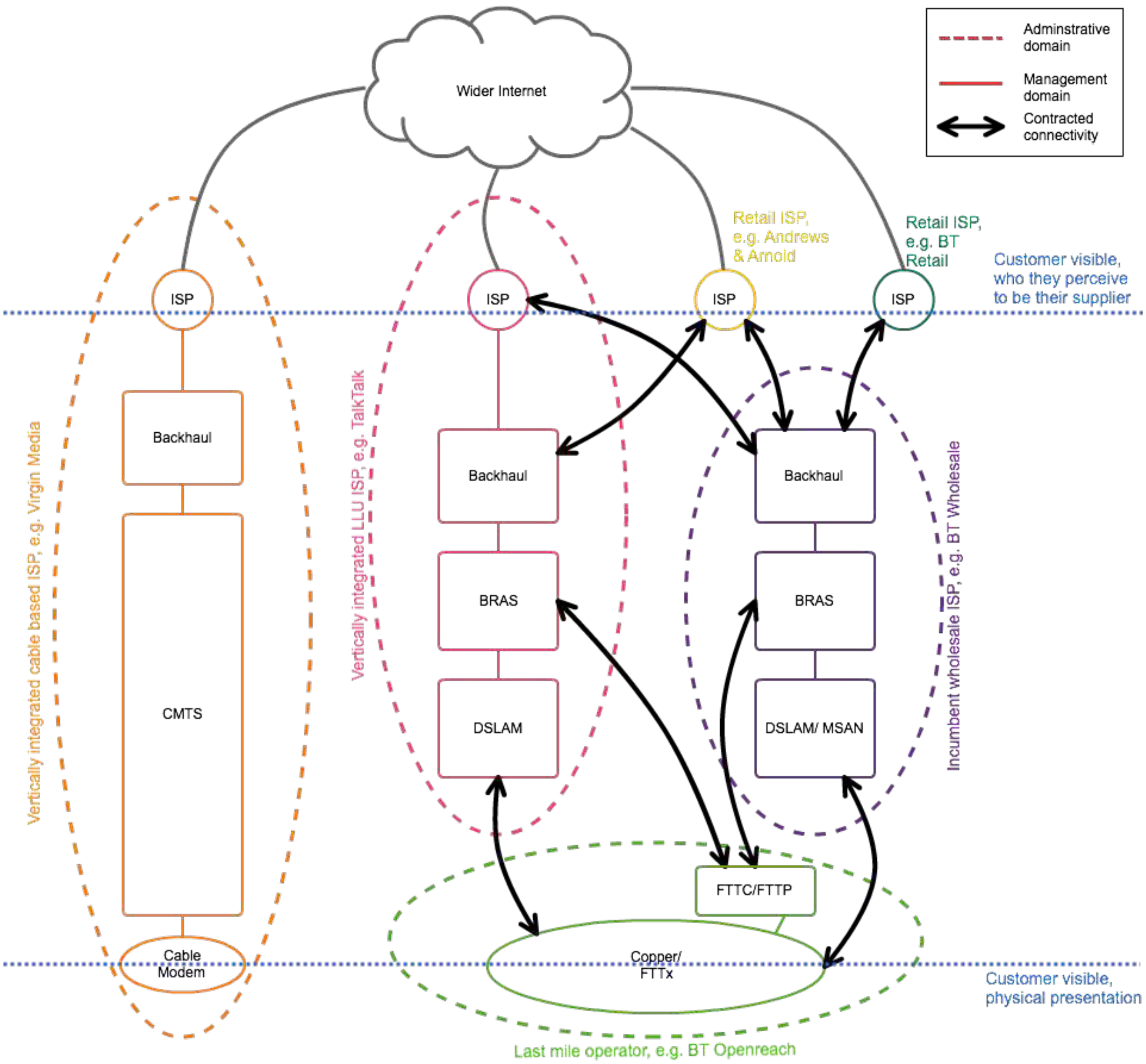


Figure C.1.: Representation of the administrative and management boundaries in UK broadband provision (wireline)

Connection Type	**Capacity Delivery Strategy (to/from final active component)**	**Upstream Access Strategy**	**First/Last queueing component**
ADSL	Dedicated time slots (capacity) in both directions. Capacity environmentally constrained.	Encode packet into cells, place in time slots.	DSLAM/ MSAN
VDSL/ FTTC	Dedicated time slots (capacity) in both directions. Capacity environmentally constrained.	Encode packet into cells/frames, place in time slots.	Street DSLAM
FTTP[a]	Dedicated capacity in the upstream (to OLT). Downstream overbooked.	CPE must shape. Allocation to peak.	OLT
FTTP[b]	Variable capacity in both directions. The medium is contended but has upstream coordination.	Request upstream capacity (time slots), when granted fill with packet(s).	OLT
Cable	Variable capacity in both directions. The medium is contended but has upstream coordination.	Request upstream capacity (time slots), when granted fill with packet(s).	CMTS
3G/LTE	Variable capacity in both directions. The medium is contended but has upstream coordination. Capacity is environmentally constrained.	Request upstream capacity (time slots), when granted fill with fragmented packets(s).	(e)NodeB
Satellite	Variable capacity in both directions. The medium is contended but has upstream coordination. Capacity is environmentally constrained.	Request upstream capacity (time slots), when granted fill with packet(s).	Downstream - base station. Upstream - VSAT.
WiFi[c]	Variable capacity in both directions. The medium is contended. Various environmental constraints.	Distributed coordination (detection of non-delivery leads to exponential backoff).	WiFi Access Point.

[a] BT OpenReach-style deployment
[b] Other deployment approaches, included for comparison
[c] Included for comparison purposes only

Table C.1.: Comparison of access connection performance properties.

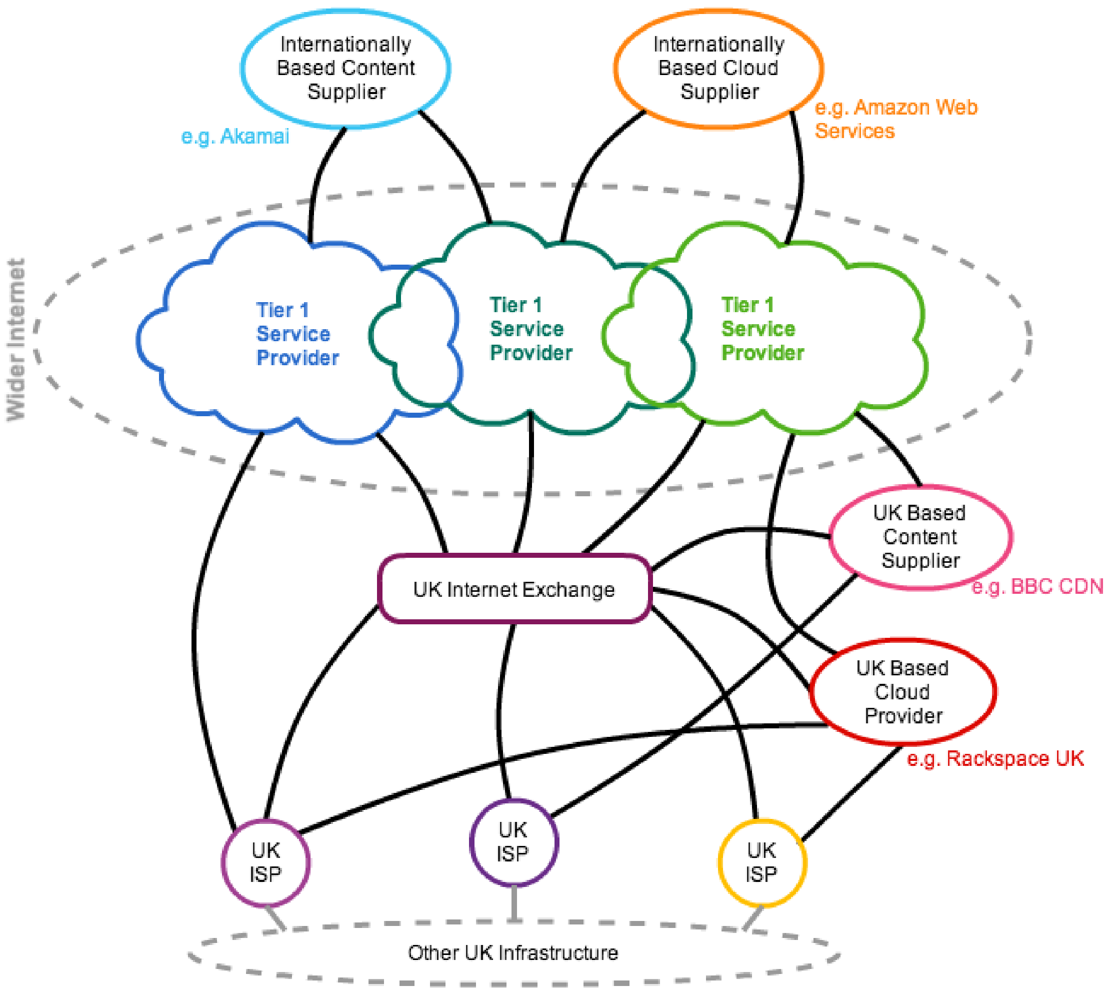


Figure C.2.: UK ISPs in wider context

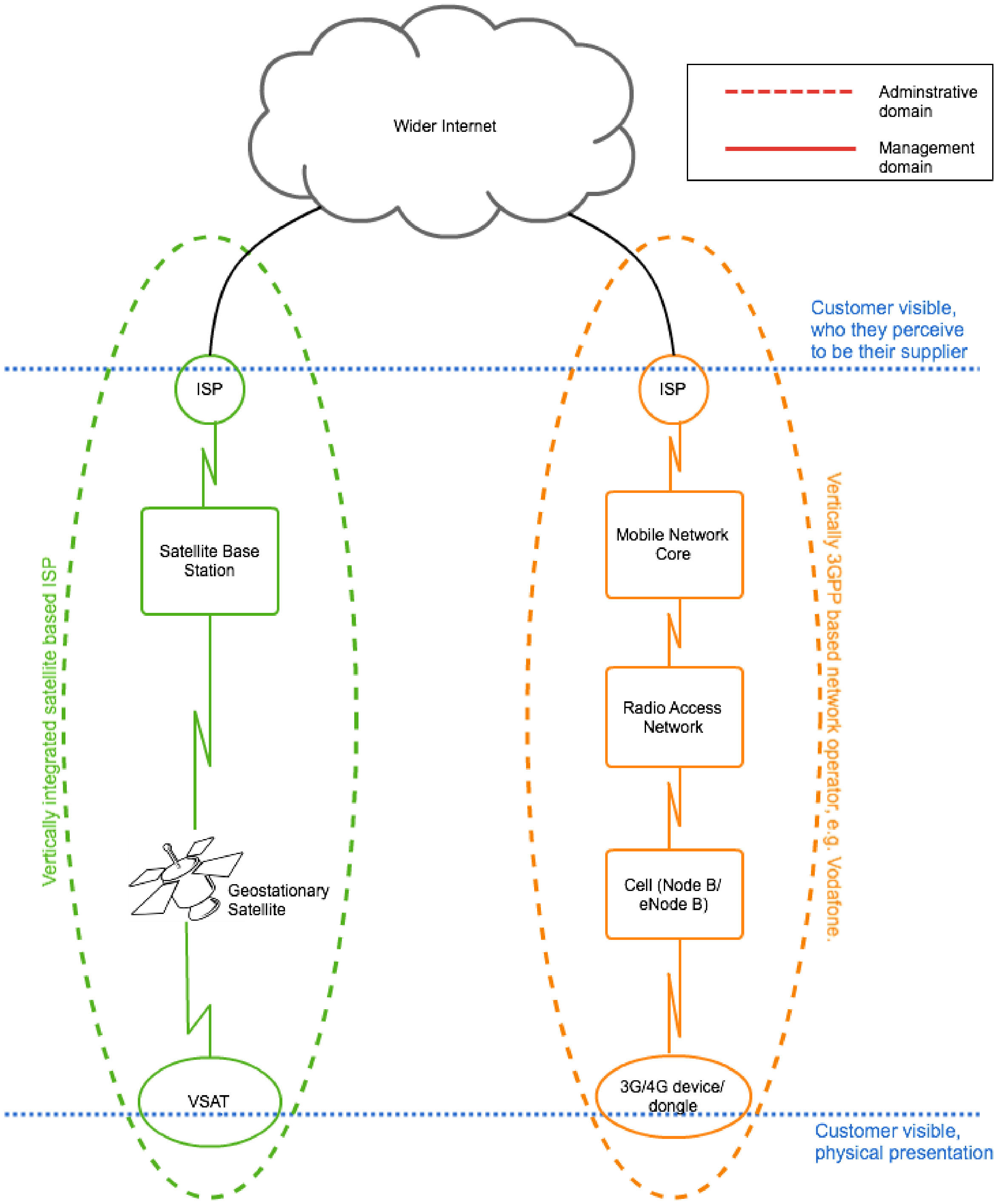


Figure C.3.: Administrative and management boundaries in UK broadband provision (non-wireline)

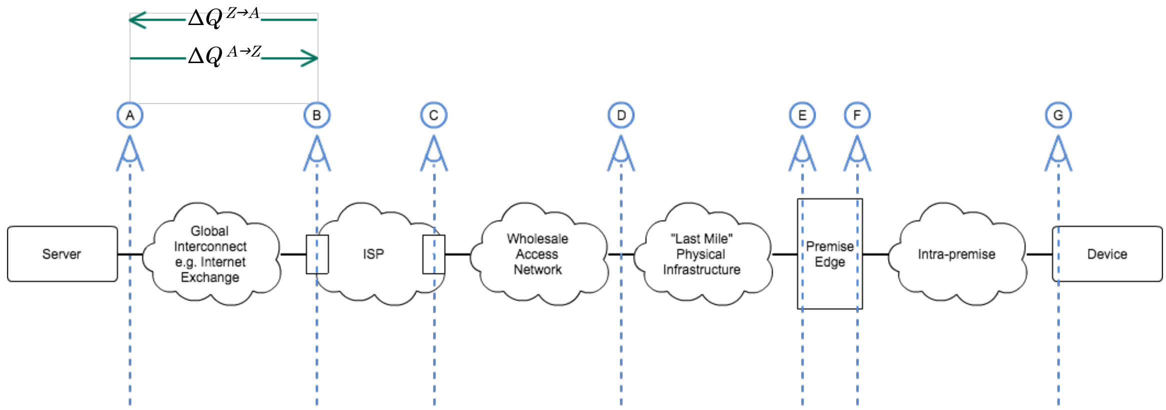


Figure C.4.: Idealised end-to-end path for typical UK consumer

tentially subject to contention and resource saturation. With the advent of LTE, the direct interaction of eNodeBs with one another (a role that was once uniquely assigned to RNCs) further complicates delivering consistent ΔQ whilst mobile.

There is a set of interlinked issues in the scheduling of the air interface resource, touched upon in §B.6. The mobile base station communicates with a number of associated Mobile Terminals (MTs). Each of the MTs may require a different fraction of the overall capacity to achieve a given throughput[8]. The load on the resource is not just dependent on the data being transferred, but also on how the relative MT location and other environmental conditions (including the utilisation of nearby cells) influences the code-space costs for carrying that data. The translocation costs are a function of all these factors.

C.2. How ΔQ accrues in the UK broadband context

As discussed in §A.2.1, applications produce outcomes by exchanging information between protocol peers (e.g. between client and server), and the only aspect of this translocation that affects an application's outcome is the ΔQ experienced by the corresponding traffic flows[9]. As has already been discussed, the $\Delta Q^{A\to Z}$ between two points 'A' and 'Z' is the 'sum' of the ΔQs accrued along the path between them[10]. So, the structural component, $\Delta Q^{A\to Z}_{|\mathrm{G,S}}$, is going to be determined by the actual path taken; while the variable component, $\Delta Q^{A\to Z}_{|\mathrm{V}}$, is going to be determined by the contention at the individual hops. Each such multiplexing point is a location at which $\Delta Q_{|\mathrm{V}}$ both forms and is distributed over the set of competing flows, as discussed in Appendix B. Since ΔQ is conserved (i.e. it cannot be undone), this means that the performance of an application is dependent on the composite effect of the journey of its traffic through different management and administrative domains.

Figure C.4 illustrates an idealised end-to-end path for a typical UK broadband connection. It identifies the major boundaries, some of which represent administrative domains (e.g. $C \leftrightarrow E$ - retail ISP, or $A \leftrightarrow B$ - an Internet exchange), and some management boundaries (e.g. $D \leftrightarrow E$ - DOCSIS headend in cable systems or xDSL in an LLU provider).

The overall experience delivered to the user (via a device) is constrained by the combination of the performance of the application and the performance of the translocation between the components thereof. The application is split, with a portion in the user device (at 'G') and

[8] The quantity of time/frequency code slots needed depends on the signal-to-noise ratio of the corresponding radio bearer.

[9] Note that QoE is solely based on *delivered* ΔQ, and that bounds on delivered ΔQ are not currently offered by any ISP (or administrative domain in the end-to-end path).

[10] Technically, this is the transitive closure of the appropriate convolution operation, i.e. $\Delta Q^{A\to Z} = \Delta Q^{A\to B} \oplus \Delta Q^{B\to C} \oplus \ldots \oplus \Delta Q^{Y\to Z}$.

a portion in the server[11] (at 'A'). The translocation that affects the application performance comprises two unidirectional ΔQs: $\Delta Q^{A \to G}$ and $\Delta Q^{G \to A}$.

Although the delivery path may be composed of many different technologies, their only effect on the user experience is the way in which they affect the ΔQ over each (unidirectional) end-to-end path[12]. As an example, consider the difference between a DOCSIS-based and an ADSL LLU-based ISP service. In ΔQ terms, these would substantially differ from each other only along the bi-directional path $D \leftrightarrow E$ (the last mile physical infrastructure). The DOCSIS service uses a distributed contention domain, which requires coordination (logically occurring at D) for traffic flowing in the $E \to D$ direction. This coordination is performed on-demand and takes a certain amount of time (as discussed in §B.6). Thus, while for the ADSL LLU, $\Delta Q^{E \to D}_{|\mathbf{V}}$ is effectively zero, for DOCSIS it can be several milliseconds. This difference may be irrelevant for bulk data transfer[13], but significant for other applications such as interactive online gaming[14]. Even though the service time for a given packet ($\Delta Q^{E \to D}_{|\mathbf{S}}$) may be three times larger for ADSL than for DOCSIS (i.e. cable can provide higher uplink speeds), the $\Delta Q^{E \to D}_{|\mathbf{V}}$ on DOCSIS can be a factor of 5-10 times greater, thus dominating the ΔQ budget in this direction.

Given that the influence of ΔQ on QoE is over the entire end-to-end path, there is a further complication to address. Although Figure C.1 on page 79 describes the UK side, protocol peers are not just in the UK (as illustrated by Figure C.2 on page 81). An application cannot determine whether excessive ΔQ is accruing in the UK or elsewhere. Many CDNs (even international ones) have UK-based servers to reduce RTTs, but the performance of (and contention on) links around the world still has a bearing on UK user application outcomes.

C.2.1. Specialised services

A 'specialised service' is one that is delivered along a path that does not terminate in the general Internet. It is commercially attractive to run both specialised services and general Internet provision over a common infrastructure; this is especially true for the delivery of such services to consumers, where "last-mile" costs tend to dominate.

Given that the last mile typically has the greatest capacity limitation, there is a strong desire to benefit from statistical sharing. Static allocation of capacity to a specialised service would inevitably mean less available for general internet connectivity. Statistical sharing implies that there is a performance coupling between such services and general internet use. This coupling can have detrimental effects in both directions. Static allocation of capacity would mean that the consumer would experience the capacity penalty at all times, irrespective of whether the specialised service was in use or not.

Today's Internet connection offerings are not sold with any lower bound on their effective performance. Specialised services are likely to be "value-added services" (i.e. those that attract a commercial premium). Many services that are currently offered as value-added (e.g. video conferencing) have stringent ΔQ requirements[15].

Traffic management would be needed to create an appropriate level of performance isolation (and quality trading) required for reliable application outcomes. An early approximation of such TM is already present in the UK market[16]. Given that there is no contractual quality

[11] Many application clients need to interact with a variety of different servers; this analysis applies to each interaction separately.

[12] This ΔQ will depend on the offered load at which traffic constraints bite; such constraints can be due to physical issues, such as maximum achievable data encoding rates, or due to explicit rate limitation.

[13] Bulk data transfer forms the basis for most consumer 'speed test' services.

[14] Interactive gaming represents a use-case for which small differences in ΔQ can have a large impact on the quality of the user experience. See, for example, `http://goo.gl/iqFCpl`.

[15] By this we mean that they have an upper bound on the ΔQ that they can tolerate while still delivering an effective service.

[16] E.g BT's TV Connect service as described in SIN 511, see Appendix D. Other LLU providers will have had to make similar design and implementation choices; descriptions of their choices are not publicly available.

floor for the general Internet, the performance effect of the operation of such specialised services in the last mile is an area that may need investigation in further work.

C.3. Management domain interfaces in the UK

The UK is unusual in that, due to its market structure and regulation, several of the technical interfaces that would be internal in a vertically integrated ISP have become externalised. BT is required to publish "Supplier Information Notes" (SINs) that contain technical information needed to connect to the constituent services offered by BT OpenReach and BT Wholesale. They also make some qualitative (but not quantitative[17]) descriptions relating to traffic management. Thus it is possible to analyse such SINs to extract what statements (if any) they make regarding queuing/scheduling, traffic management, etc.. This is done in some detail in Appendix D.

The SINs themselves do not contain enough information to know the PROs of BT's network elements (i.e. they lack the detailed technical parameters and behaviour descriptions needed to ensure reliable use of their services). However, they state (both explicitly and implicitly) that an ISP should use traffic management to avoid incorrect operation. In fact, they point out that failure to manage correctly certain control traffic could result in the loss of connection to one or even all of an ISPs customers (as described in §B.2).

Such documents are interesting because they expose the interfaces between management domains and highlight the extent to which TM is essential to keep network systems within their PRO. They also reveal that existing specifications do not provide any means to predict or control the emergent ΔQ. They embody an implicit assumption that bandwidth is fungible, which is not the case in a PBSM context. Compounding this, services used in the delivery of consumer broadband rarely provide minimum bandwidth or delay guarantees, particularly in the upstream direction.

In conclusion, the analysis of the BT SINs shows the presence of translocation performance hazards. There is no reason to believe that other UK market providers do not have the same issues. These hazards are present both for network control traffic (affecting system stability) and end-user traffic (affecting application outcomes). Mitigating these hazards is one reason for the deployment of TM in the UK broadband infrastructure.

C.3.1. Potential points of TM application

While (as pointed out in Appendix B) every output port of a switch or router is a point at which packets may queue[18], in practice most of these implement default FIFO behaviour. Points at which non-FIFO TM may be effectively applied correspond to points of ingress and egress between distinct management and administrative domains. Even without this, however, the limited rate of interfaces tends to shape traffic, resulting in the smoothing of bursts. At other points, where both the fan-in[19] and total data rates are low, non-FIFO TM will deliver little benefit.

Let us consider an example of the path from an ISP to its customers via BT Wholesale. The use of non-FIFO TM makes sense[20] at the egress from the ISP to BT Wholesale (point C in Figure C.4). Assuming that this rate-limits the ingress streams to BT Wholesale, there is little to be gained by using non-FIFO TM in most of their network (the path C $\rightarrow$ D in the

[17] BT does make more quantitative information available in commercially confidential 'handbooks'. A flavour of the additional quantitative information can be found in an edition of the FTTC handbook that has become public `http://goo.gl/d6Y5q1`. Several other handbooks with "in commercial confidence" markings can be found on BT's websites through a simple web search.

[18] Thus each such point is where some form of scheduling decision is made.

[19] Where the fan-in is the number of active input ports, or sources, destined for a particular output port, or destination, within a switch or router.

[20] This is because: (a) the connection capacity is finite; and (b) charges are based on averaged peak bandwidth.

Figure would represent multiple paths in this case). However, as traffic from multiple ISPs is dispersed towards different BRASs (points D), there is a potential for correlation (due to demand from the corresponding sets of end-users in a particular geographical area). This creates a performance hazard; mitigating this would require non-FIFO TM to be applied by BT Wholesale.

Figure C.5 is an annotated version of Figure C.1 that illustrates locations where non-FIFO TM might be usefully applied in a UK wireline context[21]. Note especially the different coloured arrows that distinguish the level of aggregation at which TM might be applied.

It is important to consider what 'positive detection' of differential traffic management would mean in a UK context. Knowing that there may be differential management occurring somewhere along the path between an end-user and the internet does not identify which management / administrative domain it occurs in, which could be:

- before the ISP (even outside the UK);
- within the ISP;
- after the ISP;
- in a local network (depending on router settings).

Thus it is a challenge simply to determine whether the 'cause' is within the UK regulatory context.

C.4. Summary

In this Appendix, we have seen how the market structure of broadband in the UK creates quite complex delivery chains, particularly in the wireline case[22]. We have discussed how ΔQ (the observable performance characteristic of the delivery chain) accumulates along the path. We have explored the issue that effects due to remote parts of the network may not be distinguishable from those due to more local causes.

We have used the particular situation in the UK to analyse some aspects of management domain interfaces (explored in more detail in Appendix D), in particular the requirement for TM to be applied to keep sections of the network within their PRO. We have identified potential points at which non-FIFO TM might be usefully applied along the delivery chain, both for this reason and others.

An important point to reiterate is that, even if TMD could detect differential TM, the structure of the UK market (coupled with the current state of TMD research) prevents one from being able to ascribe responsibility to any single body[23].

[21] Establishing which forms of TM applied at these various points might be detectable would require a laboratory-based study, including an emulation of an appropriate subset of the UK broadband infrastructure.

[22] Mobile and satellite providers remain, for the most part, vertically integrated.

[23] As it stands, one could not even be sure such differential TM was being applied in the UK.

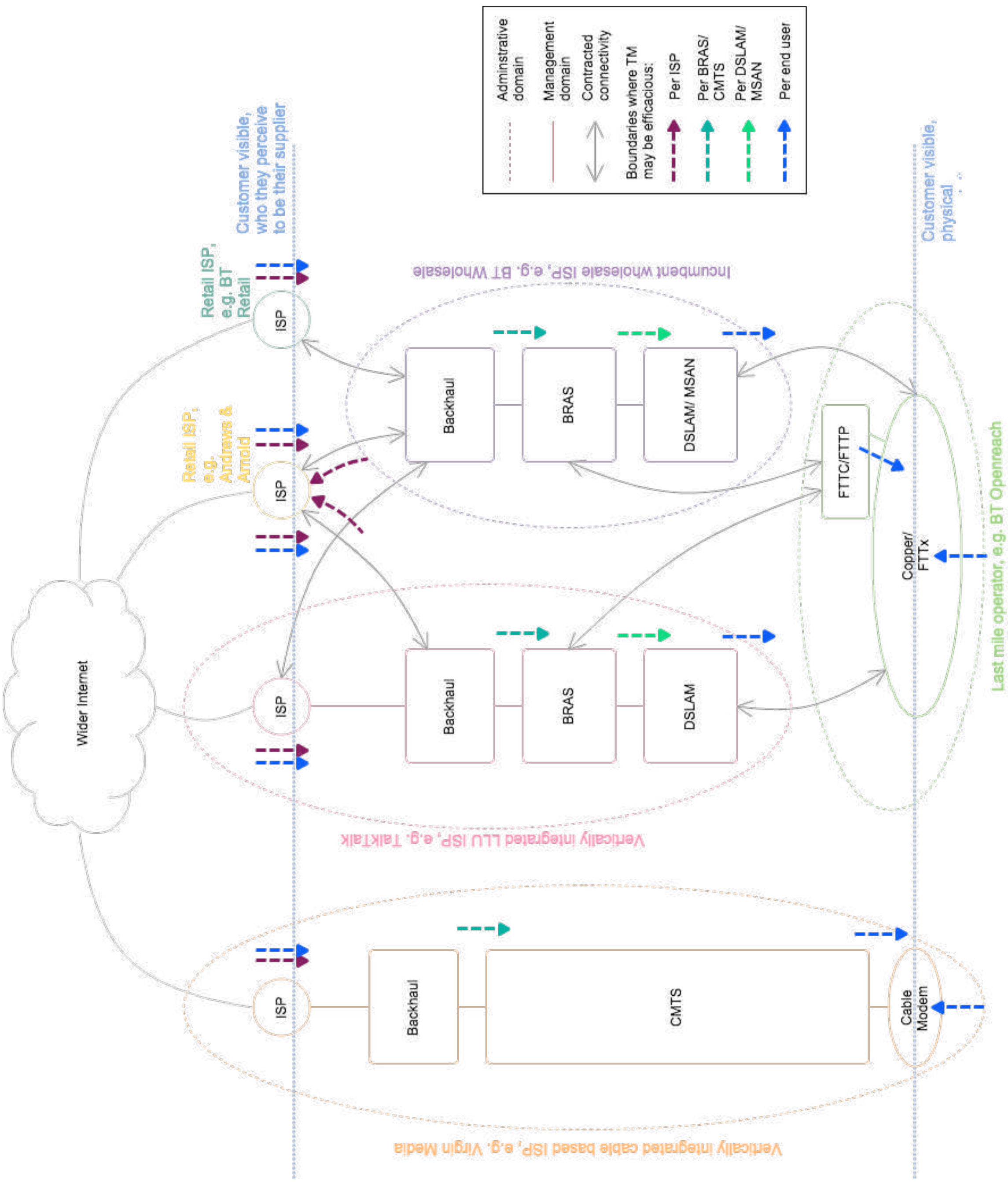


Figure C.5.: Potential TM points in the UK broadband infrastructure (wireline)

D. Analysis of BT SINs

The UK is unusual in that, due to its market structure, several of the interfaces that would be internal in a vertically integrated ISP have become externalised. Consequently, BT is required to publish "Supplier Information Notes" (SINs) that contain much of the technical information needed for other parties to connect to the constituent services offered by BT OpenReach and BT Wholesale. The SINs also make some qualitative (but not quantitative[1]) descriptions relating to traffic management. Thus it is possible to analyse them to extract what statements (if any) they make regarding queuing/scheduling, traffic management, etc.. The currently available SINs have been analysed and the following extracted[2]:

SIN	Title	Version	Comments
472	BT Wholesale Broadband Connect (WBC) Products Service Description	472v2p6	There are interesting issues relating to the inter-path and inter-user effects of Content Connect[3] and TV connect[4]. The potential QoE effects resulting from these may be something bodies like Ofcom would be interested in, particularly as there is no discussion about quality isolation in this SIN.
498	Generic Ethernet Access Fibre to the Cabinet (GEA-FTTC) Service and Interface Description	498v5p1	In §2.1.5.1, BT's interpretation of the priority code point of the ethernet frame is described. As a result of this interpretation, there are effectively only two drop-precedence levels in their system. These levels are used for both intra- and inter-consumer $\Delta Q_{\vert \mathrm{V}}$ management. The GEA product maintenance traffic "has priority" over the end-user traffic and the multicast offering is in the highest available priority level, both of these facts present a performance hazard. There is insufficient information provided within this SIN for it to be possible to know whether any observed ΔQ is within design limits. In addition, those design limits are not made fully clear in this SIN.

[1] BT does make more quantitative information available in commercially confidential "handbooks". A flavour of the additional quantitative information can be found in an edition of the FTTC handbook that has become public `http://goo.gl/d6Y5q1`. Several other handbooks with "in commercial confidence" markings can be found on BT's websites through a simple web search.

[2] Note that all references to sections in this Appendix are to sections of the corresponding BT SIN, not sections of this document.

[3] Content Connect is BT's CDN service.

[4] TV connect is BT's CDN for TV streaming video.

SIN	Title	Version	Comments
385	IP Connect UK Service Description	385v2p8	Although there is reference to IP Precedence marking and DSCP marking there is no description of what methods might be used (this SIN contains less detail in this respect than SIN 498). There is, however, an acknowledgement of the effect of fragment size on delay (particularly for low-speed frame-relay-based access services). The capacity made available as SDU carriage is not articulated, only PDU costs are quoted (and even here there is still ambiguity).
471	BT Wholesale Broadband Managed Connect Shared Service Description	471v3p3	This SIN acknowledges that the bandwidth measurements used are taken at layer 2, and include all protocol overheads. It also states that this is the bandwidth billing metric. Policing occurs at 110% of purchased load. Inaccurate packet marking leads to loss. If the wholesale customer (e.g. an ISP) does not use TM to prevent this policing, there is a service stability hazard.
492	Ethernet Access Direct (EAD) inc. EAD Enable and Ethernet Access Direct Local Access Service & Interface Description	492v1p8	SyncE effects are outlined, thus giving some indication of resulting ΔQ effects. There is a focus on loss notification but no other TM consequences.
495	BT Wholesale Broadband Connect Fibre to the Cabinet Service	495v1p1	PPP is used to support BRAS profile (sync rate) information exchange. This requires significantly lower PPP/L2TP time-outs (<20s), increasing the overall baseline end-to-end cost of this service.
503	Generic Ethernet Access Multicast Service & Interface Description	503v1p2	This SIN asserts that multicast is carried in a separate VLAN (§3.1). Multicast is assigned a relatively high urgency – 'level 3' (§3.1.5). The CP has the responsibility to shape and manage the effects of the multicast traffic delivered on the rest of the traffic (even though its traffic pattern may not be visible) - §3.1.6 & §3.1.6.1. There is no policing (at present) on the FTTP – this may represent an inter-end-user performance coupling hazard. (§3.1.6.2).

SIN	Title	Version	Comments
506	Fibre to the Premises (FTTP) Generic Ethernet Access Service and Interface Description	506v1p2	In §2.1.5.1, the SIN describes the "prioritisation" mapping. This is a single marking that embodies both 'urgency' (the requirement for low latency) and 'cherish' (the requirement for low loss rate). When expressing urgency, 4 is the most urgent and 0 is the least. To achieve this 7, 6, and 5 are remarked to 4. When expressing cherish, there are only 2 levels - 0 or any other value. This two-level cherish marking (§2.1.5.1.1) is used to resolve inter-end-user resource contention. The wholesale customer needs to keep traffic with non- zero marking {7,6,5,4,3,2,1} within a "prioritised rate". This rate is product dependent. The upstream is managed using strict priority queueing (§2.3.6), with 4 urgency classes {{6,7}, {4,5}, {2,3}, {0,1}}. This particular management only occurs within the CPE, the rest of the path has no explicit traffic management (§2.3.5).
509	BT Wholesale Broadband Connect (WBC) Fibre to the Premise (FTTP) Service & Interface Description	509v1p2	This SIN covers purely interface, and other non-performance effecting, issues.
511	BT Wholesale TV Connect (TVC) Service & Interface Description	511v1p5	This SIN is a technical description of the service interface. The only explicit performance information contained in it is that the stream is a MPEG-2 single program transport stream (§5.2) with video bit rates of 2.5Mbit/s, 3.0Mbit/s, 7.5Mbit/s and 10Mbit/s, and audio bit rates of 128kbit/s or 224 kbit/s. In each case these figures represent minima, the real frame cost (i.e contention for the common resource) will be higher. TVC is expected to be carried over the multicast service (§6.3). There is a loose performance guarantee for the BRAS/IGMP query – general queries are made every 125s with a maximum response time of 10s. Specific queries can made at a rate of one every 10s, with a response time of 8s (§6.3).
482	BT IPstream Connect Service Description	482v1p13	This SIN is a service description - it describes how 'speed' can be measured as part of the diagnostic measures (§5.1), and the management domain boundaries (§3.1).

SIN	Title	Version	Comments
485	BT IPstream Connect Office, BT IPstream Connect Home, BT IPstream Connect Max & BT IPstream Connect Max Premium Products Service Description and Interface Specification	485v1p2	This SIN describes services in terms of their available sync rates (§4.1). It is worth noting, however, that all of the products in this specification are now legacy services.

D.1. Caveats relating to bandwidth measures

When interpreting the statements in these SINs (as in the vast majority of technical documentation in this area), it is important to note precisely where traffic management is being used, or where costs are being calculated/accrued on data flows. Measurement/management is based on *the size of the protocol data unit*[5] at a specific point in the end-to end-path, including any protocol overheads. Thus there is no uniform way to compare "bandwidth" at one location (say the ISP ↔Wholesaler boundary) with that at another (say the BRAS ↔ DSLAM boundary). The fungibility of bandwidth in a circuit-switched environment such as TDM is not present in broadband.

As the size of any PDU is likely to change many times along the end-to-end path[6], the actual user data rate delivered by a reported "bandwidth" can vary substantially[7]. One way BT Wholesale addresses this fact is by distinguishing between "SYNC" rate (the rate at which signalling is occurring over the final connection to the premise) and "BRAS" rate (the rate at which traffic is shaped in order to avoid queuing in the DSLAM).

[5] A protocol data unit (PDU) is the composite of the protocol headers and the service data unit (SDU). The data relating to an application is within the SDU, but there may be several additional intermediate protocol layers.

[6] This size change is because a user IP packet is encapsulated/de-encapsulated by various transport technologies as it traverses its path.

[7] The most extreme case the authors have encountered was in a ADSL-based VoIP system where (through a configuration choice) voice packets, by only one octet, occupied two ATM cells rather than one. This meant a 96% overhead, and thus halved the effective capacity of the system.

E. Additional Literature

There is a further body of relevant literature, as represented by the citation graph in figure E.1. This was discovered by: making an initial survey; referencing all papers cited by the initial cohort; and then searching for further papers citing those.

Interesting material includes [33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 32, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 21, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110].

The techniques studied in detail in §2 are marked with square nodes.

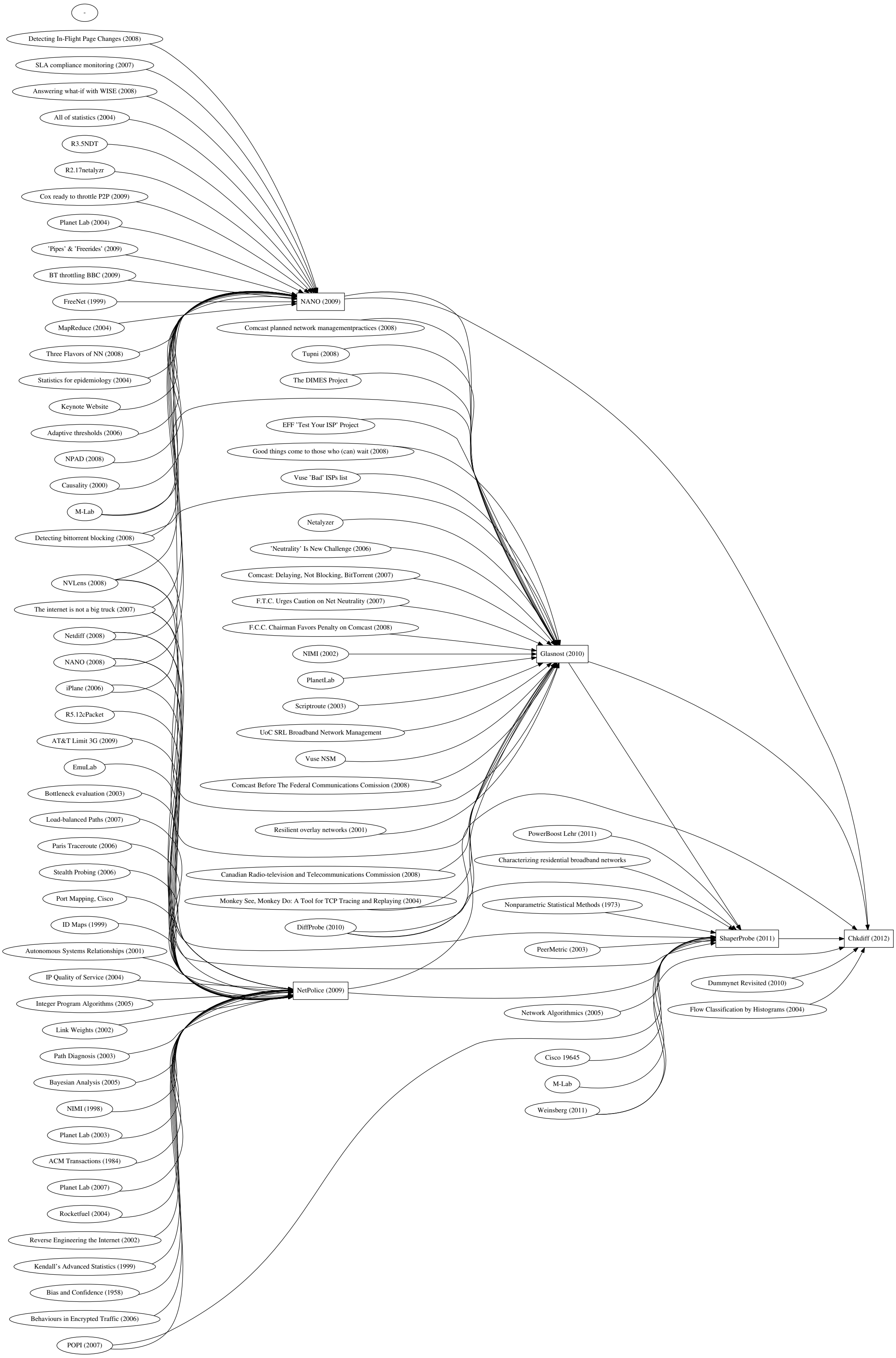


Figure E.1.: Citation relationship between relevant papers

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄六

Actual Experience: Investigation of Internet Quality of Experience

Quality benchmarking and
Digital supply chain management

23rd July 2015

Investigation of Internet Quality of Experience

for

Ofcom

Actual Experience plc
The Tramshed, Beehive Yard, Bath BA1 5BB
t: +44 1225 731 428
e: enquiries@actual-experience.com
w: www.actual-experience.com
Company number: 06838738
VAT number: 971 9696 56

List of Figures

1 Executive summary

Over the past two and a half years, Actual Experience has worked with Ofcom to analyse consumer Internet services across the UK.

This work has been based on the premise that, even though access to fast and superfast broadband has steadily increased, consumers still see evidence of poor and variable quality with the services they use. Ofcom has asked Actual Experience to investigate popular services, such as web browsing, video streaming, voice and other services to provide information around how and why consumers continue to see instances of poor quality.

Previous reports focused on that investigation and Actual Experience's methodology, culminating in information published alongside the 2014 Infrastructure Report. This showed results suggesting that access speed, although necessary, is not sufficient to guarantee a high quality Internet service. In fact, with access above around 8-10Mbps, speed generally ceases to become the dominant factor in determining service quality. Actual Experience's methodology of looking across digital supply chains – the end-to-end set of providers and infrastructures between consumers and content – exposed areas of quality-affecting issues in home environments, in ISPs and in network providers and content providers. This set of data is useful not only to Ofcom, but also to consumers wishing to understand the service they've purchased and to ISPs, infrastructure providers and content providers who work to provide higher quality services to their customers.

This latest report considers a specific aspect of the Actual Experience analysis. Actual Experience measures *digital experience quality* – how the consumer perceives the digital product – Facebook, iPlayer, Skype or any other service. No technical metric, including speed, can link directly to consumers' perception of quality and Actual Experience uses digital Voice of the Customer scores (VoCs) to achieve this. Digital VoCs are derived from a decade of British academic research and are used by customers worldwide to understand the quality of the digital services they deliver – be they within business, government, ISPs or content providers. Ofcom has asked Actual Experience to undertake further work to directly correlate these digital VoCs against surveyed consumer opinion – proving that the analysis produced closely matches consumers' perception of their Internet services.

This report shows the following results:

- Clear correlation between Actual Experience's analysed scores and surveyed consumer opinion, with confidence levels of 98% or better
- That clear correlation exists not just at the level of overall digital experience quality, but also for specific applications and differing consumer perceptions at different times of day
- As in previous reports, there is proof that both surveyed perception and analysed quality improves with superfast broadband packages but that increased access speed is insufficient to guarantee a consistently excellent standard of Internet service quality.

Alongside previous work demonstrating methodology and digital supply chain analysis, this report shows Actual Experience's approach to be a robust and reliable method of not only understanding the reasons for consumers' perception of their Internet service, but also delivering a statistically significant dataset to help any interested party improve the quality of the UK's Internet services.

2 Introduction

2.1.1 Actual Experience's engagement with Ofcom

Actual Experience has been working with Ofcom for several years, with a focus on proving a methodology that can be used to assess digital experience quality for fixed line broadband consumers in the UK. Specifically, previous work has been in response to ITTs 28-2013, and ITT 31-2012, culminating in publication of the response to the latter alongside Ofcom's 2014 Infrastructure Report[1].

This previous work concentrated on proving the Actual Experience methodology, both in terms of the science underpinning the quality scoring (digital Voice of the Customer) and analysis of digital supply chains to show where issues exist that are detrimental to consumers' use of the Internet.

2.1.2 Aims and goals

Against that background, this report builds on that analysis with a substantially larger subscriber base and focuses on correlations between Actual Experience's digital VoC scores and consumer perceptions of Internet services gathered from a pre-deployment survey. The report aims to show the following:

- Demonstration of where correlations exist between customer perceptions of quality and Actual Experience's analysis;
- An understanding of variations in consumer perceptions of quality across different services, packages and times of day;
- Demonstration that the methodology is robust, repeatable and applicable to future analysis of Internet services at scale

[1] http://stakeholders.ofcom.org.uk/binaries/research/technology-research/2014/performance-eval.pdf

2.2 Digital services considered

The digital services chosen for analysis represent both a set of popular services for UK consumers, thus assisting in the best possible survey results, and a range of different sensitivities and behaviours across the digital supply chains analysed. The following set of websites and services have both been analysed by Actual Experience and used within the consumer survey:

- Web browsing
  - Google
  - The BBC
  - Amazon
  - Gov.uk

- Streaming Video
  - BBC iPlayer
  - YouTube
  - Netflix

- Consumer voice-over-IP, such as Skype, Hangouts, etc.

2.3 Access types considered

2.3.1 Technology

No restrictions have been placed on subscribers in terms of their access technology and the following have been seen in accordance with the requirements of the ITT:

ADSL variants
VDSL (FTTC)
FTTP/H
Cable
Other (e.g. legacy dial-up, satellite)

When considering the broadband package type, these fall into two main categories – legacy ADSL up to and including ADSL2+ and superfast services – FTTx/VDSL and cable, e.g. HFC[2].

2.3.2 Package speeds

Users registering for the service were required to state their broadband speed. Options offered were <2Mbps, 2-5Mbps, 5-10Mbps, 10-40Mbps and >40Mbps.

However, when actual line speeds have been analysed, the latter two bandings have been changed to 10-30Mbps and >30Mbps in-line with Ofcom's current definition of 'superfast' broadband (>30Mbps)

---

[2] Hybrid Fibre-Coax, an access infrastructure technology used by cable TV/broadband providers

2.4 Subscriber base

Geographic distributions are shown in Figure 5 below.

Package types are shown in accordance with the definitions in §2.3.1, Technology above. There are 1,483 subscribers for whom there was declared data from the point of registration.

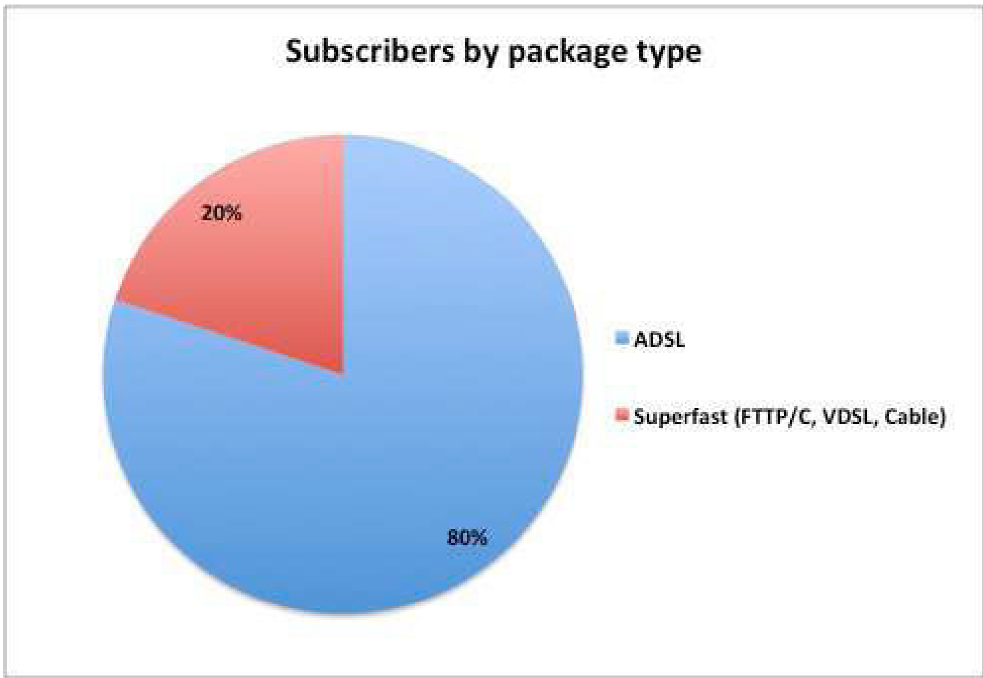


Figure 1 – Subscribers by package type

In their 2014 Infrastructure Report, Ofcom found that 71% of homes had ‘standard’ broadband connections as defined above, whereas 29% had cable or ‘superfast’ packages. This suggests that our survey results are representative of UK consumer broadband with respect to package type, but somewhat skewed towards lower speed packages. This is consistent with work that has been done in recent months with a number of rural broadband provider groups, aiming at providing good data to justify infrastructure upgrades for consumers with currently very low speed and legacy services.

2.4.1.1 ISP

Here we see the breakdown of subscribers by ISP. There are 14 providers for whom there is good data (the ISP can be positively identified) at the time of writing and the most common of these are shown below.

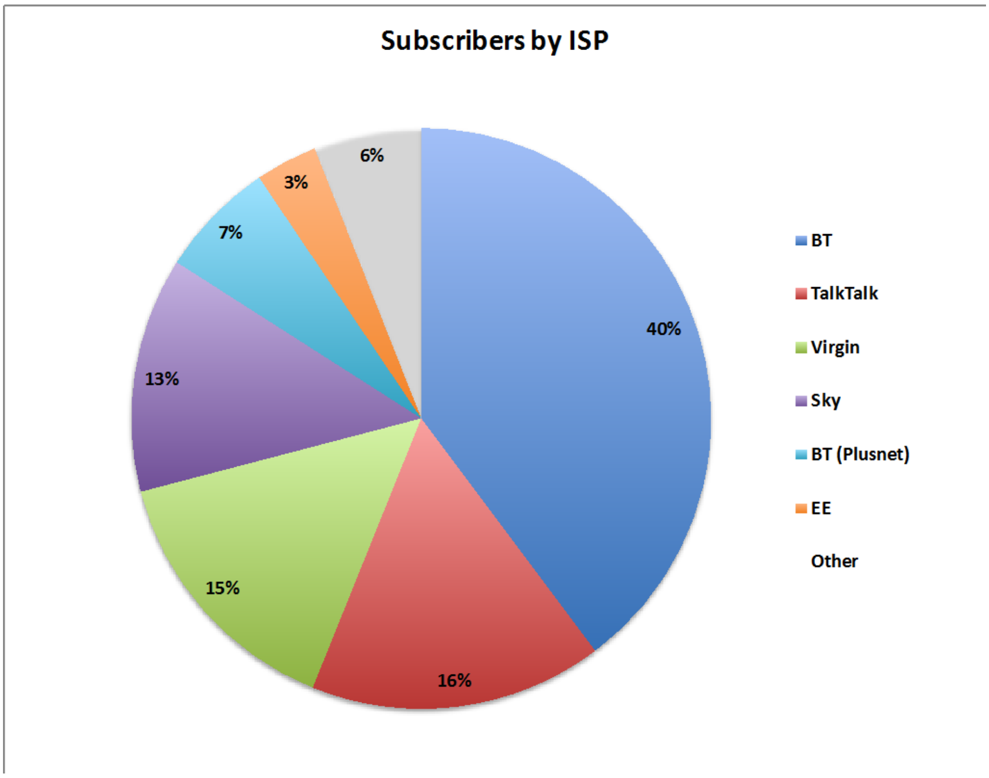


Figure 2 – Subscribers by ISP

3 Internet Digital Supply Chains

The Internet is a global mesh of interconnected providers of network, content and infrastructure services, all of whom must work together to enable consumers and businesses to work, play, transact, interact and communicate on-line.

When a user consumes a piece of content, be that a corporate service or, for example, BBC iPlayer at home, they must connect through this complex infrastructure to the location where the content is held. The set of multiple providers and infrastructures between them and their content can be thought of as a digital supply chain.

Visibility and understanding of how providers in these digital supply chains operate together is critical to understanding consumers' digital experience quality – in the same way as physical supply chains demand that factories or logistical elements must work together to deliver high-quality products to consumers.

A consumer's perception of their service comes from the many interactions between these many parts of the digital supply chain – not simply how well or poorly one bit is performing. Seemingly benign variations in different places may combine to push Internet service quality over an 'experience cliff' – the point at which the consumer notices a degradation in service quality. One practical impact of this complexity is that, whereas behaviours in one supply chain can push quality over that cliff, the same behaviours may not do so in another supply chain. This creates the variability in the quality of the service that the consumers receive and illustrates why an 'outside-in' analysis – a view explicitly from the users' perspective of their service – is essential. To achieve that, the analysis must extend across the whole of the supply chain between the consumer themselves, right through to the relevant content.

Actual Experience delivers such a digital supply chain analysis to organisations around the world – providing an analogous level of visibility to that which has existed across physical supply chains for many decades. In the context of our work with Ofcom, this involves the use of digital Voice of the Customer analysis and enumeration of the digital supply chains that exist between consumers involved in the project and the content of interest to Ofcom.

These digital supply chains contain many elements belonging to the multiple providers who either serve or carry content to consumers. In this report, as in those delivered to Ofcom previously, these have been simplified and grouped together into the four broad areas shown overleaf.

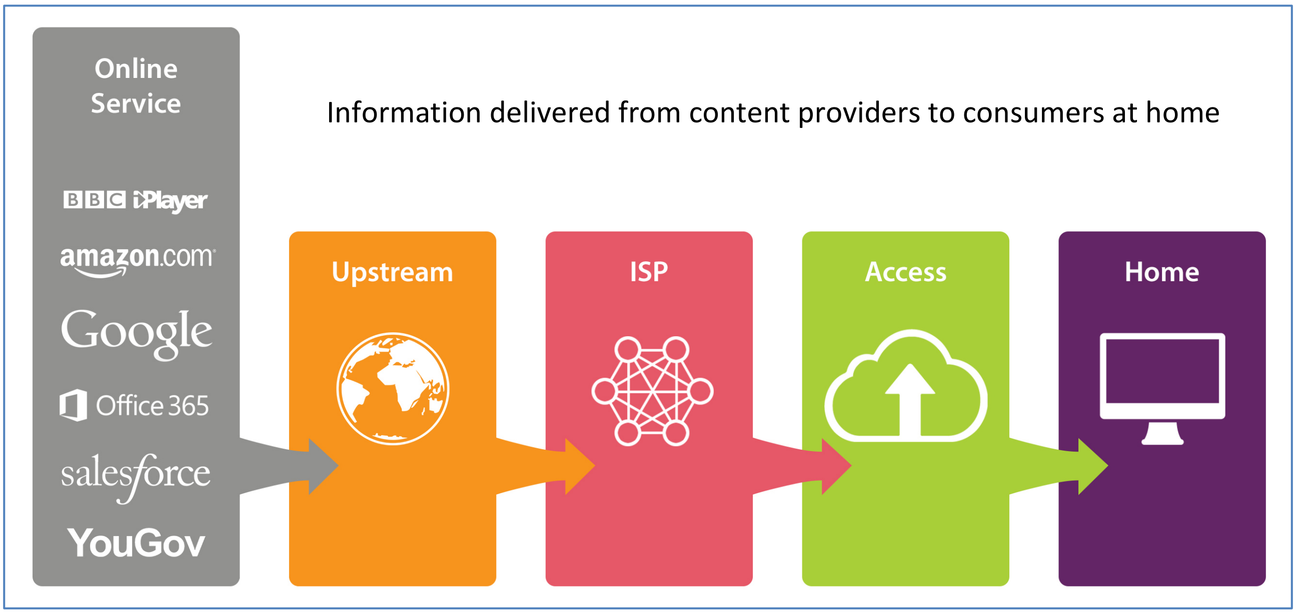


Figure 3 – Delivery of content across Internet digital supply chains

Home: The consumer's home environment, consisting of the router or device supplied by their broadband provider, along with Wi-Fi or other mechanisms within the home, through which the consumer connects to the Internet.

Access[3]: The broadband provider's local infrastructure – physically connecting the consumer to the provider's network, often along with many other consumers in that geographic area.

ISP: The regional and national networks of the broadband provider, upstream of the local infrastructure considered in the *Access* segment.

Upstream: Other network providers and content providers that form the part of the digital supply chain between the broadband provider and the content itself.

[3] Technical note: The access regime lies between the WAN-facing port of the consumer's home router and the first IP (layer 3) element in the ISP's network, and incorporates all intermediate transmission elements in between.

4 The Measurement Methodology

Additional detail on Actual Experience's methodology is available in the report published alongside Ofcom's 2014 Infrastructure Report[4].

4.1 Background from previous work

Actual Experience is the commercialisation of 10 years of academic research, led by Professor Jonathan Pitts, at Queen Mary University of London (QMUL). This research, and the advanced analytics produced, are based on the contention that when considering the performance of the providers across digital supply chains, it is critical to measure and understand how consumers perceive the effects of those providers' behaviour on the quality of their services. In the corporate context, this relates directly to the ability of users to be productive and is thus of profound economic value – something that is equally true when considering the social and economic productiveness of 'Digital Britain'.

To understand the quality of services delivered, it is critical that analysis is conducted continuously – snapshot measurements cannot be considered a proxy for a consumer's experience of a product over any period of time. Within this context, there are three components to an Actual Experience deployment: Digital Users (DUs), the Analytics Cloud and the Web UI.

- A DU is a small piece of software that resides on a device in the same location as the users whose digital quality is being analysed. That device can be a user or standalone PC (Mac, Windows or Linux) or, for example, a Raspberry Pi. The DU takes continuous measurements that are then analysed to produce the digital Voice of the Customer (dVoC) score and understand the performance of the digital supply chain.
- The Analytics Cloud is the heart of Actual Experience – where all analysis happens – and consists of high-powered computational clusters located in secure datacentres. The Analytics Cloud processes the results of DU measurements through sophisticated algorithms to produce two things: firstly, the dVoC scores and secondly the correlations required to benchmark digital supply chains – identifying sources of underperformance that affect the quality of the digital products delivered.
- The Digital Supply Chain Director consists of a set of dashboards and is how results are presented to customers. Access is via any modern web browser, and customers are presented with three core pieces of information:
    - A benchmark of how good digital product and service quality should be
      (i.e. if everything worked well across that digital supply chain)
    - A continuous digital Voice of the Customer score
      (near real-time digital experience quality, analysed for every active Digital User)
    - Continuous analysis of the digital supply chain
      (over time, to find the things that affect the digital Voice of the Customer and hence, for enterprises and providers, allow effort to be focussed to greatest effect on improving quality and consistency)

[4] http://stakeholders.ofcom.org.uk/binaries/research/technology-research/2014/performance-eval.pdf

4.2 The BbFix® Project

BbFix is Actual Experience's consumer recruitment programme – offering a version of our analysis and results, free of charge, to consumers around the world. The subset of UK subscribers for this report was drawn from this crowd-sourced user-base.

In the BbFix version of the UI shown below, the three core pieces of information are summarised for each application as:

- The digital Voice of the Customer – current, monthly average and the best level of quality that the supply chain can support
- A quality clock – a 24-hour clock – showing the continuous quantification of actual digital quality over the past 30 days
- Where it's going wrong – a visual indication of weaknesses in the digital supply chain (divided into Home, Broadband Provider, and The Internet) that affect digital quality

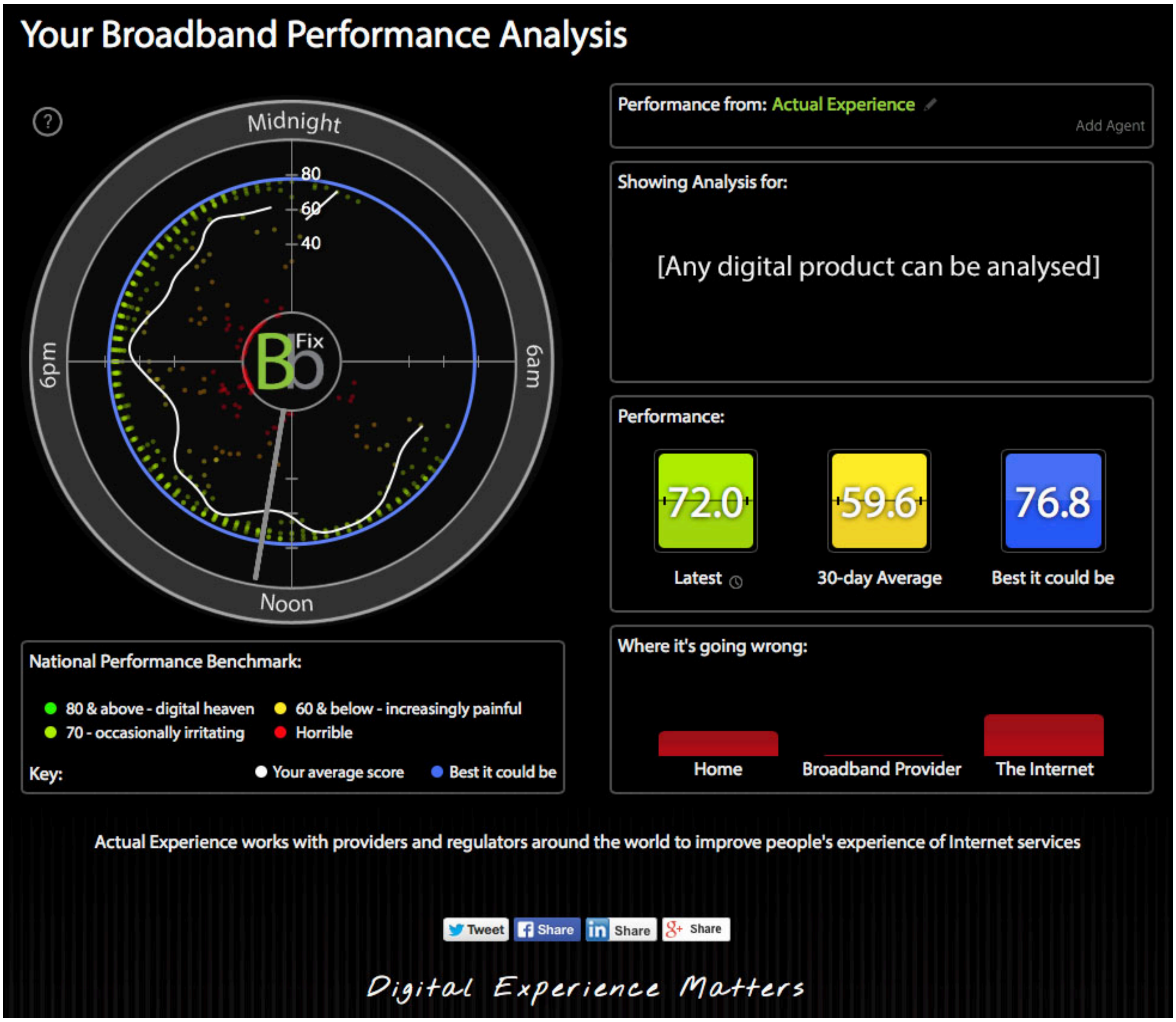


Figure 4 - The dashboard accessed by subscribers

5 Results based on the survey and participant analysis

The survey was designed in collaboration with Ofcom and from a base of approx. 1,800 BbFix subscribers, 1,344 were available for the survey mailing (after taking into account opt-outs, unsubscribes, etc.). From that group, there were 319 respondents, equalling ~24%.

5.1 Survey validity as a representative sample of UK broadband users

5.1.1 Geography

The two graphics below show firstly the distribution of the BbFix base, from which the survey was drawn, and secondly the distribution of survey respondents.

It can be seen that the survey respondents are broadly representative of the base from which they were drawn and show a reasonable spread of users across the UK.

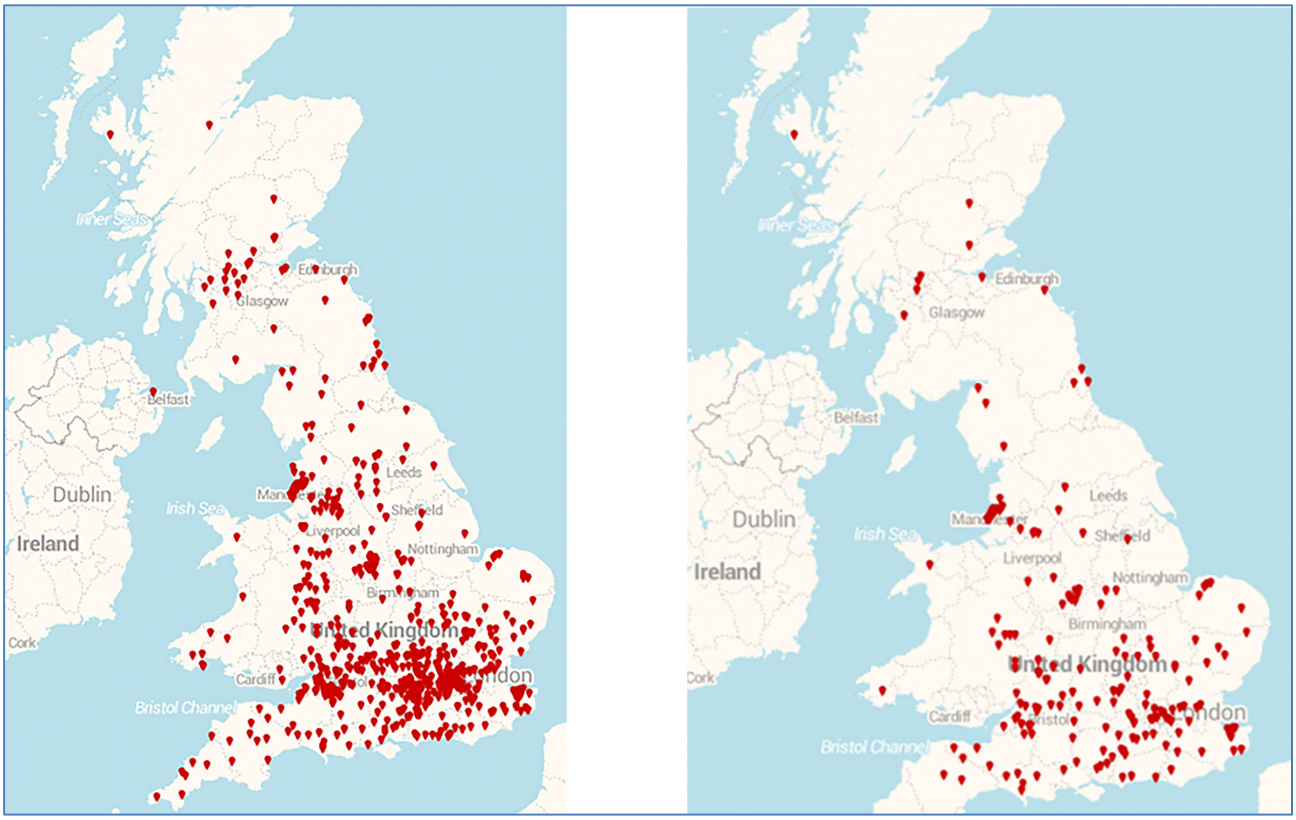


Figure 5 – The BbFix user base (left) and respondents to the survey (right)

5.1.2 Broadband package type

For the 319 respondents to the survey, we saw a breakdown of package types as follows. Standard broadband is taken to mean 'traditional' ADSL (i.e. ADSL, ADSL2 and ADSL2+) with package speeds of up to 24Mbps. Superfast broadband considers FTTH/P, FTTC/VSDL and similar services with package speeds greater than 30Mbps. Cable relates primarily to Virgin Media's superfast broadband product set.

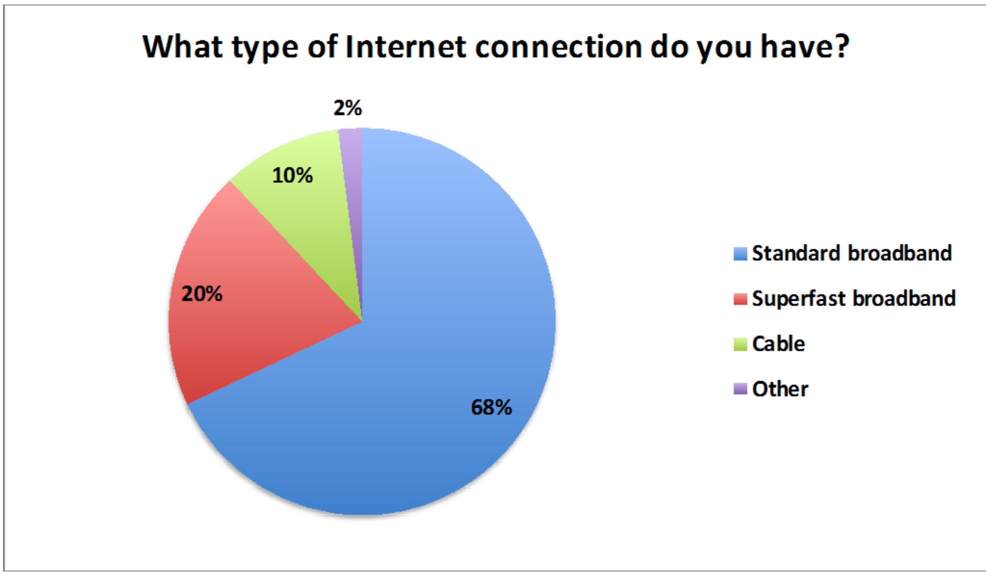


Figure 6 – Broadband package type

In their 2014 Infrastructure Report, Ofcom found that 71% of homes had 'standard' broadband connections as defined above, whereas 29% had cable or 'superfast' packages. This suggests that our survey results are representative of UK consumer broadband with respect to package type, although the proportion of the entire BbFix base with superfast broadband is somewhat lower than represented here (see Figure 1, above).

5.2 Application usage

There are approximately 24.6M households[5] in the UK and 76% of those have a broadband Internet connection[6] – some 18.8M households.

5.2.1 Streaming video

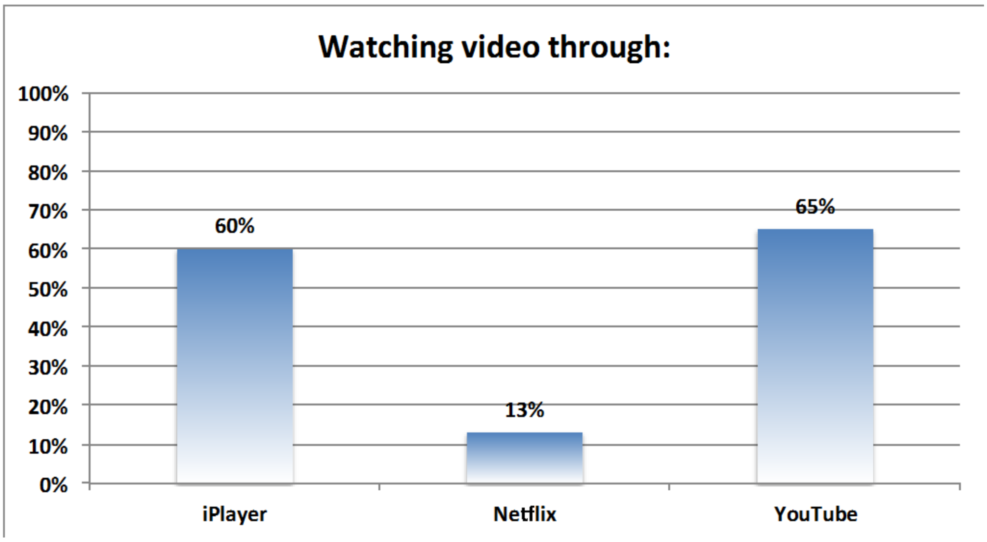


Figure 7 – Streaming video usage

Netflix is estimated at 2.8M subscribers as of early 2014[7]. Assuming one subscription per household, that equates to just under 15% penetration of the UK residential broadband market. In our survey, we saw approx. 13% of users reporting their use of Netflix, consistent with that reporting.

The BBC does not disclose its statistics for the number of households consuming VoD via iPlayer, but Ofcom's 2014 Communications Market Report[8] found that iPlayer users outnumbered Netflix users by 5x in 2013 and just under 3x in 2014. The results of our survey show just over 4.5x more iPlayer users than Netflix.

YouTube was the most popular of the video services amongst respondents and is currently the UK's third most visited website (after Google and Facebook)[9].

[5] ONS, 2013
[6] 91% of the 84% of households with Internet access use fixed broadband, ONS, 2014
[7] BARB Establishment Survey, 2014
[8] http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr14/2014_UK_CMR.pdf
[9] www.alexa.com, March 2015 analytics

5.2.2 Web browsing

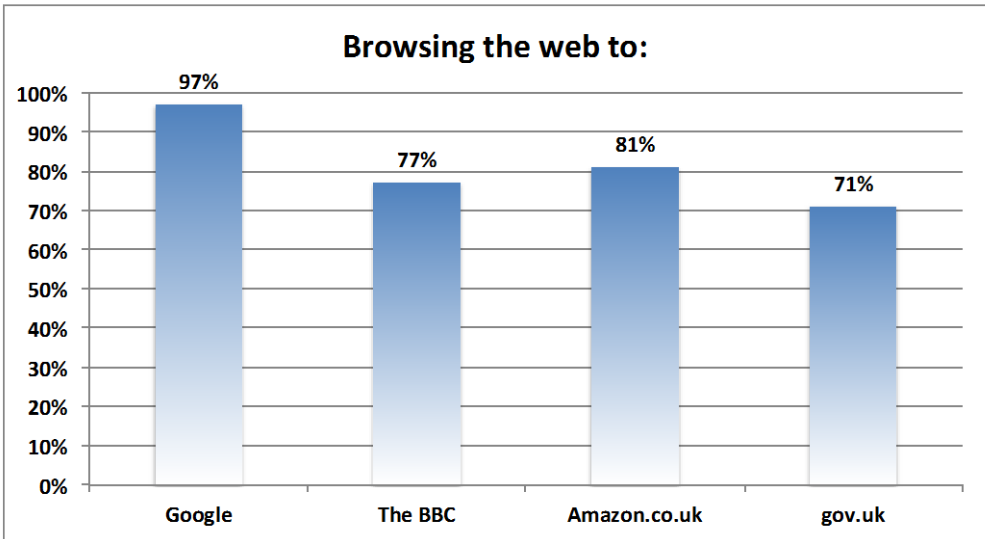


Figure 8 – Browsing usage

In 2013, the BBC reported that Google dominated the search engine market in the UK, with 90% of desktop searches[10]. 97% of our survey respondents used Google in some form.

Alexa[11] reports that the BBC websites receive approx. 32M unique monthly visitors, of which approx. 47% are from the UK. A complicating factor is that there is no breakdown of fixed broadband access or access via mobile devices, but our results of 77% of households, i.e. approx. 14.5M households looks to be credibly representative against this statistic.

In 2013, the Telegraph reported that Amazon was the second most popular retail site in the UK (behind eBay), with 12% of all online retail site visits. Alexa[12] reports almost 40M unique monthly visits with approx. 64% from the UK. Our survey results look to be consistent with these observations, but perhaps suggest that there is more mobile access to Amazon than to the BBC websites.

Ofcom reports that 40% of adults find information about government services online and 28% "complete government processes online"[13]. Any overlap between those two groups is not clear in the report. Amongst our respondents 71% had used the gov.uk website in some form.

[10] http://www.bbc.co.uk/news/technology-23318889
[11] www.alexa.com, March 2015 analytics
[12] www.alexa.com, March 2015 analytics
[13] Ofcom, 2014

5.2.3 Voice

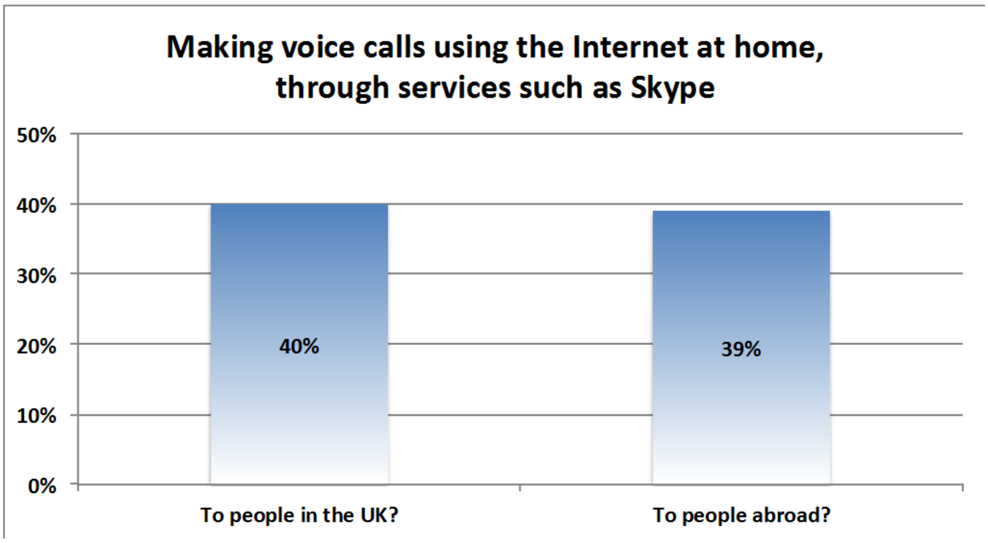


Figure 9 – Voice usage

In 2013, the ONS found that 32% of adults with Internet access used Internet voice services such as Skype, an increase of almost 47% from the previous year[14]. In our survey, almost 55% of respondents used Internet voice with a usage split as shown in Figure 9. Our survey results would seem consistent with a continued but somewhat lower growth rate over the last three years (the ONS data relates to a 2012 survey). Of that 55%, consumers were almost evenly split between those making UK calls and those making international calls.

[14] ONS, 2013. Data for VoIP services not published in the 2014 report

5.3 Perceived quality of connection

5.3.1 General

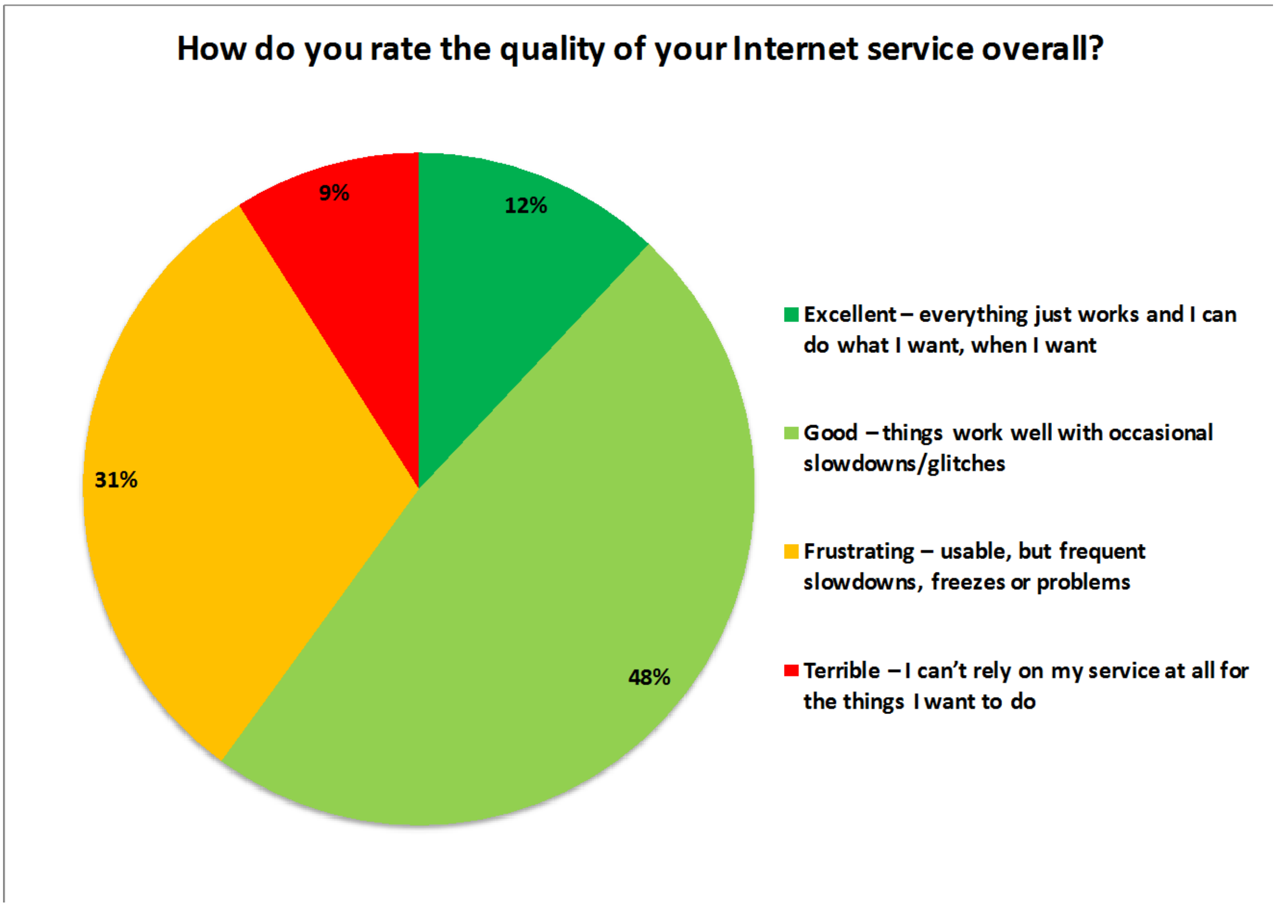


Figure 10 – Overall perceptions of Internet service quality

In answer to the general perception of their Internet services, we can see that 60% of respondents feel that their services are good or excellent, based on the definitions in the key above.

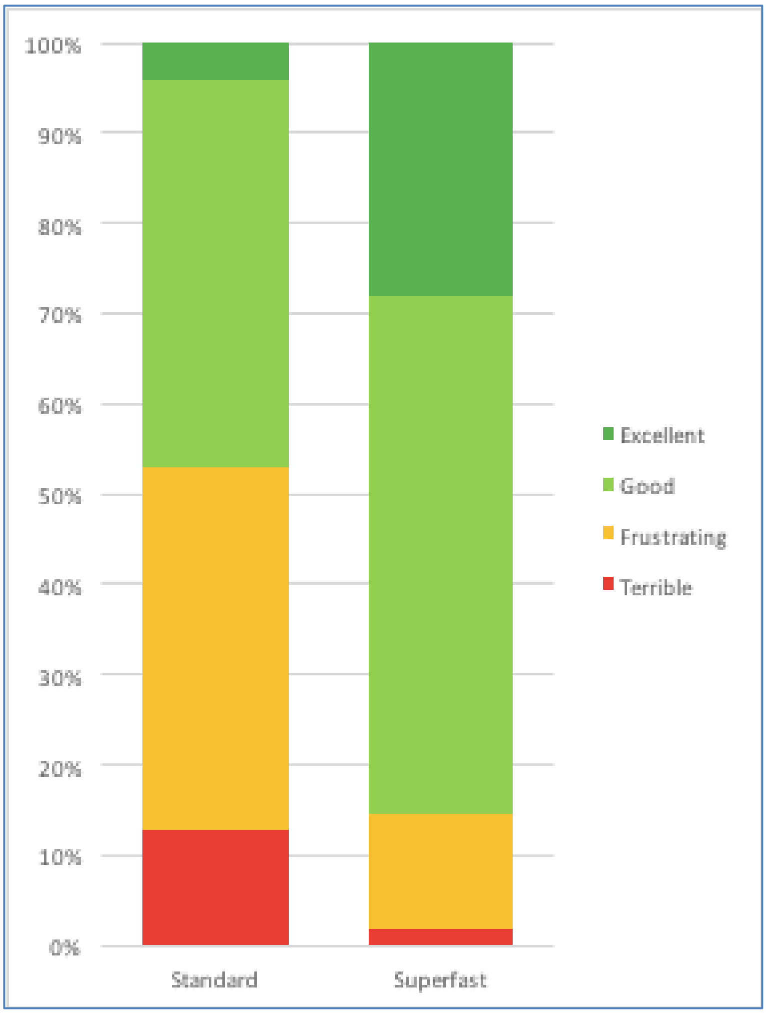


Figure 11 – Percentage of opinions across package types

Figure 11 shows an intuitive result – that as the package speed increases, so we see an improvement in the perception of the service. This does not mean, however, that a high-speed package alone delivers a good opinion. 15% of superfast customers still rate their service as *frustrating* or *terrible*.

This underlines the conjecture that speed is a necessary, but not sufficient element of a high-quality service – only 28% of superfast customers rate their overall service as *excellent*, dropping to 3% of those with standard broadband packages.

The following sections consider more granular scenarios based on specific consumer activities and different times of day.

5.3.2 Application specific

Here we consider general perceptions of quality, but broken down by application.

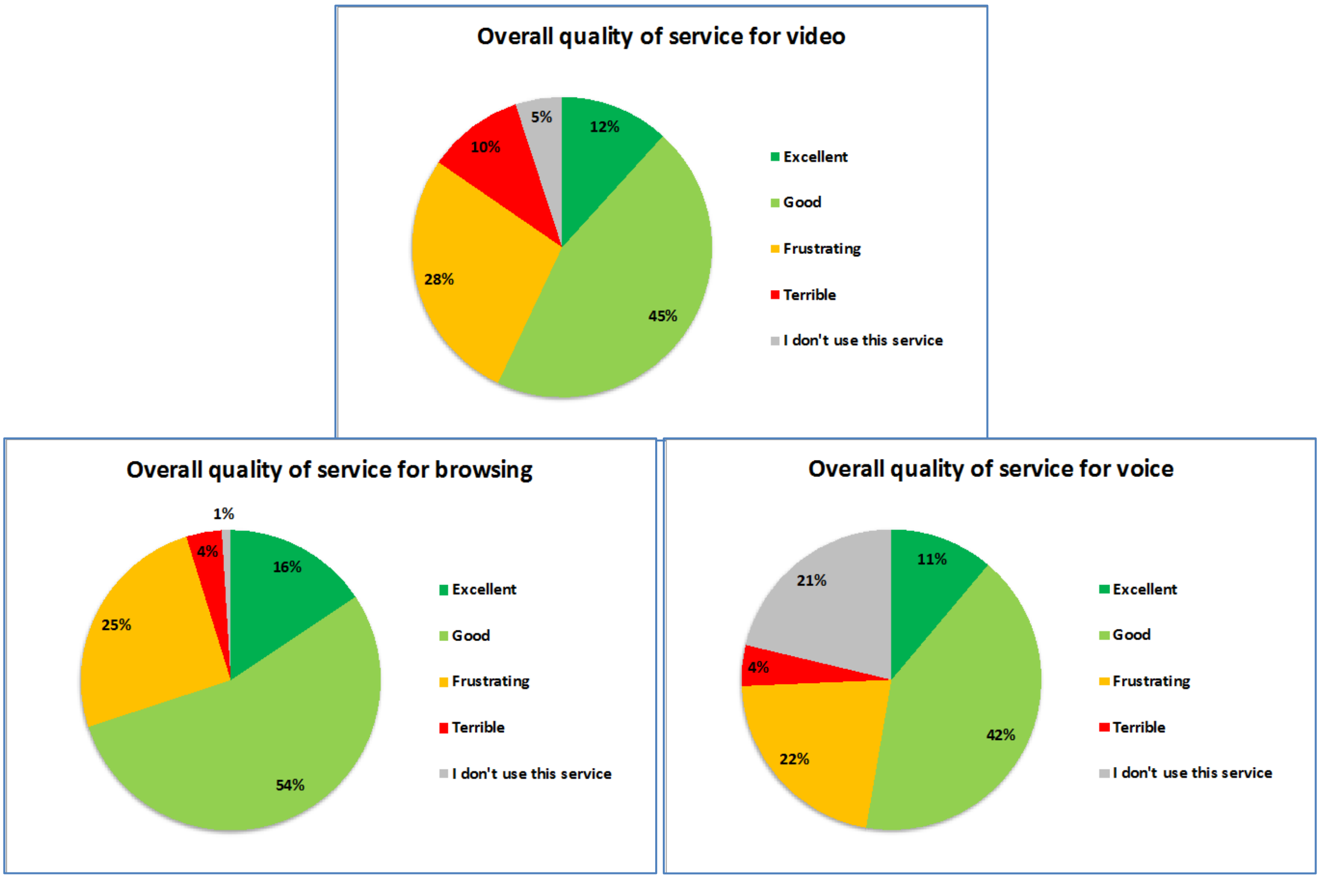


Figure 12 – Perceptions of quality by application

Broken down by application, and considering only the proportion of respondents who use that service, we see that:

- Browsing is the most common activity (99% of respondents using one or more of the websites listed). Some 95% of respondents view video through one of the services we asked about, whilst 79% of respondents used VoIP services for either national or international calling
- The greatest level of satisfaction is with browsing (71% good or better of those who used the service), followed by voice (67% good or better) and finally video (60% good or better)
- Further, significantly more people rated their video experience 'terrible' than for voice or browsing (11% vs. 4% for browsing and 5% for voice)

5.3.3 *Time of day variations*

This section looks at the results above at a greater level of granularity – to understand what, if any, variations occur at different times of day.

5.3.3.1 Video

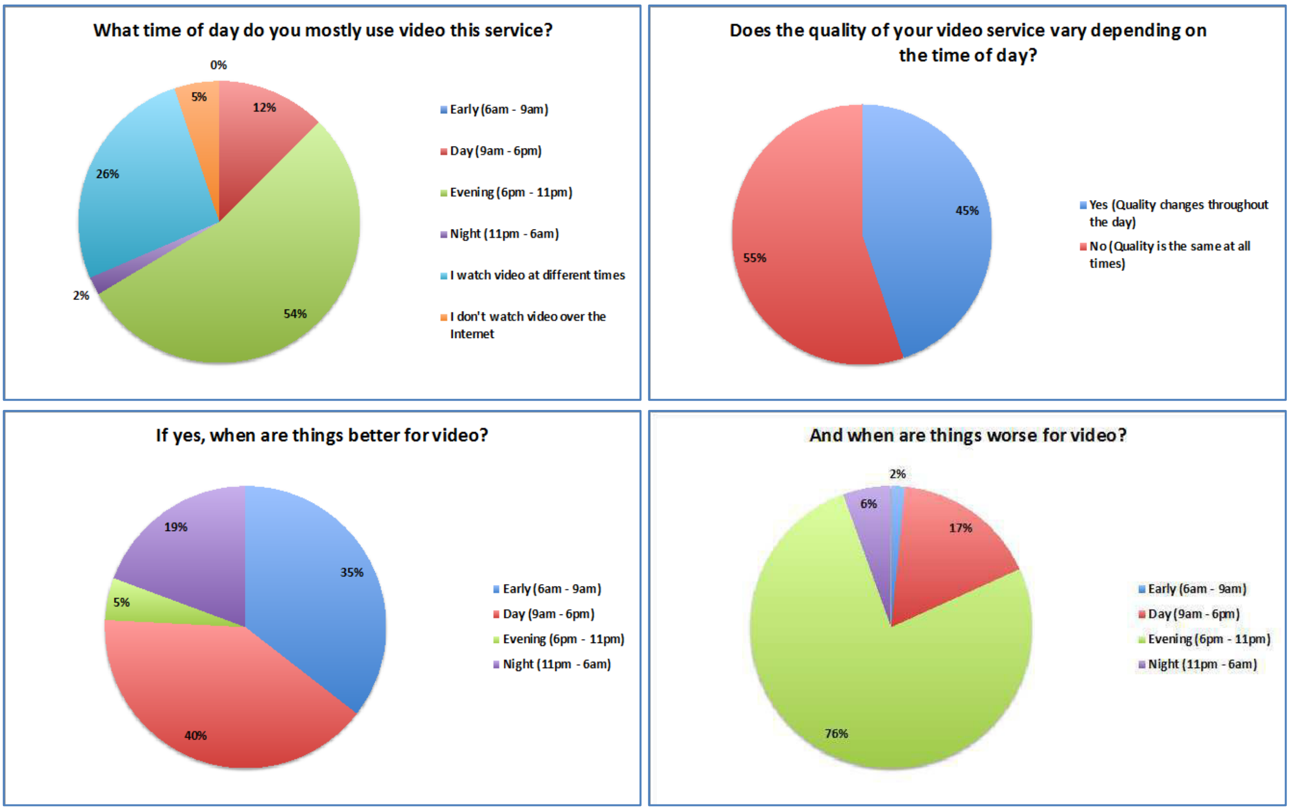


Figure 13 – Perceived daily quality variations for video

As might be expected for consumer video services, we see by far the largest group use the services in the evening. This is consistent with the BBC's analysis of its iPlayer service, which suggests that peak time viewing is a little after 10pm, having risen throughout late afternoon and evening. Viewing requests drop off sharply after that time[15].

However, against that background, almost half of respondents felt that quality was variable, with the vast majority experiencing the worst service when they would be most interested in using it – throughout the evening peak time. Conversely, the survey base was equally split in terms of seeing better quality during morning and day times, with a somewhat smaller segment considering late night to offer the highest quality viewing experience.

[15] BBC iStats, October 2014, slide 16

5.3.3.2 Web browsing

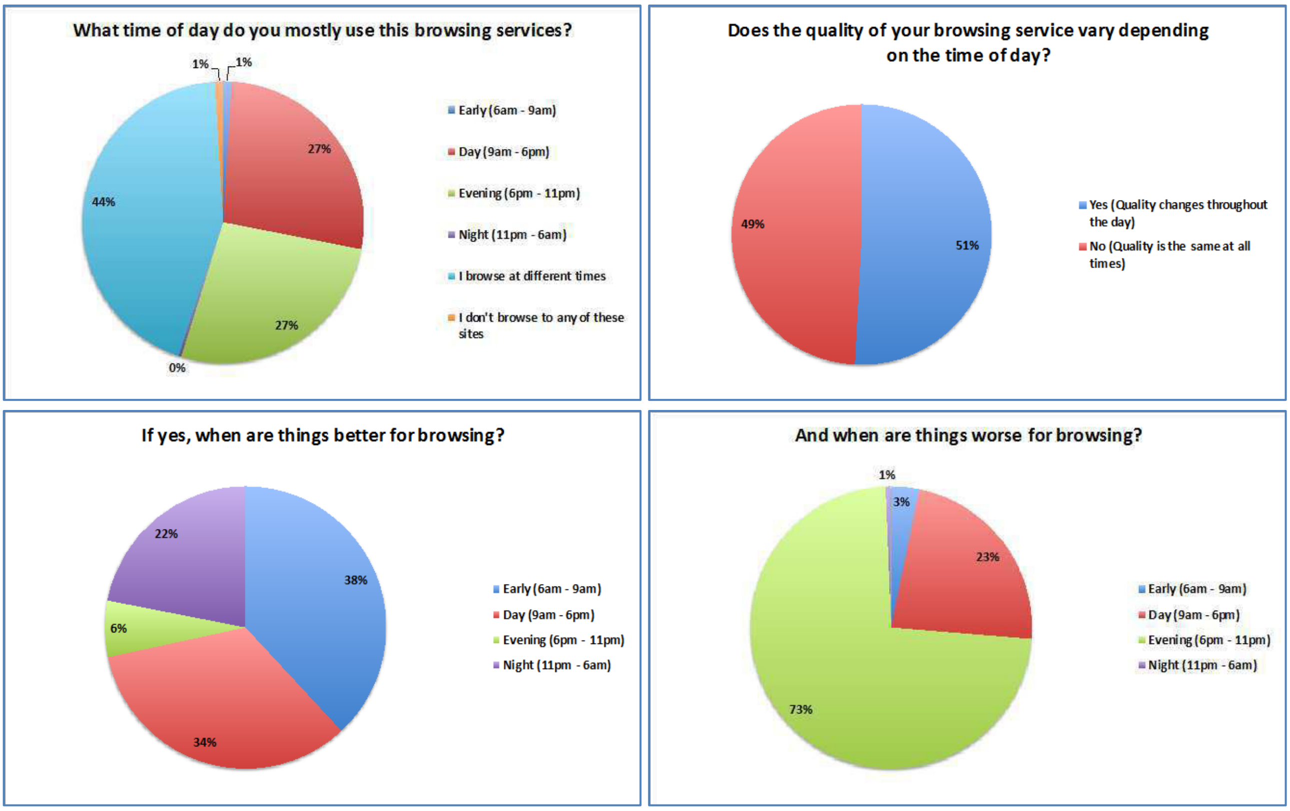


Figure 14 – Perceived daily quality variations for web browsing

Usage patterns for web browsing are more spread throughout the day than for video services, with the largest group browsing across the various time periods queried in the survey, albeit with almost none of the respondents mostly using the web sites listed in the early morning or late night.

With respect to quality, we see an almost identical pattern to video, both in terms of when services are better and when they're worse. This clearly implies that the behaviours that cause our subscribers to perceive a drop in quality are not application specific, but look to be common across their usage. This is a highly relevant guide for subsequent analysis to understand where and why such issues occur.

5.3.3.3 Voice

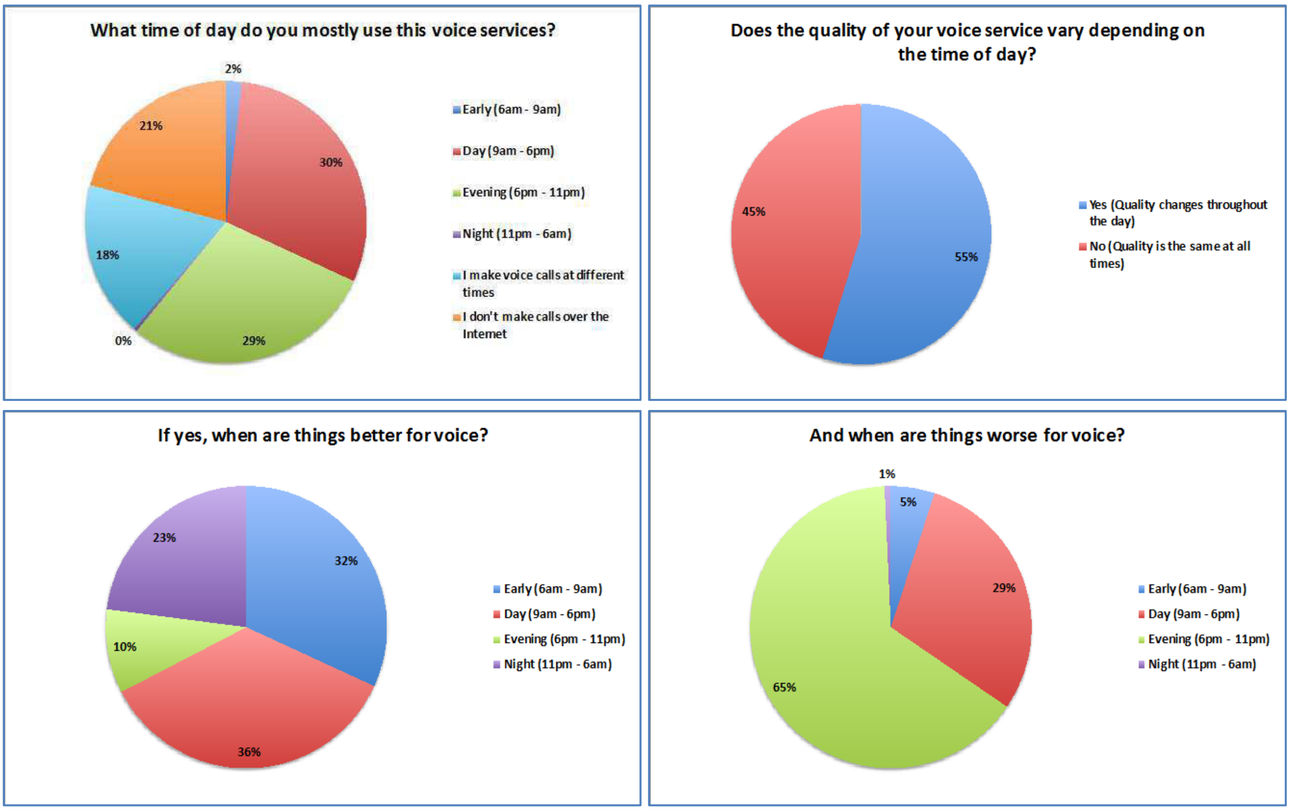


Figure 15 – Perceived daily quality variations for voice

For voice services, we see similar patterns emerge. Usage is almost entirely during the daytime and evening and a little more than 50% see quality variations. These variations look to be very similar to those previously seen, with a significant majority seeing poorer quality of calls in the evening. This further corroborates previous comments that issues affecting service quality appear to be general, rather than confined to one particular service.

5.4 Analysis of survey results

From the results above, we can see that general perceptions of Internet service quality are often good, but that there are significant variations at different times of day, with rather less marked variations between the services used by respondents.

The following sections will consider these results in more detail; looking for correlations with Actual Experience's quality analysis and variations between types of users, based on criteria such as broadband package type and usage at different times of day.

5.5 Correlation of survey results with Actual Experience analysis

5.5.1 Methodology

The survey used to gather user perceptions was agreed with Ofcom and answers were received from 319 of the 1,344 subscribers invited (see §5 above). Apart from the digital services and the other questions chosen, an important consideration was to avoid any potential bias based on previous exposure to the Actual Experience BbFix Project. Although all survey respondents were active users of BbFix, none had seen analysis of the digital services questioned in the survey prior to responding. Only after all responses were received did digital experience quality analysis of these services commence.

Users' analysis was matched to the digital services they indicated they used in the survey and only this data has been used to produce the results below. E.g. if a subscriber indicated that they used iPlayer, voice services, Amazon and Google, then those results, but no others from their analysis contribute to the dataset used for this section.

This section is concerned with correlating users' perception of service quality with Actual Experience's analysed results. To provide a good understanding of digital experience quality, two distinct dimensions must be considered:

- Firstly, an assessment of the typical quality as it is experienced at the point of consumption (e.g. when actually watching a video or browsing a web-site).
- Secondly, an assessment of how the accumulation of specific events affects perception of quality over time. This is a measure of how instances of poor quality affect the overall perception of a service, depending on the frequency of their occurrence.

Actual Experience's analysis captures both the digital Voice of the Customer scores over time – allowing a general correlation to be made – and also specific behaviours in the digital supply chains that result in moments of poor quality. It is the accumulation of these that relate to the second point above.

The requirement for these two types of analysis can be seen from the examples below. Here, we're looking at two users' 'Quality Histories' – a plot of how their digital VoC varies over a period of time.

Figure 16 – Actual Experience screenshots showing digital VoC history

In the first case, the BbFix subscriber rates their overall Internet service quality as *Terrible*, and in the second case the rating is *Excellent*.

Each dot on the plot is the instantaneous dVoC score for that point in time and the figure in the top right hand corner of the screenshot is the mean dVoC score. It's quite obvious from the *Terrible* service that the many instances of very low dVoC are the things that combine to give the low score, in spite of a period of good behaviour at the beginning of the period shown. In the *Excellent* example, such behaviour is entirely absent. Thus capturing the frequency of these events, as well as their overall effect – the mean digital VoC – is a critical metric, both for correlating with user opinion and understanding the cause.

Another way to express this is to show how the instantaneous digital VoC scores are distributed. The following shows the proportion of time spent in each of the five categories.

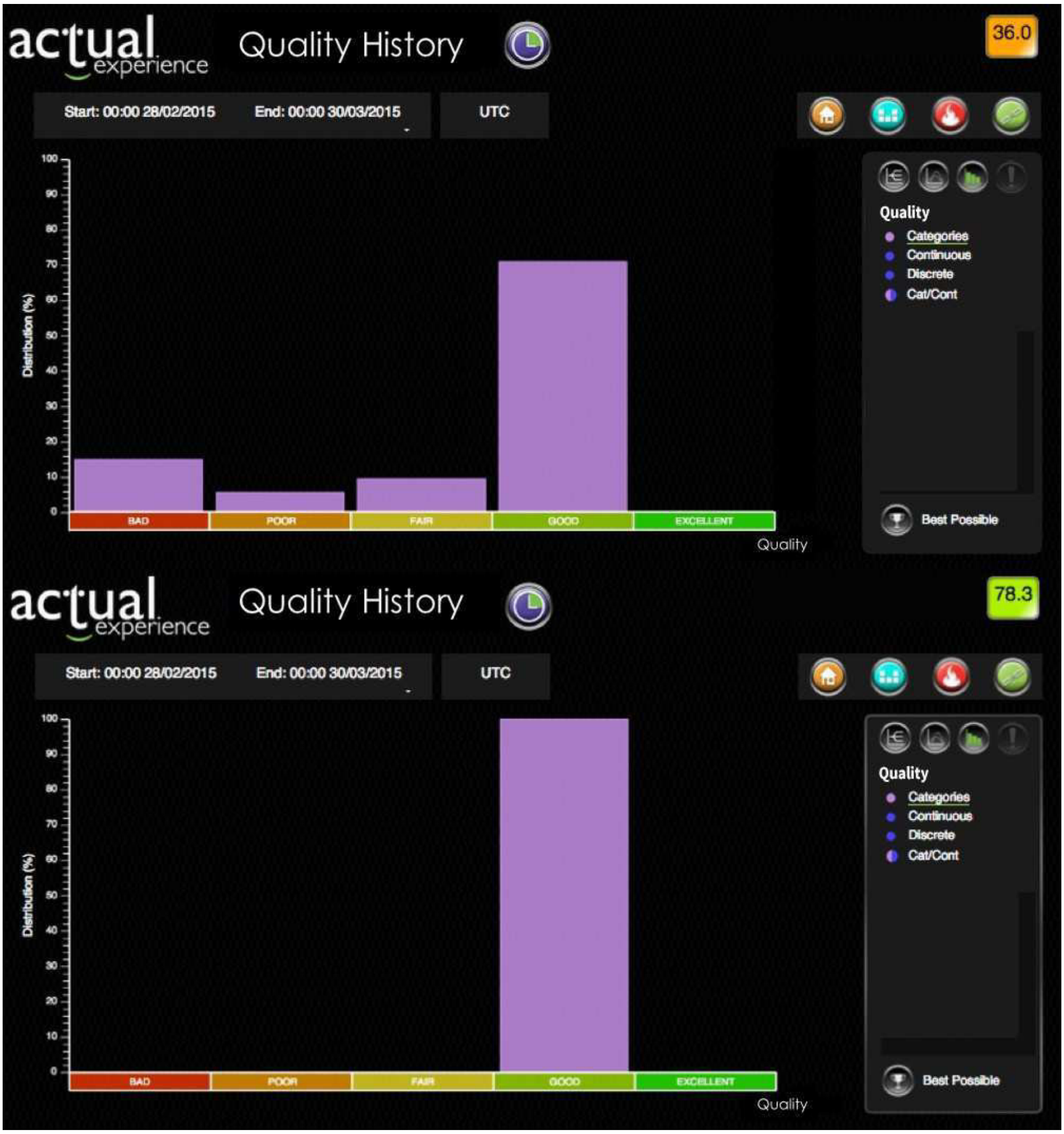


Figure 17 – Actual Experience screenshots showing digital VoC category distributions

With these views, the difference between *Terrible* and *Excellent* is clear: it is the lack of consistency in the delivered digital quality, as evident in the significant proportions of time in the *Fair*, *Poor*, and *Bad* categories of the digital VoC scores. Previous work by Actual Experience[16] has shown that those experiencing poor digital experience quality over low-speed broadband are often subject to significant variations in digital supply chain performance.

---

[16] http://stakeholders.ofcom.org.uk/binaries/research/technology-research/2014/performance-eval.pdf

In the following analysis, the percentage of time where digital VoC scores equate to poor quality is analysed against the surveyed opinion. This shows how the digital supply chain behaviour correlates with the overall quality perception.

The accumulation over time of poor quality events can be evaluated in a variety of ways, e.g. as a mean value, or standard deviation, or as a measure of the distribution. In the physical world, this is analogous to the ability of, for example, a manufacturer to deliver a defect-free product to its customers. This emphasises the consistency of the product across many instances and in the digital world, assessing the consistency of *digital* products requires language explaining the frequency of these behaviours in the digital supply chains.

Here, the concept of 'poor quality time' is used to express that concept of digital consistency in an intuitive way. It equates simply to how many minutes in any given hour a consumer might expect things not to be working well.

The analysis in this document compares two measures of accumulated experience: the mean of the digital VoC scores, and the poor quality time result.

5.5.1.1 Statistical relevance of sample group

As explained in the previous section, users' digital experience quality analysis was matched to the digital services they indicated as used in the survey. In addition, only those dVoC scores from the time periods users indicated as being 'mostly used' were included in the correlation analysis. Hence survey responses informed not only the configuration of what DUs measured but also when their scores were considered as being relevant for analysis.

Finally, an overall measurement volume threshold was applied, such that DUs with fewer than 42 hours of accumulated measurement time across the period of measurement gathering (approx. 2.5 months) were not included in the analysis. This is equivalent to an average of a little over 30 minutes of usage per day, and was applied to avoid bias in subsequent analysis (e.g. if only a very short period of very poor results were seen, this could not be considered unambiguously indicative of that consumer's broadband performance).

For all of the results presented in the following sections, separate correlation analyses were conducted to compare surveyed opinion of quality with digital VoC scores (expressed as poor quality time).

With the exception of two subsets (voice, and superfast/cable) these analyses demonstrate that the correlation is substantive (varying from moderate to strong) and statistically significant at the $p<0.001$ level. For voice and for superfast/cable, the correlation of surveyed opinion with the accumulation of poor quality time is of moderate strength and statistically significant at the $p<0.02$ level.

Hence, it is reasonable to conclude that the correlations observed in the sample group are both substantively significant and statistically significant. This clear correlation between Actual Experience's analysed scores and surveyed consumer opinion exists not only at the level of overall service quality, but also when analysed by application type, package type, and time of day.

5.5.2 *Overall service quality correlations*

The plots in this section show some of the key results of this report – the correlations between consumers' perceptions of service quality and digital Voice of the Customer scores derived from Actual Experience's analysis.

The opinion bandings represent the subjective replies (see §5.3 above) to the survey carried out before analysis started.

The plots below show the results of the analysis, split across the survey opinion bandings:

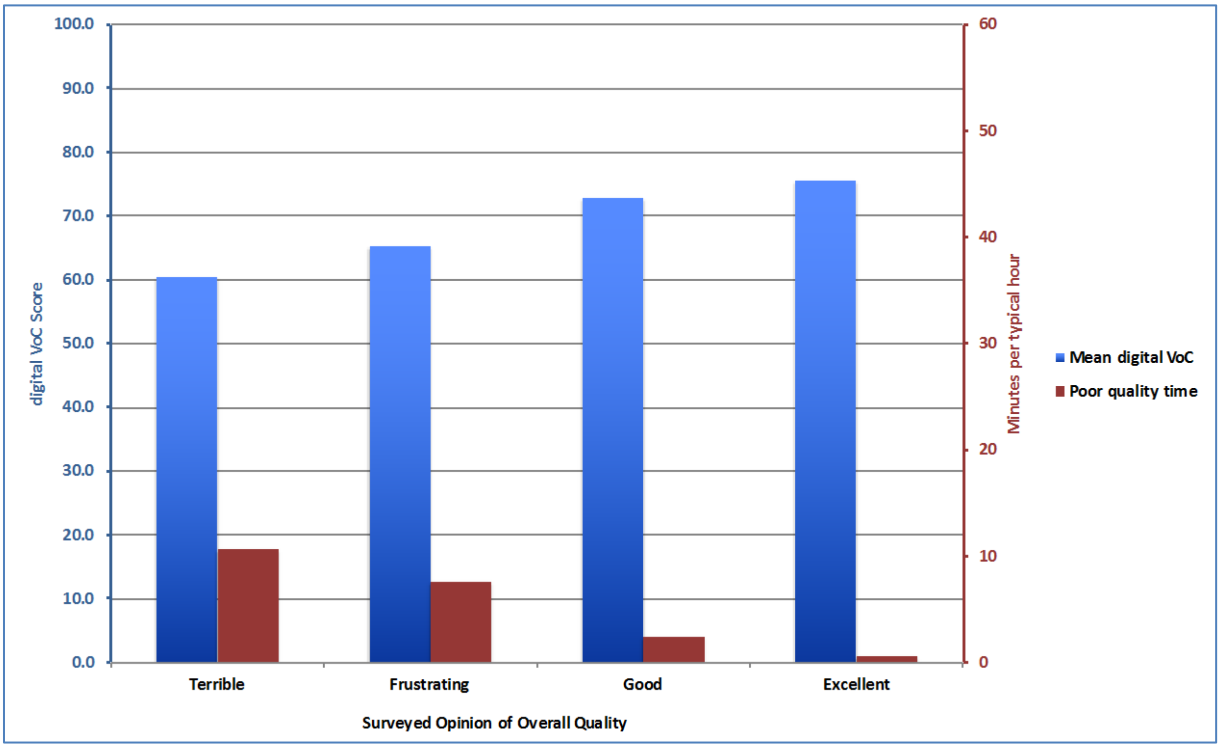


Figure 18 – Digital experience quality analysis overall in each opinion category

Figure 18 shows the change in digital VoC results as we look at respondents' data from each of the four overall opinion categories. The mean digital VoC is taken across the whole measurement period – from the 1st March to the 16th May.

The other category shown – *Poor quality time* – represents an important concept as described in the preceding section. Intuitively, we understand that many instantaneous moments of poor quality build up a perception *over time* of a generally poor quality service. For example, expecting video streams to buffer or freeze once or twice each time a stream is played, or expecting a web page to load slowly every now and then suggests that one will view the service as less than ideal. This measurement represents just that – instantaneous events, not averages, that would represent a moment of significantly reduced digital experience quality to the consumer. The accumulation of these over time, expressed as typically expected minutes per hour, is shown above.

These two results taken together – the change in mean digital VoC, alongside the poor quality time analysis clearly show a correlation with the surveyed opinions. Simply put, higher mean digital VoC + reduced poor quality time = a better user perception of the service.

This result is the first strong indication in this report that Actual Experience analysis does indeed correlate with the views of consumers.

Now we look at those results in more detail. Figure 19 below looks specifically at the typical amount of time a consumer would expect to see poor quality. As discussed above, this shows the linkage between the accumulation of poor quality behaviours and the effect on surveyed consumer perception.

The lines show the frequency distribution of digital VoC scores equating to instances of poor quality, with the 25-75th centile distribution shown as the blue/grey blocks. The line inside the centile blocks denotes the median frequency score in each banding.

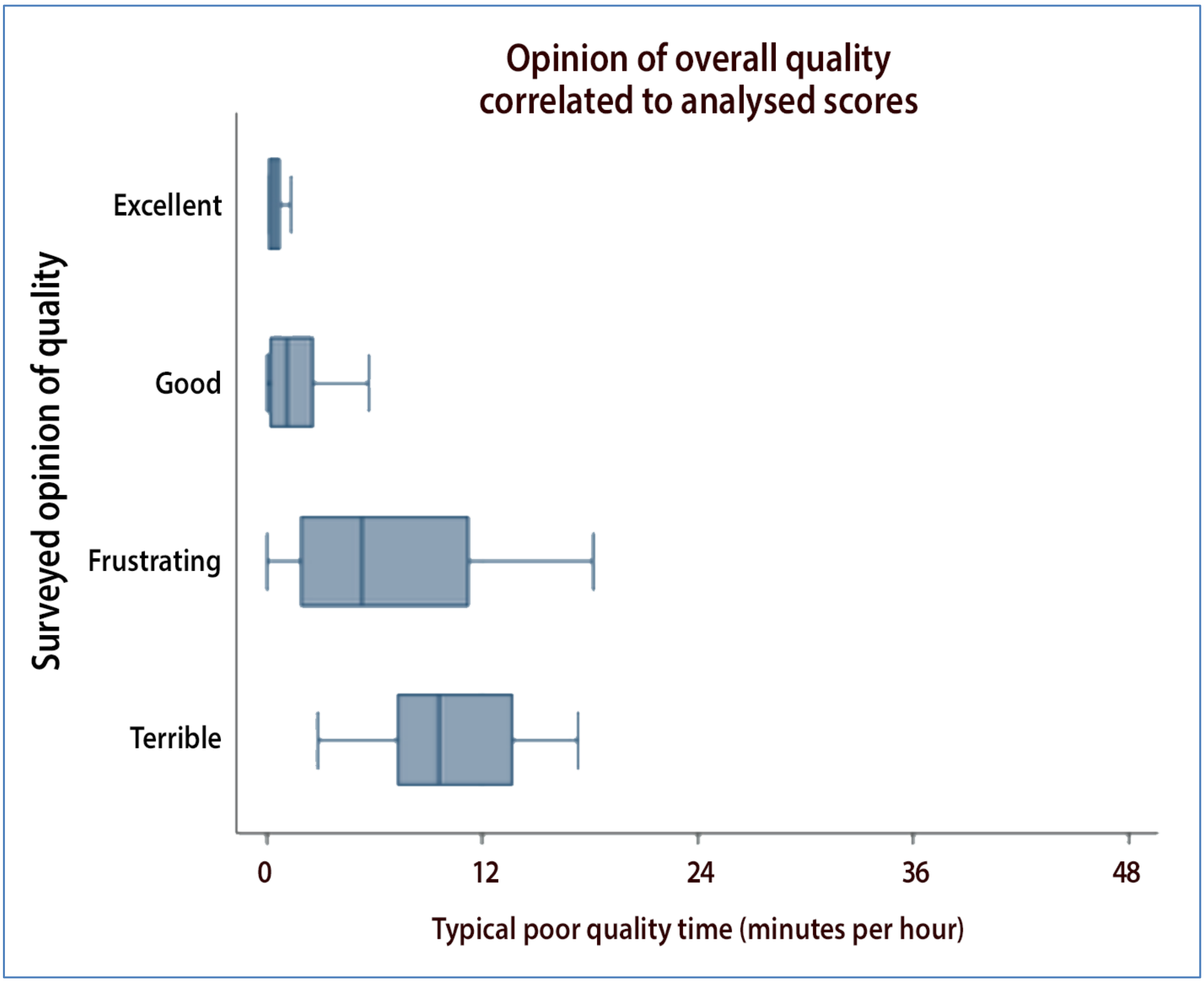


Figure 19 – Correlation of poor quality time with surveyed perception of overall service

As we've already seen, a clear correlation exists between the accumulation of poor quality time seen by Actual Experience and their cumulative effect on subscriber perceptions. Having characterised this, we can now consider the same survey responses from the perspective of mean digital VoC analysis.

In a similar fashion to Figure 19, the lines represent the range of mean digital VoC scores with the 25-75th centiles and median values for that distribution. We can therefore see both the range of results and the variation within each band.

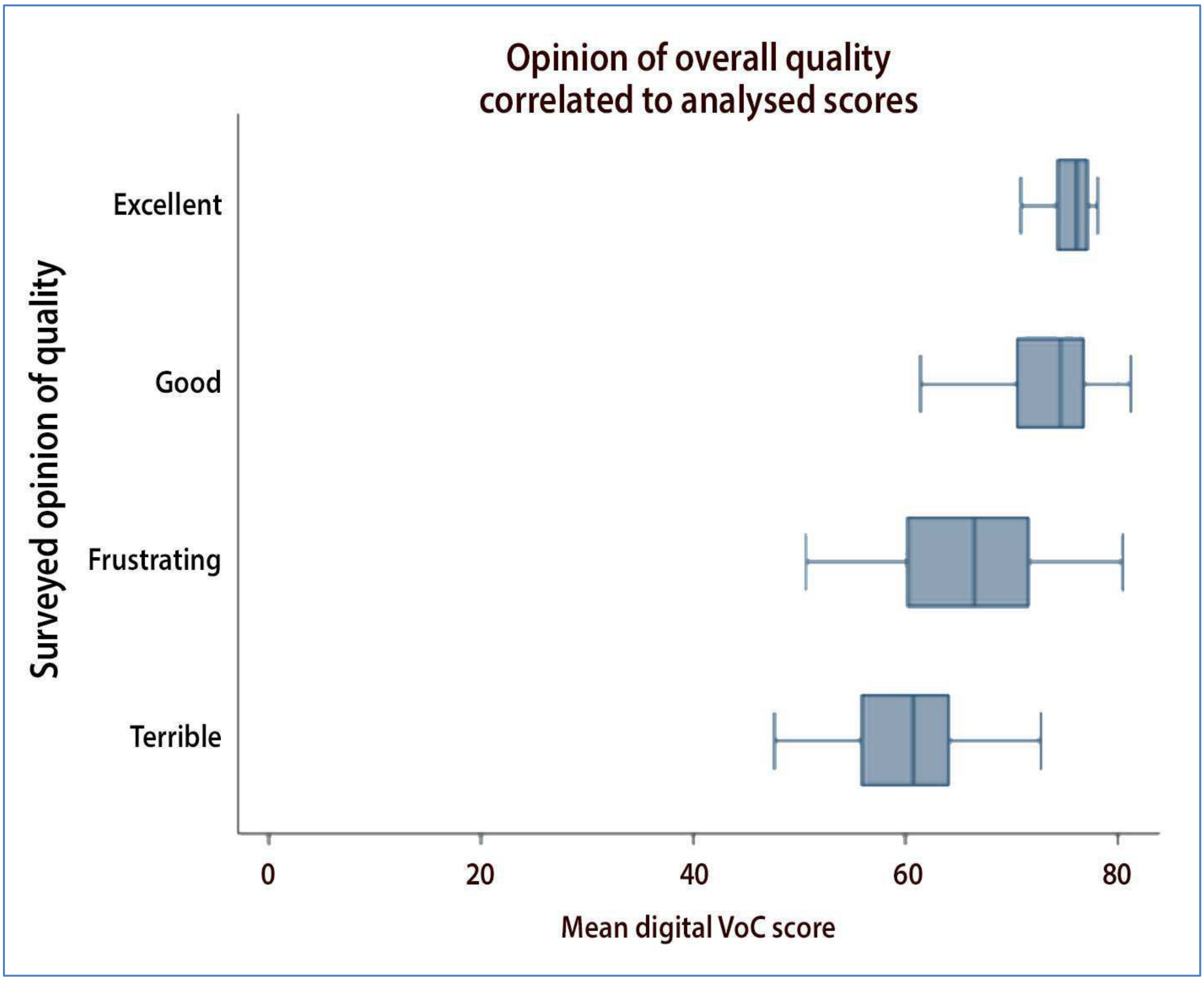


Figure 20 – Correlation of dVoC with surveyed perception of overall service

These results show a clear linkage between Figure 19 and Figure 20, as seen in Figure 18, demonstrating that the reduction in mean digital VoC is driven by increases in poor quality time. This makes sense in the context of how the survey questions are framed, e.g. *"How do you rate the quality of your Internet service overall?"*. The key word is "overall", and it's clear from the analysed quality scores that it is this accumulated experience over time, i.e. the frequency with which quality is affected, that drives people's opinion of their Internet service.

We can now draw the following conclusions relating to how both the accumulation of instantaneous digital supply chain behaviours and mean digital VoC analysis correlates with the survey respondents' perceptions of quality:

- **Excellent banding**. For respondents who consider their digital experience quality to be *Excellent*, very little misbehaviour is tolerated, with the 25-75thth centile bounded less than 1 minute in a typical hour. The median point is at 40s. For the mean digital VoC analysis in Figure 20, the 25-75th centile is between 74 and 77 with the median point of 76. This therefore shows a very clear correlation between user perceptions, high mean digital VoC scores and only rare instances of poor quality.
- **Good banding**. In the *Good* category from Figure 19 we see approximately 4x the poor quality time than for *Excellent* results. This demonstrates how these more frequent behaviours in the digital supply chains are becoming noticeable to users. However, the 25-75th centile range is still fairly tightly bounded at around 2½ minutes in a typical hour with the median point at just over 1 minute. Although poor behaviours are becoming more noticeable, only a limited amount is tolerated to support this perceived level of quality. From Figure 20, we see a greater overall range than for *Excellent* results, which might be expected given the survey definition – *things work well with occasional slowdowns/glitches.* The 25-75th centile covers a slightly wider range than for the *Excellent* banding, stretching from 71 to 77. This is consistent with the increased accumulation of poor quality events that would be perceived by users as an occasional drop in service quality, as opposed to a consistently degraded experience. This result is again consistent with where Actual Experience would expect to see good but not excellent levels of digital experience quality.
- **Frustrating banding**. The *Frustrating* results in Figure 19 shows greater accumulations of poor quality time. The median shifts over 5 minutes and the 25-75th centile ranges are bounded at around 11½ minutes, as users' perception of quality decreases. Figure 20 shows a similar shift, which might be expected given that we are now often seeing more of the poor behaviours that were occurring only occasionally seen in the *Good* category. The 25-75th centile covers around twice the range of the *Good* results and three times the range of that in the *Excellent* plot. This supports the view that variability has a significant effect on the perceived level of quality, even when the digital supply chain is capable of supporting scores (and a digital experience quality) in the *Good* to *Excellent* range. The median point here is 66. Again, these results are consistent with Actual Experience's analysis of a level of quality that renders a service usable, but frustrating.
- **Terrible banding**. Finally, the *Terrible* results continue this trend, with Figure 19 showing frequent instances of poor quality – in some cases more than 1 in 5. The 25-75th centile range is bounded at almost 14 minutes in a typical hour with a median of just under 10 minutes. Figure 20 shows a somewhat smaller range than the *Frustrating* segment. This suggests that very poor perceived quality is now about a more consistent level of degradation, as opposed to just increasing variability. The median point is at 61. As with the higher quality bandings, these results are consistent with Actual Experience's analysis that would describe a digital supply chain as performing so poorly that users could be expected to simply give up, and would have to be very persistent to continue with a service at all.

Considering these results overall, it's clear that there is an unambiguously clear correlation between users' perceptions of quality and Actual Experience's digital experience quality analysis. We see consistent correlations between the frequency of poor digital supply chain behaviours, mean digital VoC scores, and user survey results, showing the effect on user perception of both instantaneous poor behaviours in digital supply chains and the accumulations of those behaviours over time.

5.5.3 Application specific quality

This section takes the overall results above and breaks them down by each application questioned by the survey and subsequently analysed by Actual Experience.

The plots show the results of the digital VoC analysis, split across the surveyed opinion categories described in §5.3 above, from the perspective of the different applications.

For each of the categories, the mean digital VoC score is calculated from the total measurement period. Then, for the same measurement period, the accumulation of quality-affecting events is shown as in §5.5.2 and Figure 18 above.

As for the overall results in §5.5.2, there is a clear correlation showing how the mean digital VoC score is tied to the surveyed opinion of each application; as mean digital VoC increases, so too does overall opinion. There is just as strong a correlation shown when looking at frequency of instantaneous poor quality, in that as that frequency reduces, opinion improves.

This corroborates surveyed opinion with Actual Experience digital quality scores: as mean digital VoCs increase and the frequency of poor quality events decreases, so perception of the service quality improves.

5.5.3.1 Web browsing

The plots below show the results of the browsing analysis, split across the surveyed opinion categories.

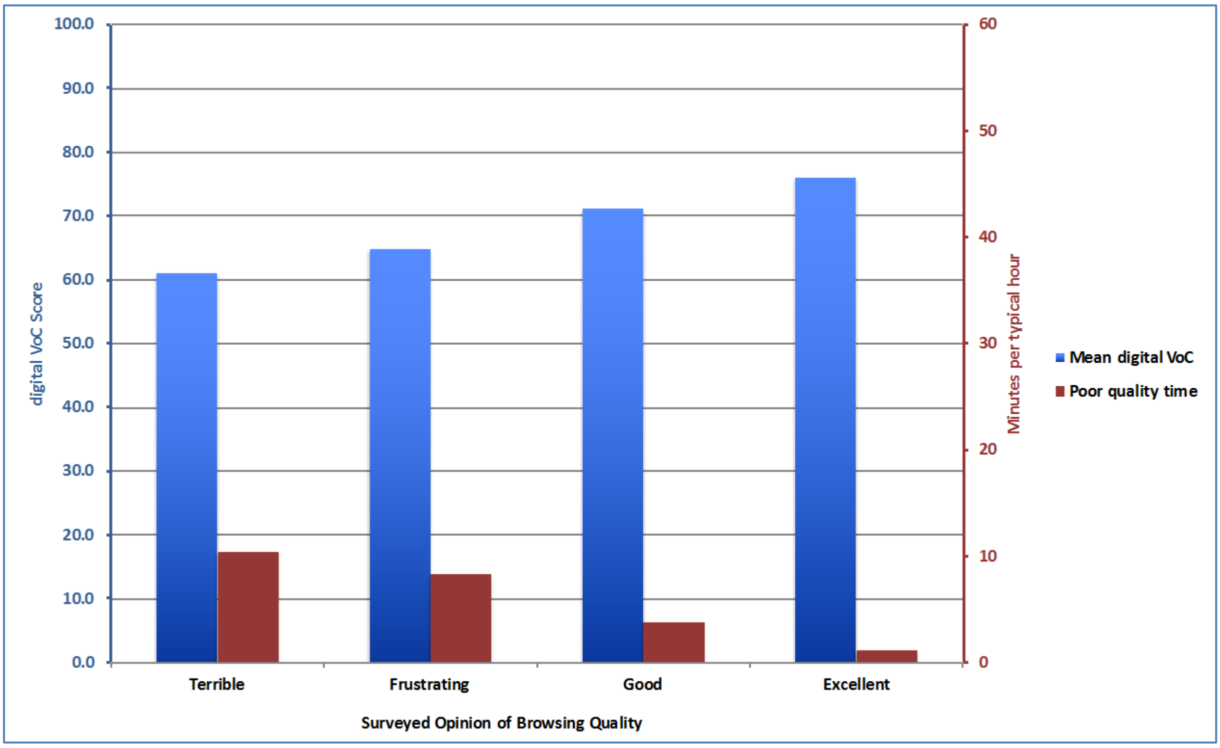


Figure 21 – Digital experience quality analysis for browsing in each opinion category

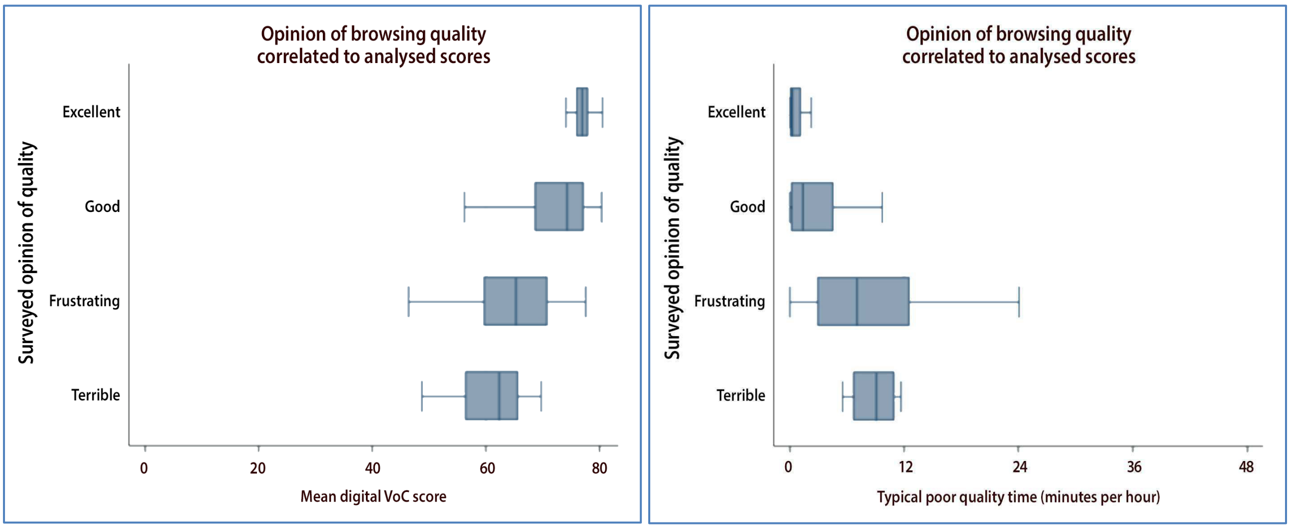


Figure 22 – Correlation of analysed quality with surveyed perception of browsing

Figure 21 and Figure 22 show the change seen in analysed results for browsing, looking from the perspective of each surveyed opinion category.

We can see from the *Terrible* category that poor quality time could be expected to be up to 12 minutes in a typical hour – in other words, for all the accumulated scores seen across the measurement period, up to 1 in 5 indicated experiences of poor quality for the consumer.

The *Terrible* opinion category also had the lowest mean digital VoC score, corroborating this surveyed opinion category.

Taking a closer look at the *Frustrating* category, we can see that a higher mean digital VoC score is achieved, but also a median of just over 7 minutes per hour of expected poor quality. This suggests high levels of variability are a regular occurrence – which simply translates to a highly frustrating experience from the perspective of the consumer of that service.

The surveyed opinion categories of *Good* and *Excellent* follow the trends seen – mean digital VoC increases, and the degree of poor quality time reduces significantly as opinion improves. This again demonstrates solid corroboration between Actual Experience data and surveyed opinion.

5.3.2.1 Video streaming

The plot below shows the results of the video streaming analysis, split across the surveyed opinion categories.

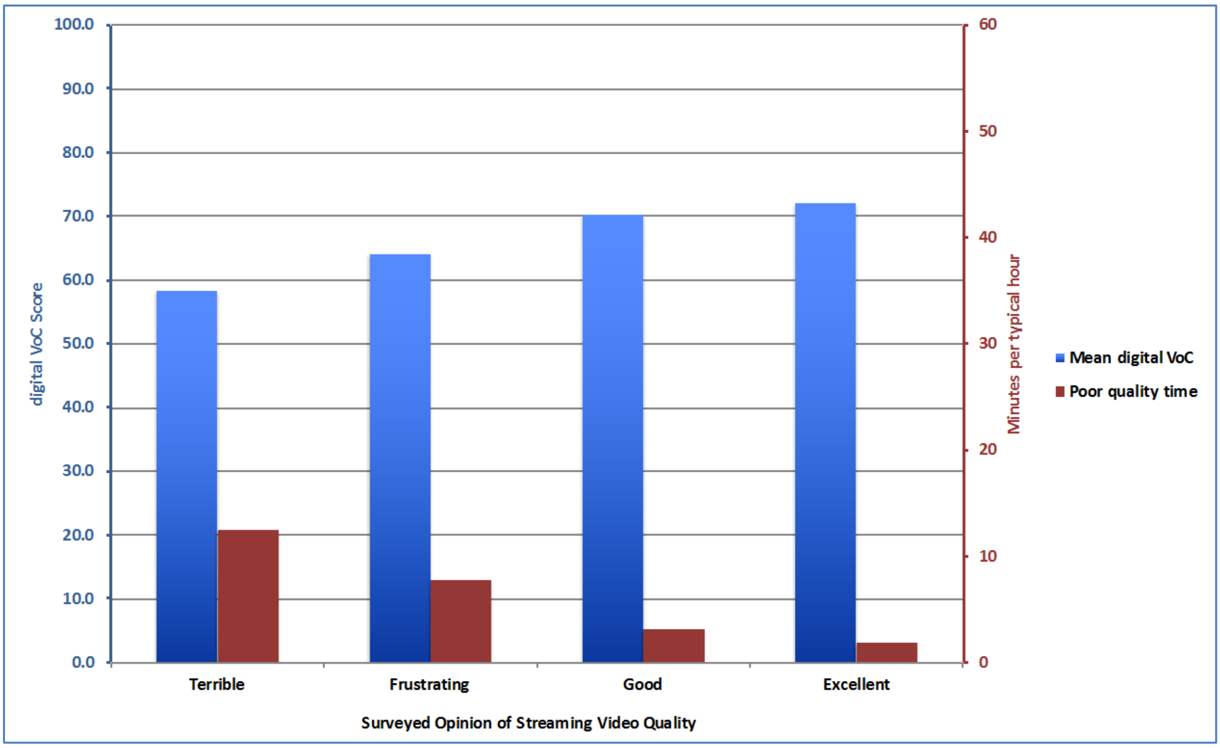


Figure 23 – Digital experience quality analysis for video streaming in each opinion category

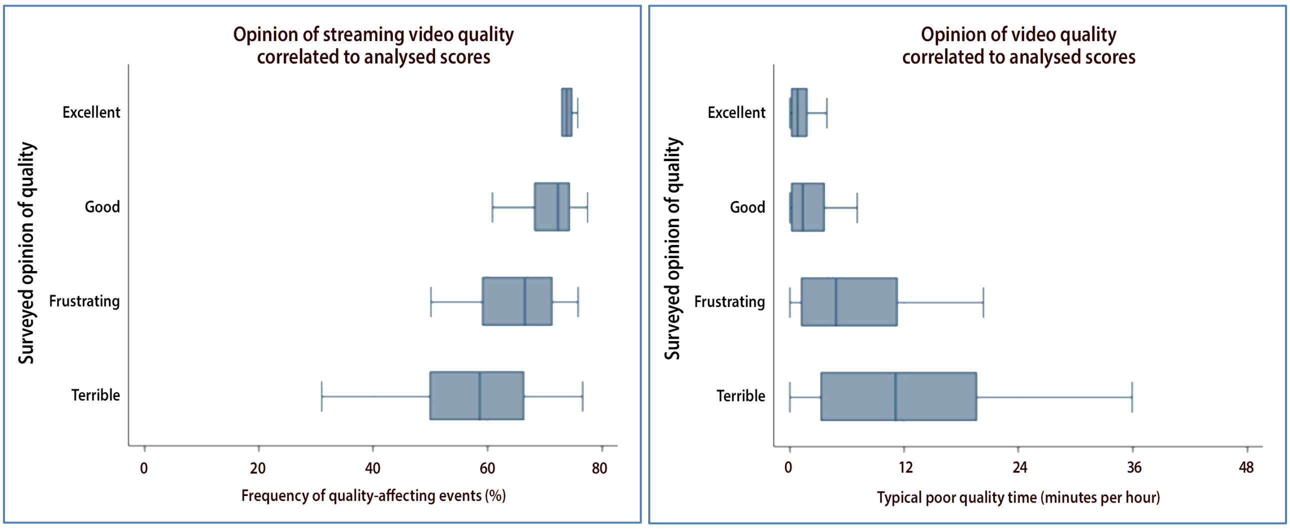


Figure 24 – Correlation of analysed quality with surveyed perception of video

Figure 23 and Figure 24 show the change seen in digital VoC video streaming results across the surveyed categories.

For *Terrible* results, we see an even greater proportion of time where poor quality would be expected – a 75^{th} centile at 20 minutes in a typical hour could only be considered a consistently unreliable service. The mean scores are also low with a mean of 58, combining with poor quality frequency results to show a clear correlation with surveyed opinion.

Taking a closer look at the *Frustrating* category, we again see that this would present itself as a highly variable perception of digital quality from the perspective of the consumer, with a mean of 64 and up to 11 minutes of expected poor quality per hour.

Good and *Excellent* follow the previous trends seen – mean digital VoC increases, and the expected poor quality time reduces significantly as opinion improves. In this latter category, the median is at less than 1 one minute per hour - showing only the smallest tolerance of 'glitches' in the service if perception is to remain excellent. This again shows clear corroboration between Actual Experience data and surveyed opinion.

5.5.3.2 Voice

The plot below shows the results of the voice analysis, split across the surveyed opinion categories.

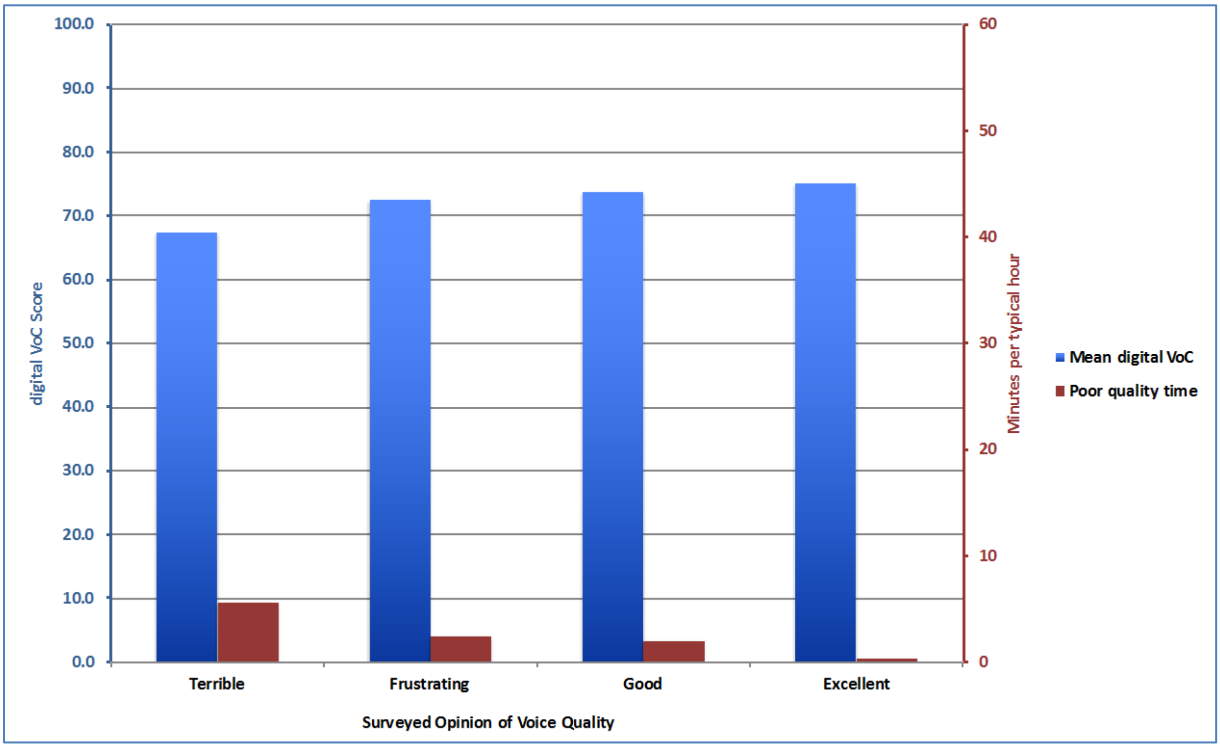


Figure 25 – Digital experience quality analysis for voice in each opinion category

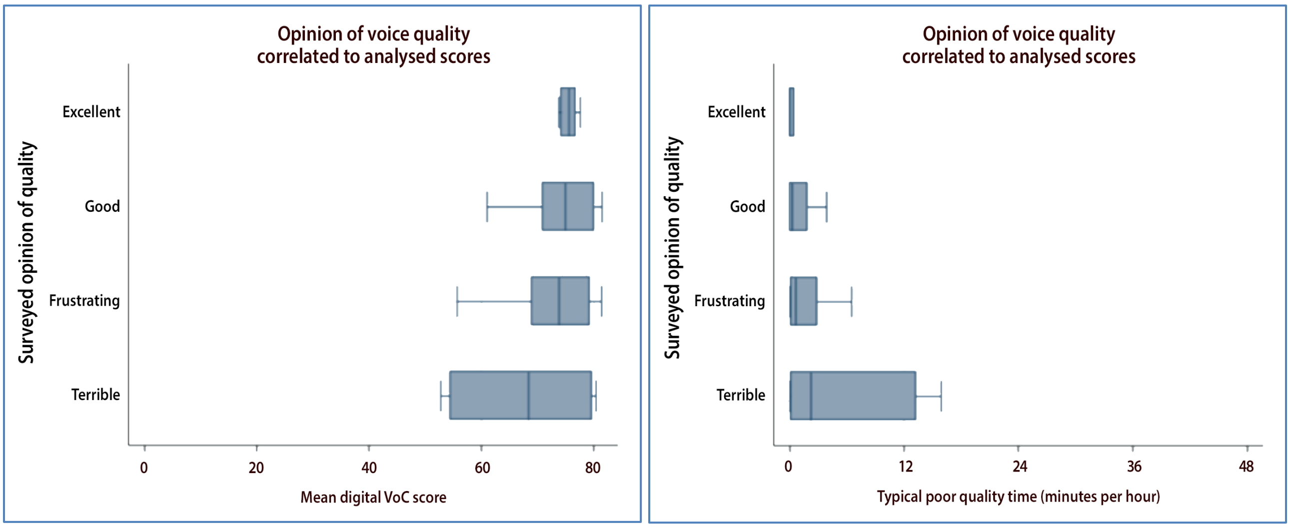


Figure 26 – Correlation of analysed quality with surveyed perception of voice

Figure 25 and Figure 26 show the changes seen in voice results, following the previous trends seen. Again, as mean digital VoC increases, and the expected amount of poor quality decreases, so opinion rises in direct correlation.

Interestingly, the plot also shows that when it comes to Voice over IP digital quality, there needs to be near to zero levels of variability – i.e. the digital VoC scores need to be at a consistently high level, continuously – in order to qualify as *Excellent*. The 75th centile for poor quality time is at less than 1 minute in a typical hour, with the median at zero.

The continuing importance of consistency – the lack of even occasional events that affect quality – is again shown within the *Good* surveyed opinion category, with the 75th centile for poor quality at a little under 2 minutes per hour and the median at just over 20 seconds per hour. This dimension looks to be the critical factor in reducing consumers' perception of the service – a median of less then 1 minute and 75th centile of 3 minutes per hour being sufficient to push opinion into the *Frustrating* category. As with previous examples, *Terrible* perceptions equate to both low mean digital VoCs (67) and very high frequencies of poor behaviour in the digital supply chains – a 75th centile of more than 13 minutes per hour suggesting a frequently unusable service.

This result demonstrates an extremely high level of consumer expectation of quality when considering an application where minimally imperfect behaviours have a noticeable affect on overall perception. This is perhaps unsurprising intuitively – we have experienced essentially perfect fixed-line voice quality for many years and VoIP products often market themselves as alternatives to those traditional services, driving consumer expectation to very high levels.

5.5.4 Time of day variations

5.5.4.1 In general

Figure 20 above shows how the surveyed perception of overall quality correlates with Actual Experience's analysis. Figure 27 below presents the same analysis, but according to time of day. Time bands are grouped according to those in the survey questions: Early (6am - 9am), Day (9am - 6pm), Evening (6pm - 11pm) and Night (11pm - 6am).

As might be expected from previous results, we see a clear correlation along the same lines: increasing mean dVoC + decreasing poor quality time = a better consumer perception.

However, what we see here is another level of granularity with the correlations. In §5.3.3 above, we saw that the evening period was consistently highlighted as when perceived quality was lowest. That result correlates clearly with Actual Experience's analysis here, with mean scores lowest and the proportion of poor quality time highest for the evening section in every surveyed category. This is particularly obvious for customers rating their service as *Terrible*, where we see the 25-75th centile bands for digital VoCs and poor quality expected for between 54 and 66, and 7 to just under 16 minutes in a typical hour respectively.

That this is most evident for those BbFix subscribers whose overall opinion was rated *Terrible*, suggests that their experience of Internet service during the Evening is the determining factor in their overall assessment. This aligns with the very clear survey outcome that the Evening is the worst time for each one of the service types and yet peak in terms of desire to use a service.

The Night period features less prominently in the correlation as a result of very few people surveyed frequently using services during the Night period.

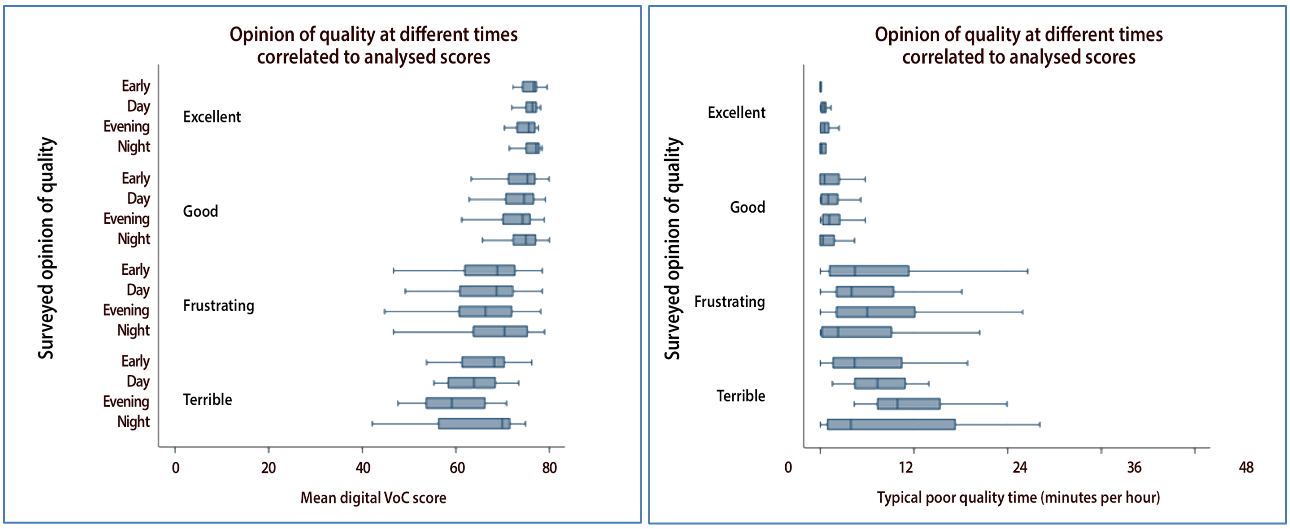
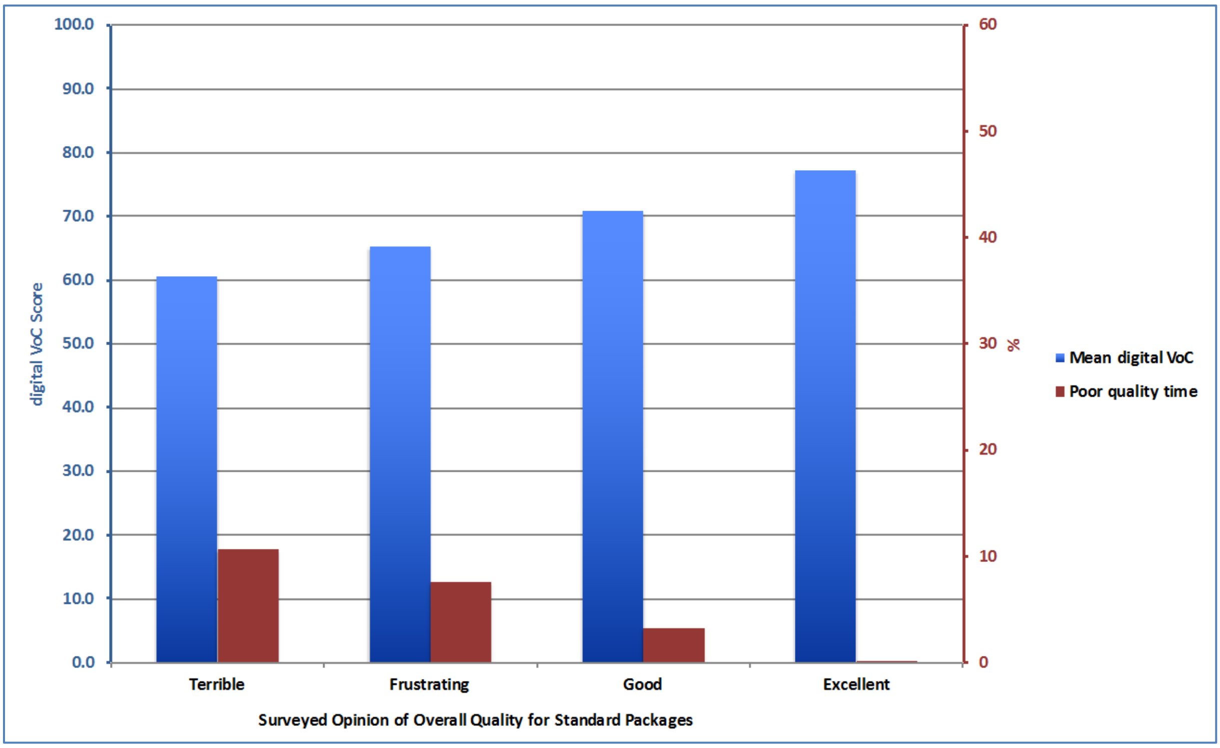


Figure 27 – Correlation of analysed quality with surveyed perception at different times of day

There is insufficient data to reliably extend these time of day results to the specific applications considered elsewhere in this report.

5.5.5 Correlations with broadband package types

The plots below show the results of the digital VoC analysis, split across the surveyed opinion categories described in §5.3 above, from the perspective of different broadband packages – classified as Standard (up to and including 24Mbps/ADSL 2+) and Superfast (at or above 30Mbps).



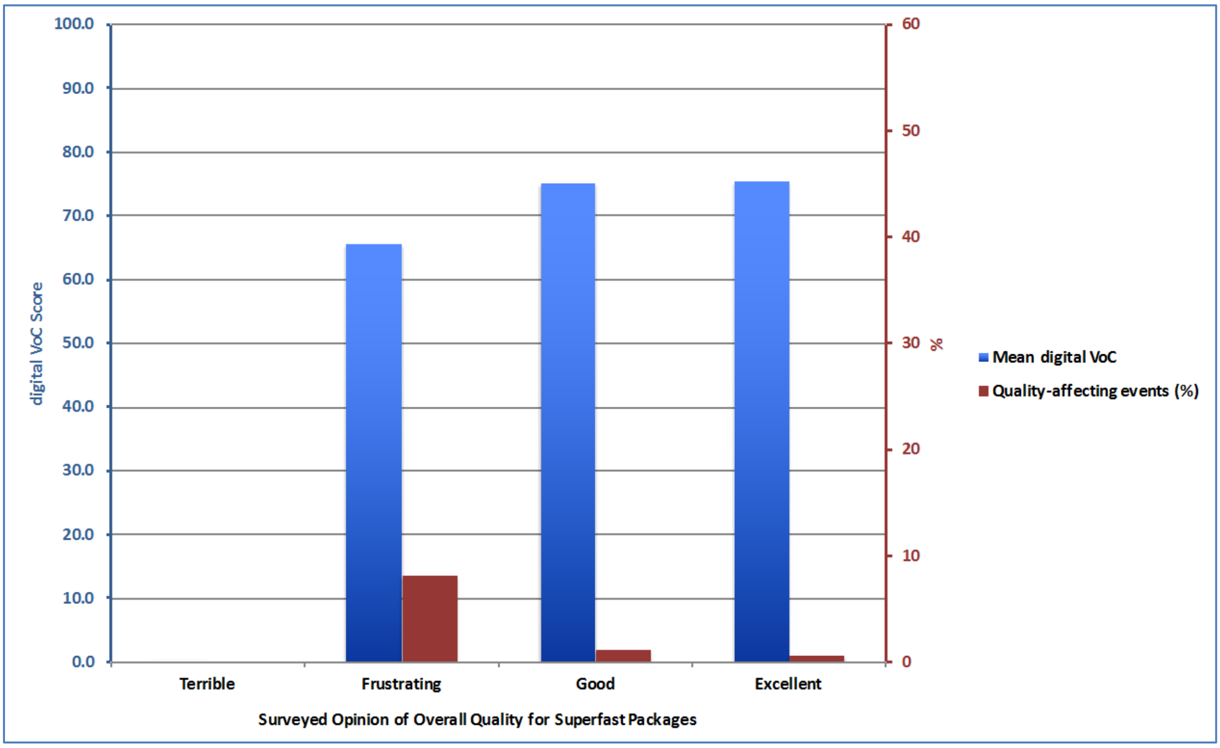


Figure 28 – Digital experience quality analysis for different broadband packages in each opinion category

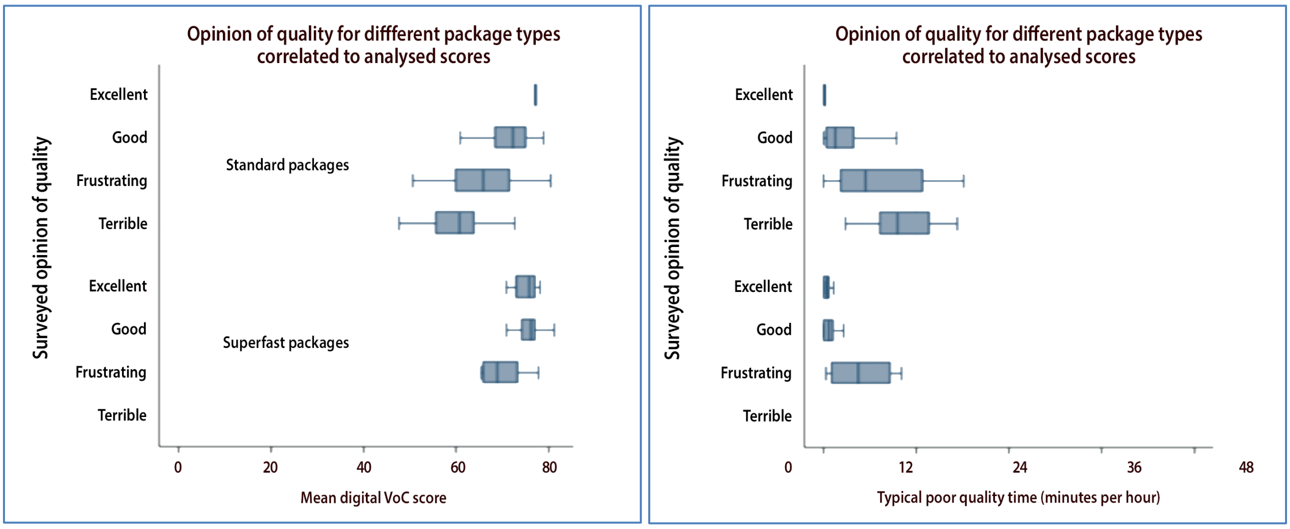


Figure 29 – Analysed quality vs. overall surveyed perception from different broadband packages

Here again we see the expected correlations with both mean scores and poor quality time accumulation.

It is not apparent that there are different expectations of what 'good' means from the different packages. For example, there is marginally *less* tolerance of poor quality instances for standard packages than for superfast (means of just over ½ minute and just under 1 minute in a typical hour respectively) for the service to be considered *Excellent*. The digital VoC means were also similar in this category – 77 and 75 for standard and superfast respectively.

The results in the *Frustrating* category were also similar – mean digital VoCs of 65 and 66 and mean poor quality of just under 8 minutes and just over 8 minutes respectively. In the *Good* category, there is slightly more of a difference – mean digital VoCs of 70 and 75, with means of just over 3½ minutes and just over 1 minute in a typical hour for standard and superfast packages respectively.

There was insufficient data in the *Terrible* category for superfast packages as a consequence of only 1.8% of survey respondents being in this group.

These results are consistent with the view that whilst a certain level of access speed (>8-10Mb from previous work) may be necessary to eliminate consistently poor quality, it is not sufficient to remove the variability – expressed here as the expected poor quality time – that leads to a frustrating experience.

5.5.6 Conclusions

5.5.6.1 General correlation between survey and analysis

This project set out to test the correlation between consumers' real perceptions of quality and Actual Experience's analysis, both overall and for several popular services.

The results are conclusive – unambiguously clear correlations – with general confidence levels of 98% between consumers' surveyed responses and Actual Experience's quality analysis.

The survey opinions corroborate the two areas of quality analysis documented here:

- Mean digital VoC scores. These indicated the quality of the consumer's Internet service overall, or for the specific application or time considered, over the two-and-a-half-month measurement period. In every category, scores increased predictably in each set of measurements as surveyed opinion improved from *Terrible*, through *Frustrating* and *Good*, to *Excellent;*
- Accumulation of poor quality time. This second dimension of the analysis is critical in understanding the trends seen in the mean digital VoC scores. It is intuitive that multiple occurrences of poor quality, however brief, will build a perception over time of a less than perfect service. Actual Experience is able to capture behaviours in the digital supply chains responsible for these events, as they occur and then consider their accumulation over the measurement period. These also clearly correlate with user opinions, and show how the level of tolerance for poor behaviour drives overall opinion of a specific service or overall Internet experience.

Taken together, these two factors are conclusive – simply put, increasing mean digital VoC, allied with decreasing times experiencing poor quality delivers an improvement in consumer perception.

The survey results corroborate the Actual Experience finding not just overall, but for the three service types considers – voice, streaming video and browsing – and for the analysis of service quality at different times of day.

5.5.6.2 Areas of difference

There are no significant areas of disagreement between Actual Experience analysis and surveyed consumer opinion. However, there are areas where there is insufficient data for statistically robust results and further areas beyond the scope of this work, that may be of future interest. These include a more granular analysis of specific services consumed across specific broadband package types and at different times of day and trend analysis of where common quality-affecting digital supply chain behaviours exist within or across specific user groups.

6 Summary and conclusions

This report has built on previous published work for Ofcom – proving both Actual Experience's methodology and that methodology's ability to identify causes of poor quality in complex Internet digital supply chains.

The work here shows that, when service quality is quantified as digital Voice of the Customer scores, there is a clear correlation with actual consumer perceptions of their services. Confidence in these correlations is typically at the 98% level.

The report shows surveyed opinion corroborating Actual Experience analysis in every area considered – overall service quality, for specific applications and for specific times of day. Differences between standard and superfast broadband packages were also examined.

Taken with the previously accepted work, Actual Experience can now present a robust, repeatable, scalable methodology to understand consumer perceptions of Internet services and analyse issues across digital supply chains that affect service quality and those perceptions.

This approach is applicable at large scale – to tens or hundreds of thousands of consumers; for multiple services of interest to consumers or business customers; for different methods of access including fixed line, mobile data and business-grade services; and for any region of interest both within the UK and globally. The analysis behind the digital VoC scores allows a deep understanding of where and how action should be taken to improve quality and Actual Experience is able to present that information, on a continuous basis, in a form suitable for use by both individuals and organisations. Regulators and government, network and content providers, consumers and businesses can each draw the conclusions and understand the actions required to assist their work in improving quality for any consumer of digital products and services.

Appendix A – Specific examples of quality-affecting behaviour

As well as the production of digital VoC scores and quality-affecting event frequencies, Actual Experience's analysis provides a rich insight into the digital supply chains connecting users and content.

The analysis of these digital supply chains delivers an understanding of the behaviours – be they transitory or systemic – that affect the quality of the services consumed.

The examples below show commonly seen issues that harm quality and thus, as seen in the results above, consumer perception of Internet services. These are 'Supplier Management' and 'Supplier Behaviour' dashboards, which show the digital supply chain ordered in terms of the responsible parties and devices most contributing to poor quality, respectively.

Figure 30 – Supplier Management dashboard

Figure 30 above shows how the responsibility for poor quality is rarely, if ever, in a single part of the digital supply chain. The red blocks indicate the qualitative responsibility of each part of the digital supply chain for the reduction in the digital VoC. However, it can be seen that some areas have a significantly greater effect that others with respect to the quality of the service. It is here that efforts to improve performance would have the greatest effect on service quality.

Actual supplier names have been removed, but are explicit on the real dashboards.

For the Supplier Behaviour examples below, we are able to see devices in the digital supply chain as opposed to organisations. The greatest effect comes from those at the top of the list, with lesser effects further down. This ordering assists interested parties in a detailed understanding where to focus time and effort both to resolve quality issues and systematically improve quality.

Supplier-identifying addresses have been blanked out.

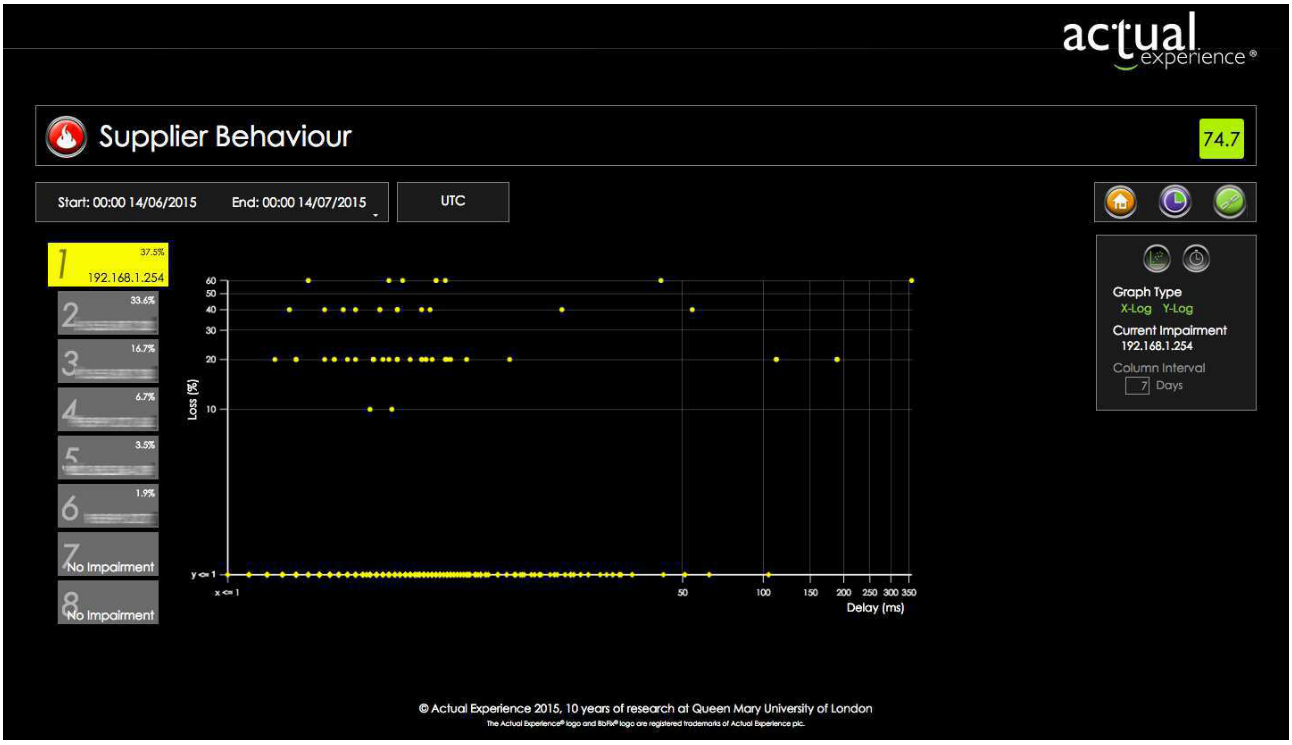


Figure 31 – Provider router significantly affecting service quality

This example shows the provider router as the primary cause of quality issues – almost 40% of the overall problem. This particular result suggests that this may be due to poor wifi performance – either as a result of local conditions (i.e. the consumer is far from the router in a thick-walled house) or misconfiguration, or simply an old router with legacy wifi standards, unable to cope with the demands of the consumer’s usage.

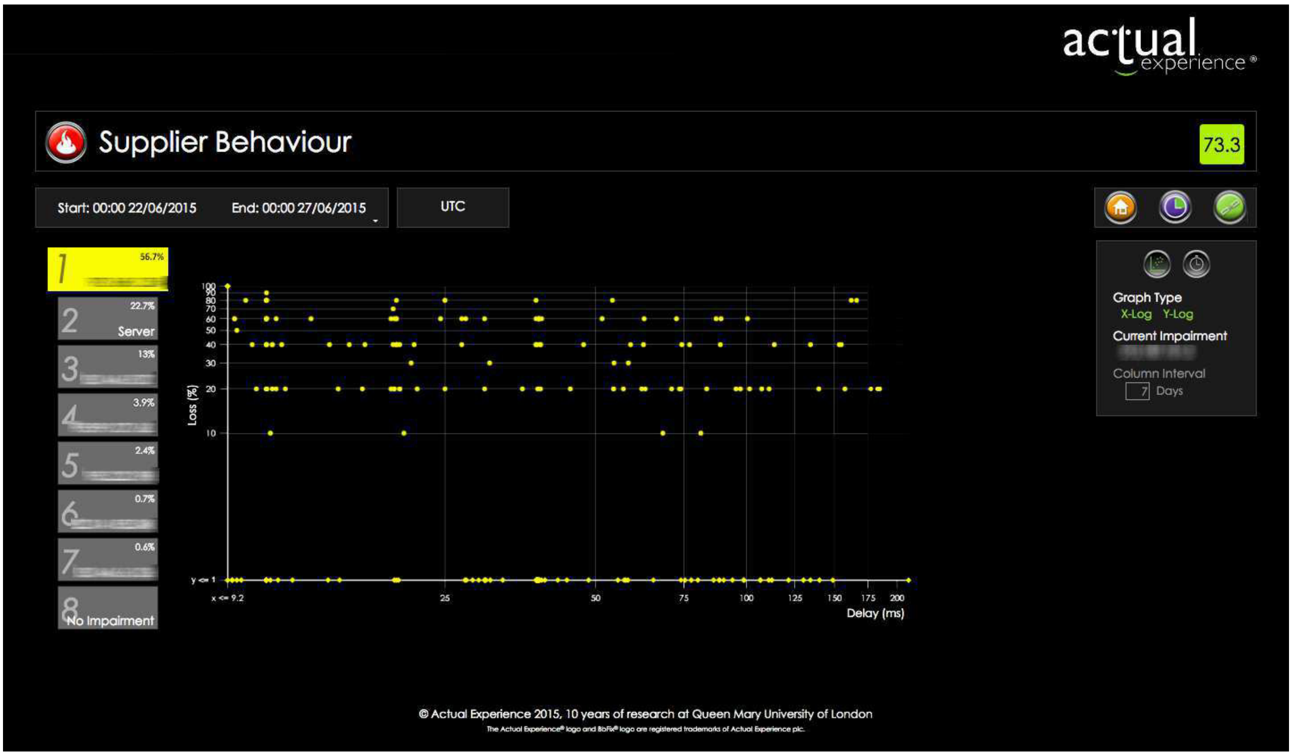


Figure 32 – Network device significantly affecting service quality

Here, we see a network device in an upstream network (identified by the obscured address) grossly responsible for the quality issues being experienced by the consumer.

In this instance, the device would look to be suffering from severe congestion – something well understood and able to be remediated now that it's apparent that this behaviour is mostly (almost 60%) responsible for the quality issues seen by consumers.

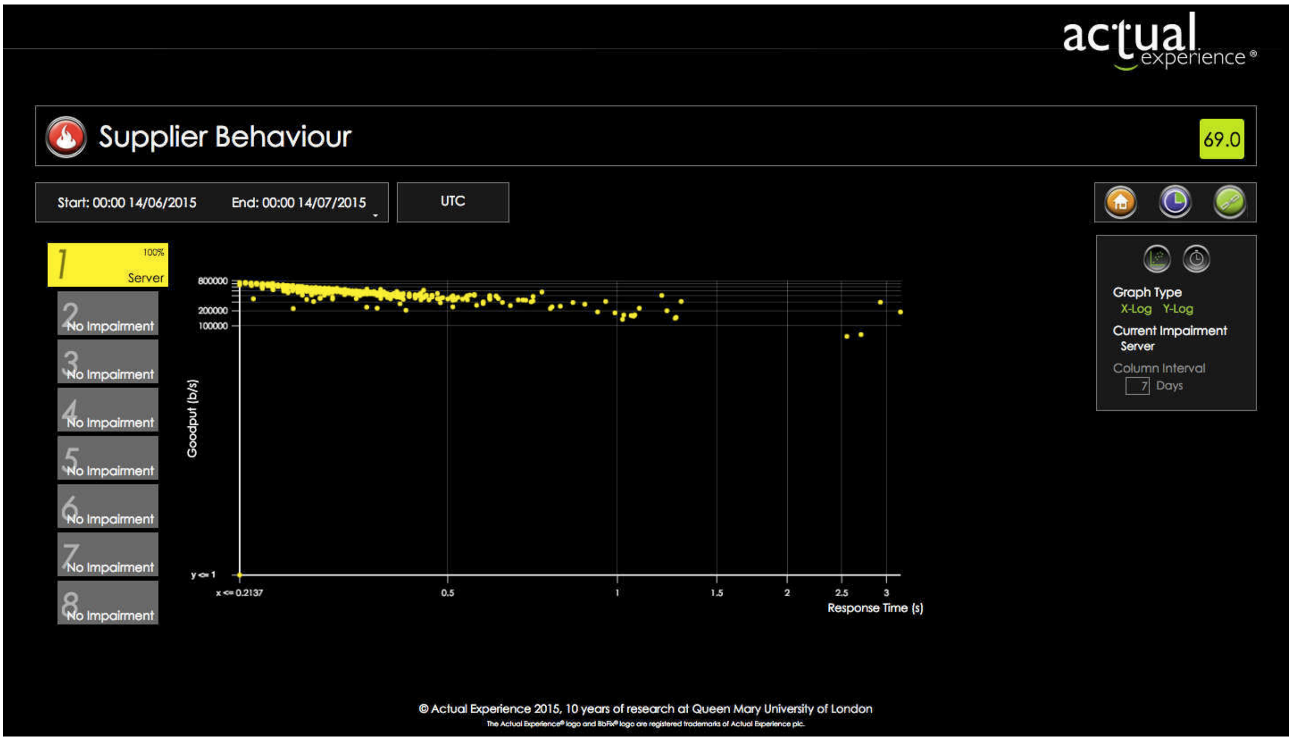


Figure 33 – Content server solely affecting service quality

In this final example, it is the server in the content provider's infrastructure that is solely responsible for quality issues. This is an example of where the broadband provider is not at all responsible for the issues seen by their customers, unless they are able to influence the content provider themselves.

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄七

Arcep State of Internet in France 2017: Quality of service, from the perspective of UFC-Que CHOISIR

Three QUESTIONS TO

UFC-Que Choisir (French consumer association)

Antoine AUTIER, *Deputy Manager of UFC-Que Choisir*

QUALITY OF SERVICE, from the perspective of UFC-Que CHOISIR

1 UFC-Que Choisir publicly paid tribute to the discontinuation of the controlled environment: could you review the reasons for this with us?

In and of itself, a controlled environment can offer advantages, particularly by focusing on the quality of networks as such, and by removing from analysis some biases stemming from the wide variety of ways in which consumers use the Internet. That being said, this type of device becomes an issue when the investigation protocol associated with it is known to the Internet service providers (ISPs) tested. The risk that they will optimise the lines tested cannot be ruled out, and as a result, the test results could conceivably overestimate the actual capacities of the networks. However, the mechanism recently stopped by Arcep gave the ISPs too prominent a role in the protocol's development, hence UFC-Que Choisir's strong reservations from the start as to the mechanism selected.

> “The major benefit of crowdsourcing lies in the fact that it enables all consumers to send data that reflect their user experience.”

2 What do you see as the advantages and limits of the crowdsourcing method?

The major benefit of crowdsourcing lies in the fact that it enables all consumers to send data that reflect their user experience; consequently, the results at the end of the chain are actually connected with the wide variety of consumer realities, enabling them to compare not only ISPs, but also all Internet technologies. What's more, the tool on which crowdsourcing is based can enable each consumer to access indicators on the quality of his or her connection and, if desired, be able to compare and contrast them with information on the alleged quality of the advertising campaigns.

As such, crowdsourcing is not without shortcomings, however. For example, if the number of consumers using the crowdsourcing tool is not high enough, this comparison may lack relevance, as the results from the ground are not representative of all situations possibly encountered by consumers.

Furthermore, and even though these biases were raised by widespread use of the tool by consumers, questions about the quality of information processing could persist.

3 How can technical results be conveyed to consumers in a clear and educational way manner?

Along with the price charge for Internet access, the quality of service provided to consumers is a key factor on the basis of which they choose their Internet access offer. It is therefore essential that they be able to access clear and relevant information on the quality of service for all fixed Internet offers. This information needs to be seen broadly, and not be limited to speeds alone. For example, the quality of the home Wi-Fi, which is currently widely used, must be carefully addressed. Similarly, IPTV quality tests of can no longer be overlooked. Furthermore, given the high stakes involved in interconnection, the quality of the Internet services espoused by consumers deserve to be the focus of a special explanatory effort.

At the same time, this approach opens up the risk of flooding customers with information, and thus muddling their understanding. It is for this reason that, beyond technical results, it is important to decrypt information and make it intelligible for consumers. This is one of UFC-Que Choisir's abiding aspirations.

methodological approaches used and the heterogeneity of the results measured. More often than not, this diversity is explained by the varying objectives sought by the various tools.

That being said, a harmonisation in the measurement methodology is important. Without minimal standardisation, it is difficult to draw up comparisons between geographical zones or between operators, analyse changes in performance over time, or allow an end user to formally compare the actual performance of his/her Internet access with those indicated in his or her contract.

This is the challenge raised at the European level by the BEREC working groups and the European Commission mapping project. Arcep takes an active part in this, contributing knowledge that will fuel the reflections of these bodies through its work at the national level and by regularly interacting with European stakeholders.

a) *Ecosystem mapping: tools available on the market*

The ecosystem of crowdsourcing metrology on the quality of fixed services is far-reaching and diverse. The following study is based on responses from ten existing players to a questionnaire sent out by the Authority as part of its call for partnerships open to any interested organisation.

The players were assigned to one of three more or less uniform groups:

The "hardware sensors": sensors located on the client side (on the box, operating an Ethernet bridge or simulating a terminal) that automatically perform quality of service measurements.

"Web testers": testers accessible online by the general public, also referred to as "speedtests", which makes it possible to measure the flow (or latency, etc.) of its fixed Internet connection.

"Other software solutions": a broader category that includes both server solutions (mscore), software agents embedded in boxes (cloudcheck) or on web pages (Radar script).

Many of the players have developed a variety of solutions and could fall into multiple categories. This is true, for example, of Gemalto, which in addition to the hardware probes owns an online tester intended for companies. For the sake of concision, the study will focus hereafter on each player's main tools, as defined in the mapping above.

Where existent, the diagram indicates, below each player's name, the commercial name of the associated tools, which are sometimes more well-known to the public.

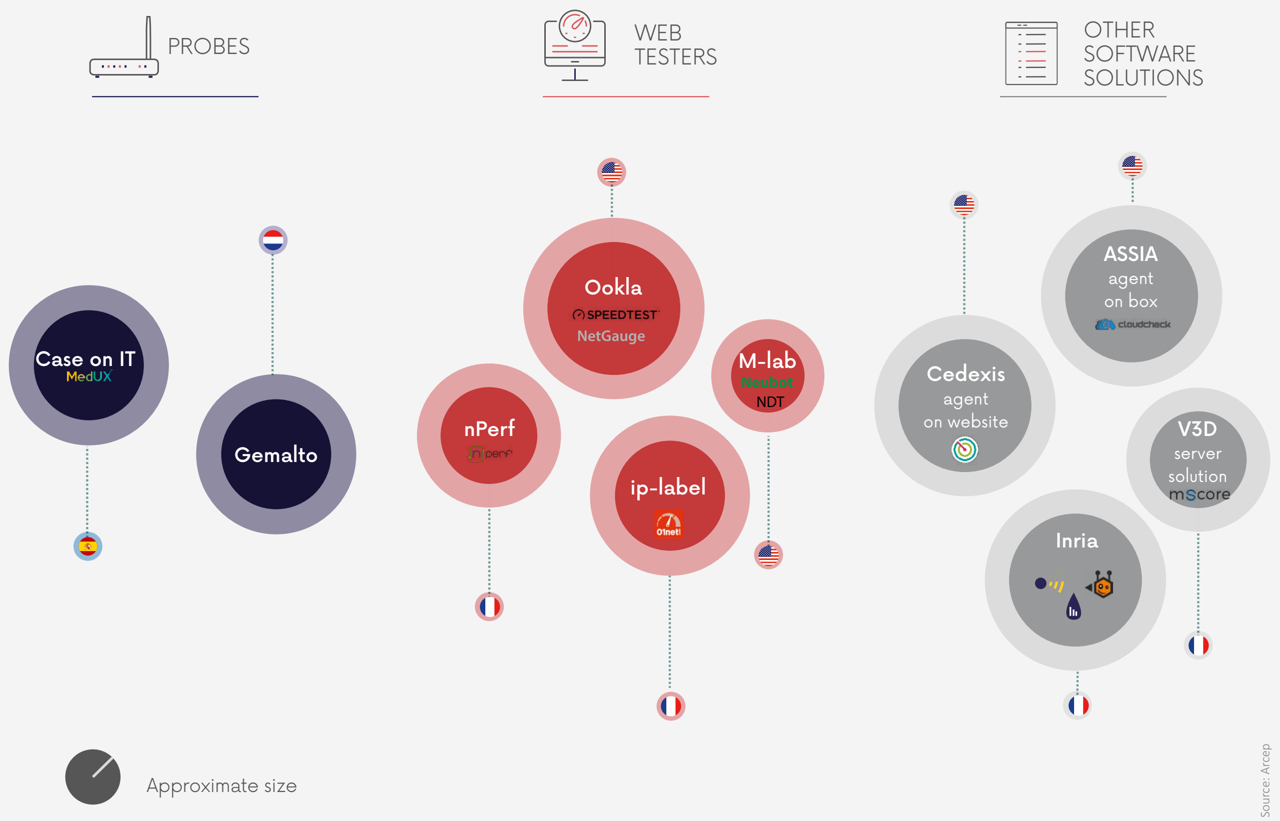
The scope of research at Inria and M-lab goes well beyond quality of service of electronic communications networks. Similarly, quality of service accounts for only a small part of the Gemalto Group's activity. On the other hand, the activity of the seven remaining players is entirely dedicated to those topics (in the broad sense). The size of the bubbles roughly reflects the size of the relevant player in approximate number of employees working on quality of service. It does not prejudge the intrinsic value of the solutions proposed.

Most of the players shown are for-profit companies. Their core business and their positioning on the value chain vary widely. Although all players operate a B2B model (business to business), some of their tools are known to the general public via regular publications of their figures and analyses, as a result of which they can often gain visibility from their corporate clients.

Five of the eight for-profit companies included in the study earn turnover that comes in very large part from ISPs; two of them report moderate to low dependency on ISPs; one (Cedexis) is almost not dependent on them at all.

M-lab and Inria, in contrast, are non-profit organisations. They develop technology offered in open-source mode and report their data in open data (for most of their tools).

A rich and varied **ecosystem**



The table below shows the main sources of revenue for players participating in the study.

The column "sales of quality of service or experience data" encompasses two types of scenarios: data sold and subsequently collected by players via their tools; and sales of marketing claims or data licensing (which make it possible for an ISP to communicate the results published by a given tool). In both cases, the data belongs to the tool that has produced them.

In other situations, players do not sell data directly but rather the metrology service, i.e. a technology or infrastructure that can, for example, be offered on white label.

Some companies also offer to manage and optimise their clients' network. These can be ISPs (this is the case of ASSIA) or content providers (this is the case of Cedexis). Cedexis's core business is quite different from that of other players: the company offers its customers the opportunity to improve the availability and speed of their website by directing their traffic to CDN platforms, clouds or data centres that show the best performance – estimated by Radar tests or other external measurement sources – at a given time and location.

Lastly, there are also other sources of income not shown in the previous table. This includes the sale of advertising inserts on websites of certain online testers.

▶ **Note:** *Inria owns 4 tools dedicated to measuring the quality of the distinct fixed services that serve different objectives (ACQUA, APISENSE, Fathom, Hostview). As their methodologies vary widely, in order to remain brief, we will not detail them hereafter in this publication.*

// Quality of Internet **access service**

Benchmarking of existing tools: methodological section

Type of activity	Commercial			Non-commercial
	Sale of QoS/ QoE sales	Sale of metrology services	*Network monitoring*	R&D
Case on IT		●●		
Gemalto		●●●		
Ookla	●●●	●		
nPerf	●●	●●		
ip-label		●●●		
M-lab				●●●
ASSIA			●●●	
V3D		●●●		
Cedexis	●		●●●	
Inria				●●●

Source: Arcep

Tests conducted and methodology employed

There are two major types of performance indicators: technical indicators (speed, latency, jitter, etc.) and usage indicators, which refer to actual uses (web browsing, video playback in streaming, peer-to-peer download, telephone/voice on IP, etc.).

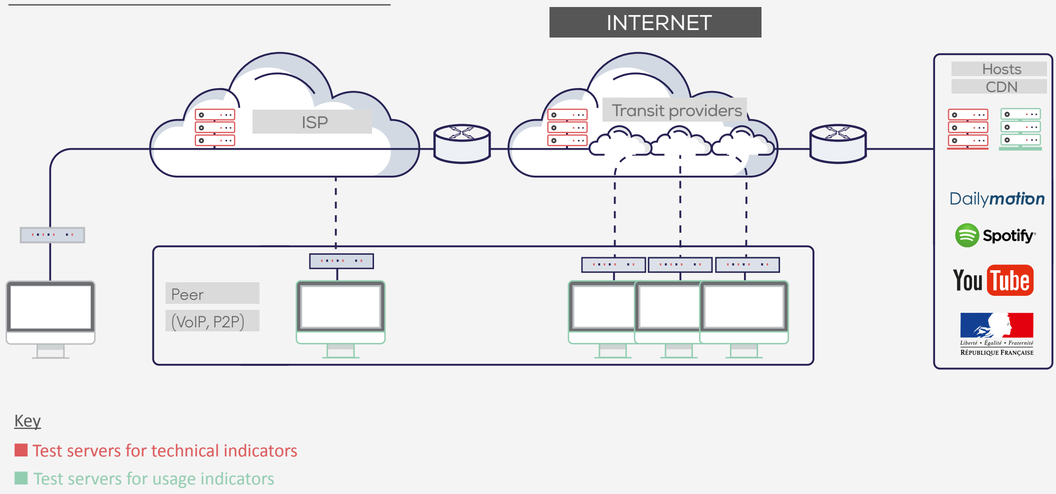
By definition, the test server on usage indicators are located at the level of the content provider's (Youtube, Skype, etc.) host. The test servers on technical indicators can be located within varying distances of the user. The closer the test server, the more the quality indicator depends only on the ISP's network performance. The controlled-environment system also showed the considerable impact of the servers' location on the indicators (more than 30% on the download speeds and over 50% on upload speeds – see section 3.1.2 b).

Technical indicators

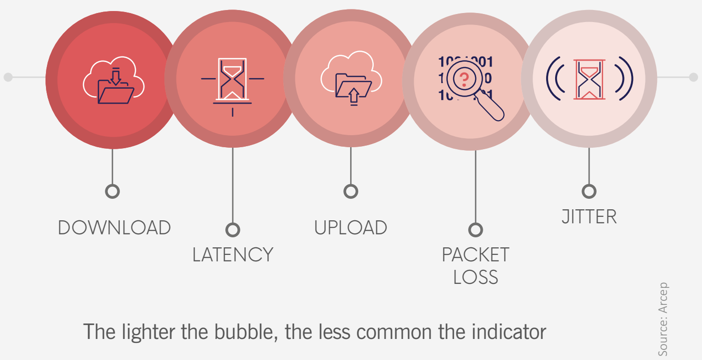
In addition to their location, the connectivity of the test servers required to measure the technical indicators can influence the outcome of the measurement. If inadequately sized, the speed measurements will be artificially capped. The test servers of the various systems taking part in this study showed relatively similar connectivity levels: approximately 100 Mbps for the old generation servers, 1 Gbps for the current servers, and 10 Gbps for the next-generation servers, which are designed to respond to the risk of saturation that could arise when performing simultaneous tests based on very high-speed technologies.

However, the total number of servers varies greatly from one tool to the next. While the ip-label device contains only one server, nPerf and Ookla have more than 300 and 6,000 across the world respectively.

Location of the test servers



Most measured technical indicators



The vast majority of devices select, by default, the server that is closest to the user – toward which latency is the lowest. Mechanically, the larger the number of servers, the more the server selected by default is likely to be located in the user's ISP network. When servers are located outside of the ISP networks, it is important to ensure that the servers have a similar connection between the different ISPs, to avoid possible discrimination. Each of these factors contributes to explaining the differences in results between the systems.

The tools most frequently measure the same technical indicators: download speed, upload speed (except at Cedexis), the latency (and sometimes its derivative, jitter), and packet loss (measured in all players except nPerf, ip-label and Cedexis).

The throughput is calculated by dividing a volume of data sent from the server to the customer (download speed) or from the customer to the server (upload speed) by the total transmission time.

File transmission can take place in monothread or in multithread (parallel use of individual threads or "simultaneous sessions"). While the measurement in monothread is closer to a usage indicator (download of a file that would be hosted on the test server), that in multithread can help saturate the link and therefore estimate the capacity of the line.

It is important that a variable be set to limit the duration of the test, whether the volume of the file transmitted or the transmission time. The technology tested (which is rarely known in advance) is of importance. If the file volume is very large but the test is performed via long xDSL lines, for example, the test will be very lengthy and tend to discourage the user initiating it. Reciprocally, if the file is small and a very high-speed technology is tested, it will be downloaded very quickly, and the flow curve will not exceed the phase known as slow-start (gradual flow increase planned by the TCP protocol): the measured speed will thus not be representative of the actual speed available. When the test duration is pre-set, it is necessary to determine the time needed to reach

// Types **of tests**

Benchmarking of existing tools: methodological section

Speed measurement methodology	Protocol	Encrypted flow	Monothread or Multithread	Fixed variable	Displayed value(s)	Slow-start included in the result displayed (respectively)
Case on IT	FTP ; HTTP	yes*	Mono*	conf.	Avg; Max	no ; no
Gemalto	IP	yes	Multi	t = 10 sec*	Min ; Avg ; Max	yes* ; yes* ; no
Ookla	TCP ; HTTP	no*	Multi	conf.	Avg**	no
nPerf	TCP	yes	Multi	t = 15 sec	Avg ; Peak***	yes ; no
ip-label	TCP	yes	Multi	t = 7 sec*	Max	no
M-lab	TCP ; HTTP	yes*	Mono	t = 10 sec	Avg*	yes*
ASSIA	TCP	no	Multi	t = 5 sec*	Avg 98^{e} percentiles* max	yes; no ; no
V3D	TCP ; UDP	no*	Mono*	V = 5 Mb* or t = 10 sec*	Avg 10^{e} et 90^{e} percentiles	conf.
Cedexis	TCP ; HTTP	conf.	Mono	V = 100 ko	percentiles	yes

Source: Arcep

Key
conf. : configurable
* Recommended or default value (the variable is configurable).
** Average calculated on a dataset excluding rates in the fastest 10% and the slowest 30%.
*** The peak speed is defined as the average of the rates calculated on 30% of the test, the window selected being the best (generally at end of test).

cruising altitude, without discouraging the user. The slow-start phase, very often included in the measurement, can then be taken into account or excluded subsequently from the calculation of the average flow computed over the duration of the test (in which case the listed speed is higher). The question does not arise when, for example, the maximum speed reached over the period is displayed. The choice of exposure value is of major importance when the results are intended to be presented to the general public, often highlighting only one number in particular.

While all the tools measure round-trip latency, some use the TCP protocol and measure the time elapsed between sending a request and receiving the tracing request (Round Trip Time or RTT) [25], others use the UDP protocol and measure the time

[25] Except Cedexis, which measures the time between the start of an HTTP request and the start of receipt of the query, on a query where DNS resolution and the establishment of the TCP connection are already established.

// Types **of tests**

Benchmarking of existing tools: methodological section

LATENCY measurement methodology	Protocol	One-way or round-trip?	Time-out	Number of samples	Displayed value(s)
Case on IT	ICMP	Round-trip	conf.	min. 1	Min ; Avg ; Max
Gemalto	ICMP ; TCP ; UDP		5 sec*	10*	Min ; Avg ; Max
Ookla	TCP ; HTTP	Round-trip	20 sec	approx. 10	Min
nPerf	TCP		3 sec	approx. 20	Min ; Avf
ip-label	TCP		Conf.	approx. 10	Min
M-lab	TCP		Conf.	approx. 100	Min
ASSIA	TCP	Round-trip	5 sec	5*	Avg ; 98e percentile* ; Max
V3D	TCP		Conf.	10*	Min ; Avg ; Max
Cedexis	TCP ; HTTP		4 sec	1**	N.A.**

Source: Arcep

KEY
conf. : configurable
* Recommended or default value (the variable is configurable)
** The case of Cedexis is somewhat unique in that a Radar session records only one measure per CDN, Datacenter, or Cloud tested, but then aggregates all the samples into its reporting in percentiles (10th, 25th, 50th, 75th, 90th, and 95th percentiles).

between sending a message and receiving the same message after server or client reflection, and some use the Ping command to launch a ICMP query.

The number of samples from which the displayed value (minimum, average, percentiles or maximum) is derived varies depending on the tools. To make this choice, the decision must once again be made between statistical representativeness and test duration (likely to discourage users). Time-out, or the point at which a query is considered to have failed, also has its importance: the later it comes, the more the latency tests are included in the sampling, and the higher the result displayed.

Usage indicators

Usage indicators offer significant interest. Based on actual practices, they are more representative of the user experience, and therefore more intelligible and likely to effectively inform choices on access technology or ISP. The observation is shared by most of the tools included in the study: while speed remains a factor that counts, but what matters most to consumers is whether the services they use are working properly.

Five of the devices presented measure usage indicators: hardware sensors (Case on IT, Gemalto) and three software solutions (Cedexis, Inria, V3D).

As each tool has defined its own approach, measurement methodologies are still highly varied. Not only can usage as such (web browsing, voice on IP, streaming video, etc.) be simulated or real, but the associated performance indicators also differ. While some tools continue to be based on the indicators referred to hereafter (mainly speed), others are used with new measures directly connected with evaluated use (time required to load a web page, fluidity of voice on IP or video streaming, etc.).

Measurements made on streaming-based video playback illustrate this.

For example, mscore (V3D) simulates, from a test server, a data stream comparable to a video stream, by setting a variety of parameters: average speed, inter-packet time, buffer depth, etc. It then evaluates service deterioration caused by introduced by the end-to-end crossing of the digital network on the simulated flow, by measuring technical performance indicators. These indicators are then grouped together as a single rating, using a configurable scoring method.

In contrast, other tools choose a given YouTube video imposing minimum quality and duration criteria. Some pure (video) usage indicators are collected: time required for the video to initially load before its launch, number and duration of the stalling episodes.

The indicators measured by the tools developed using Case on IT and Cedexis are relatively similar to those shown above. However, the Cedexis measurements come from all pages displaying videos players that would have deployed the Radar client, rather than from a single YouTube video. Furthermore, additional indicators are measured: time required to load video chunks delivered to users, their latency and their speed. These are then correlated with the usage indicators measured elsewhere in order to quantify the impact of these QoS metrics on the user experience.

Data processing, analysis and transmission

Once the measurements have been completed, reprocessing rules can be applied to the data collected: deleting measurements outside predefined thresholds, in the event of test server unavailability, by robots, etc.

Most of the time, measurement service providers leave their clients to make their own adjustments based on their needs. Generally speaking, with the exception of a few tools, little action is undertaken to drastically combat fraud.

The question of data transmission is twofold.

Each tester is not automatically provided with its individual data. Indeed, only one third of the tools allow access to test history. Once again, the existence or absence of this access is not determined by the nature of the tool.

The distribution of third-party data to a client and/or the general public (through the publication of observatories) raises the key question of aggregation the data, and the basic requirements for respect for privacy to be respected. It raises the problem of the representativeness of the data aggregated in this manner. The question takes on all the more meaning when the data collected is used to produce publications for the general public that can influence operators' behaviour, as is in particular the case for most web testers. To address the issue of representativeness, two main and complementary areas are to be developed:

- the volume of data collected, the order of magnitude of which is highly dependent on the nature of the mechanism deployed: tens or hundreds of thousands, where hardware sensors are concerned; a few tens or hundreds of millions, as concerns web testers; a few billion when it comes to software developed on web pages such as Radar.

- the characterisation of the data collected (geolocalisation, access technology, modem, terminal, used in the measure - see next section)

In many cases, the mechanisms used to collect a large volume of data do not make it possible to control, with any greater degree of detail – or characterise – the user environment, and *vice versa*.

Characterising the user environment

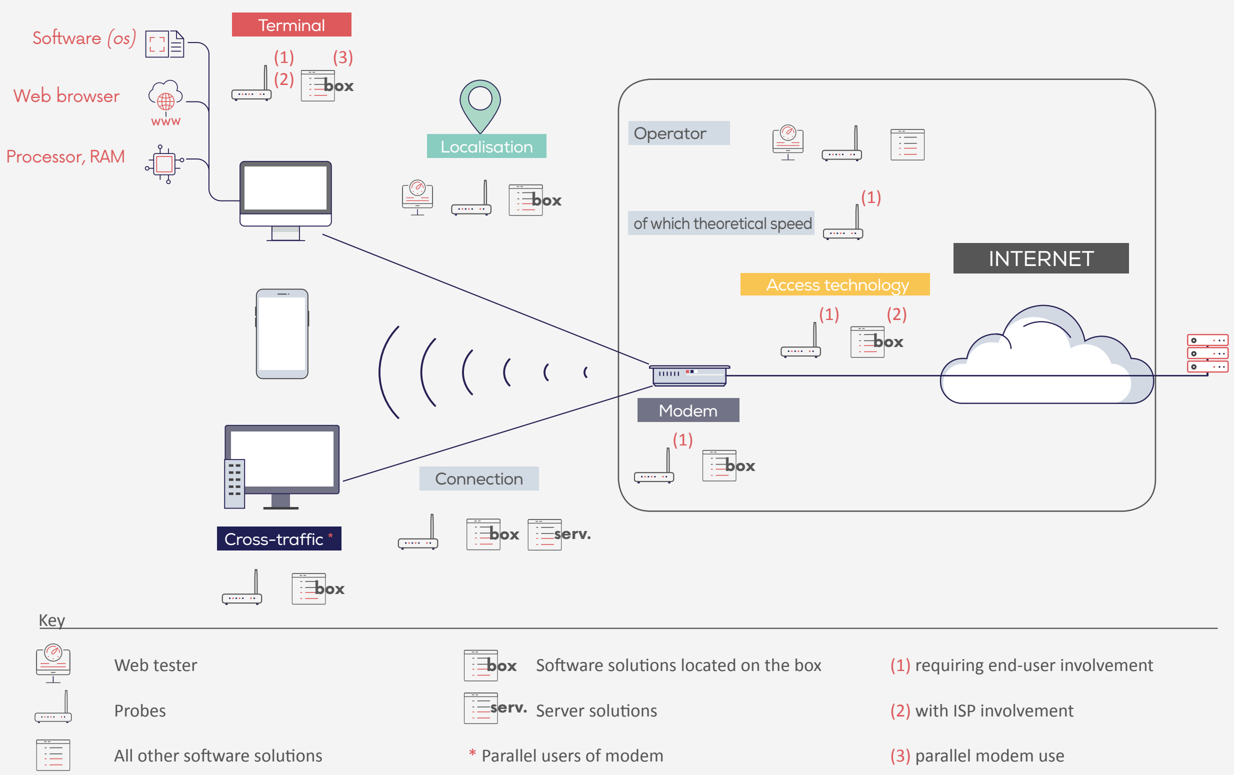
The term “user environment” covers a range of parameters some of which are more difficult to identify than others. Detection is highly dependent on the type of tool used. For instance, hardware sensors and software agents on boxes are often more able to identify them than software agents deployed in web pages or online testers.

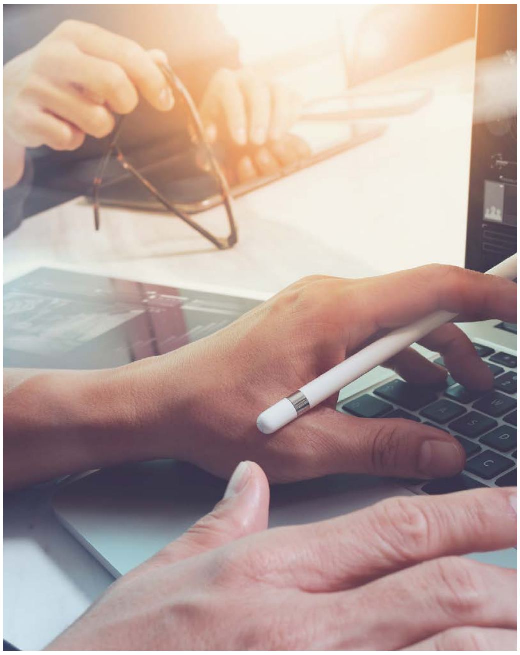
The operator and user location can both be detected by all tools, correlating the IP address of the tester and the existing databases. Other parameters are far more sensitive to identify: access technology (xDSL, coaxial cable, fibre, but also satellite); box-to-terminal connectivity (Wi-Fi, RJ45 cable, etc.); use of access by different terminals in parallel (cross-traffic); offer characteristics (theoretical speed), modem, terminal (web browser, software, processors, RAM).

In some cases, for the hardware sensors, the user’s environment settings are not detected but are fixed. For example, in the Case on IT set-up, the terminal features and its connection to the modem (cable or Wi-Fi) are pre-determined as the terminal is the MedUX probe.

It sometimes also happens that detection requires end-user participation (reporting questionnaire) or involvement on the part of the access provider (databases). In itself, all tools could therefore escalate this information if they requested it from the user and if the latter were able to respond reliably

// Characterising **the user environment**





to their request. However, because of their model, certain tools such as probes have much more direct access to end users (which, as a downside, are fewer in number) and do report this information.

Some tools focus in particular on the home network and the considerable deterioration in performance resulting from a Wi-Fi connection. This is in particular true of Case on IT, Gemalto, Ookla – which now makes it possible to launch Speedtest from one terminal to another – and ASSIA – which allows measurement of both the box to a test server and box to one or more terminals.

More broadly, a quantified evaluation of the often considerable impact of the various parameters listed – use of an obsolete version of the web browser or operating system, parallel login uses, etc. – would be greatly beneficial for the entire metrology ecosystem as well as for end users.

Mapping study conclusion

The crowdsourcing-based quality measurement ecosystem for fixed services is already very broad, and the diversity of approaches and models promising. However, significant work on the part of the entire community – ISP, measurement providers, academics, civil society, regulatory authorities, international bodies, etc. – remains to be done, particularly around the following topics:

- Sharing best practices in measurement methodologies;
- Characterising the user environment;
- Improving statistical representativeness (panel and number of measures);
- Fighting fraud;
- Developing usage indicators;
- Ensuring the trust-worthiness and impact of publications aimed at the general public.

As regards user environment control, Arcep invites in particular:

- measurement service providers to develop solutions to identify the various parameters of the user environment and incorporate them into their reporting;
- ISPs to raise their customers' awareness of the simple means available to them to optimise their network performance;
- academics to precisely quantify the impact of different user environment parameters on network performance.

Arcep also encourages the ecosystem to explore the avenues mentioned over the course of discussion exchanges with market players in order to assess their benefits and feasibility. Among the ideas suggested, it sees the implementation of random test exercises as a means of deterring fraud, and the opening of box APIs or certain operator databases to private players (measurement service providers) or public entities (regulators) to facilitate the identification of the user environment as particularly worthy of attention.

In this context, the Authority will act as a facilitator and trusted third-party to unite the community over time and stimulate the ecosystem's work around topics of general interest.

b) *Comparison of the measurement results obtained from several online speedtests*

In order to inform its thinking as it moves to crowd-sourcing, the Authority carried out a study designed to analyse the indicators measured by several popular online speedtests.

- Akostest - provided by Slovenian regulator AKOS: https://www.akostest.net/en/
- Journal du net (JDN) : http://www.journaldunet.com/test-connexion/
- M-lab's Network Diagnostic Test (NDT): https://www.measurementlab.net/tools/ndt/
- Netztest - carried out by Austrian regulator RTR: https://www.netztest.at/en/
- nPerf : https://www.nperf.com/fr/
- Ookla Speedtest: http://www.speedtest.net/fr/settings
- 01-net, put in place by ip-label: http://5g-token.col.ip-label.net/html/

Test procedure and protocol

The study was carried out over a two-week period on two test sites located in Paris and La Garenne-Colombes, and initially set up as part of the Arcep observatory on the quality of service of fixed networks. Through these dedicated lines, the various characteristics of the user environment were totally under control. The measurements were carried out directly from the boxes via Microsoft Internet Explorer 11 over the lines available on the test sites - long ADSL lines, cable (30 Mbps and 100 Mbps) and fibre, from Bouygues Telecom, Free, Orange and SFR. For each tool and over each line, data were collected on upload speeds, download speeds and latency [26].

Most tools are updated regularly. For example, Ookla launched a new version of its tool while the tests were being carried out, and Netztest (the RTR's tool) will be updated mid-2017. M-Lab hosts the measurement developed by the recently updated Internet2 consortium to support HTML5 testing. Moreover, it isinteresting to note that the Akostest, Netztest and 01-net speedtests are based on the same technology and methodology, which was developed by RTR, the only difference being the server testers and certain configurable elements.

[26] Whenever available (all performance testers except *Journal du net*).

As stated in the previous study, the location of the test servers has a significant impact on the results. All the testers choose a server by default based on an algorithm which is specific to each tooland often tends to minimise latency and/or maximise the speed achieved. When the tool has only deployed one server, this server is automatically chosen by default. The state of the network or the servers deployed at the time of the test may justify that the default test server is not always the same over time even though the test site is the same - this is, in particular, the case for the nPerf and Ookla performance testers.

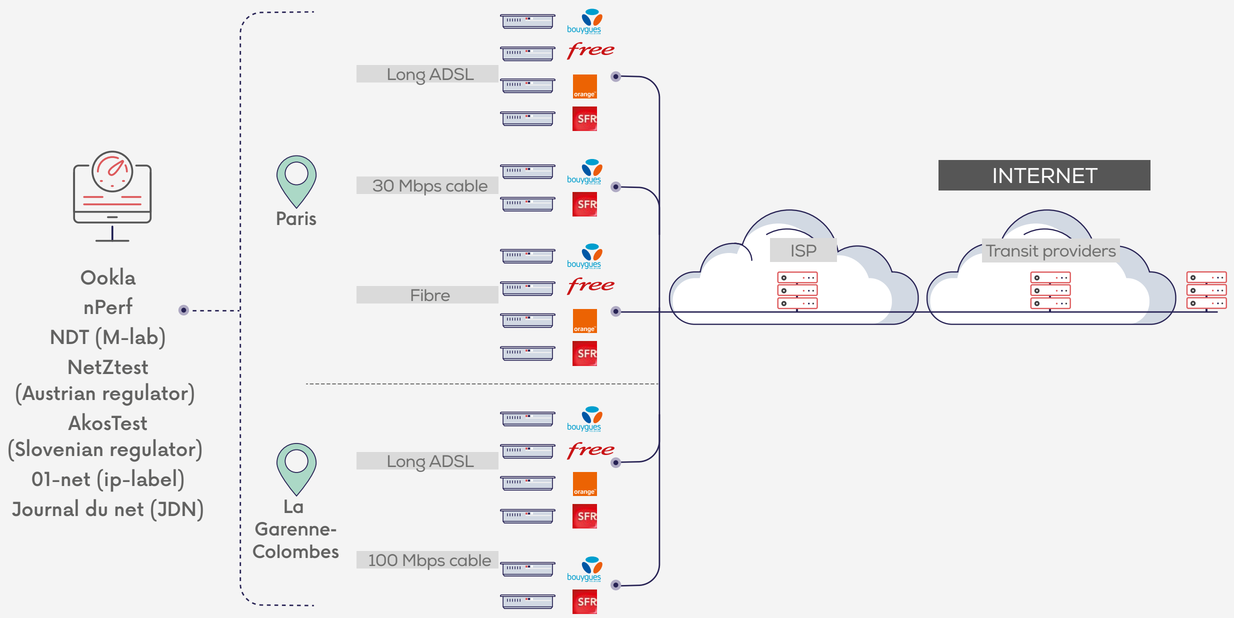
These two testers allow users to choose which test servers will be used to start the test. As can be seen in the following table, various locations were chosen in order to compare data from the largest possible number of tools by isolating the impact of the location of the test server. When the testers allowed for this, some test servers were chosen within the ISP network ("Bouygues Telecom", "Free", "SFR", "Orange") in order to analyse the possible impact on the results.

First Analysis

The Authority implemented a first level analysis. Some of the important initial findings of which are presented below. A more detailed analysis designed to show the reasons behind these findings – in the light of the measurement methodologies for example – still needs to be carried out. Subject to the agreements of the various measurement providers, the figures collected during this study will be examined over the course of workshops with all the stakeholders and may set in motion the Authority's co-construction approach.

As presented on the following graph, the median download speeds averaged over all the ISPs and obtained over fibre lines to the default test server vary significantly depending on the tool chosen. The lowest average (165 Mbps) and the highest average (901 Mbps) vary by a factor of more than 5. The values of the speeds presented by ISP (not averaged) show the same dispersion between the various tools. However, the classification of the four ISPs by download speeds over fibre remains

// Study **device**



// Test **servers**

	France								Europe				International	
Pattern generator location	Ile-de-France	Lyon	Strasbourg	Bouygues Telecom	Free	Orange	SFR	Other	Austria	Slovenia	Ireland	Other	United States	Thailand
Ookla	D	S	S	S	S	S		D	S	S	S		S	S
nPerf	D	S	S			S	S	D	S	S	S	D	S	S
NDT	D													
01-net	D													
netZtest									D					
AkosTest										D				
JDN	D													

Source: Arcep

Légende D : Default test server
S : Selected test server

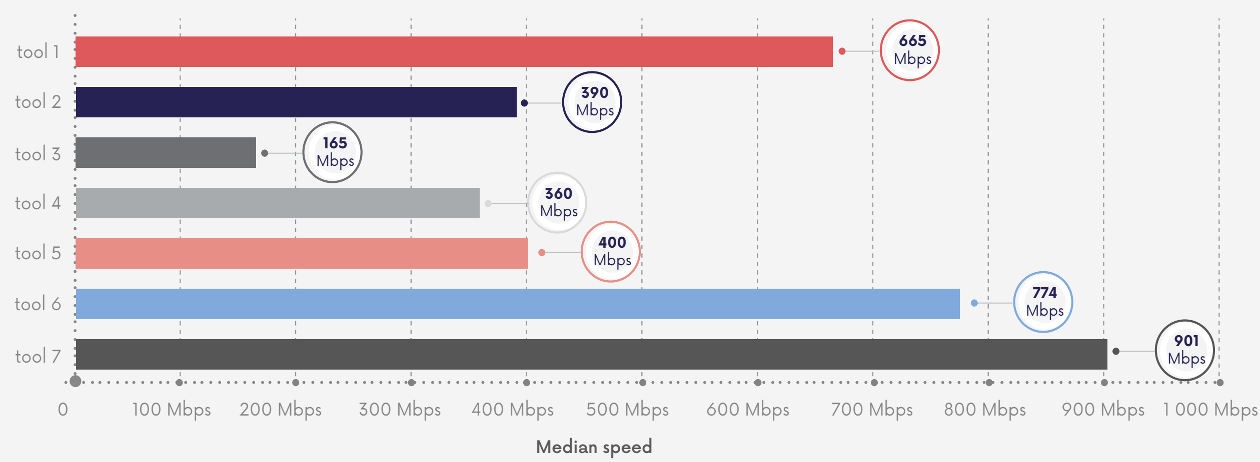
relatively stable - five of the seven tools show the same ranking, whereas the other two invert two ISPs.

The variation of the absolute values of the upload speeds over fibre is also significant - a factor of 8 is noted between the average of the median speeds of the speedtest with the lowest values and that of the one with the highest values. Contrary to the download speed, the order of the ISPs by upload speeds is not the same between the tools.

The spreads in upload speeds and download speeds observed over cable and ADSL are less than on fibre (approximately 20%). Although, as with fibre, the ISP rankings by download speed is relatively stable over ADSL, the ISP rankings by upload speeds differs depending on which tool is used. On cable, the ISP classification ranking by upload speed and by download speeds differs depending on the tool used.

// **Median download speed** depending on tool

Configuration: fibre to the home, all operators combined, test server by default



Source: Arcep

© D.R.

The considerable difference in upload and download speeds over fibre is explained in part by the location of the test server chosen by default - the further away it is, the slower the speeds will be.

Variations can also be identified with test server located in similar places. Thus, when they are at locations distant from the test site (in Europe or elsewhere in the world), the spread remains of the same order of magnitude. On the other hand, if the test server are located in France, the difference in speed is significantly less (in the region of 30%).

The choices regarding methodology appear to be an important factor. Indeed, when the comparison is limited to tools using a similar technology, the speeds measured are significantly closer.

Finally, as expected, test server located within an ISP's network often appear to advantage the host ISP – to the detriment, sometimes notable, of the other ISPs. The host ISP improves sometimes by up to two places in the ISP rankings by download speeds over fibre for example.

In addition to the average values, it is also useful to examine *ad hoc* values. Indeed, the measurements generated using certain tools show significant short-term variations.

3.1.4 Undertakings at the European level: on the road to a common measurement tool for fixed QoS

The quality of Internet access service is one of the priorities of many international regulators. It is also the focus of numerous undertakings at the European level, in which Arcep is deeply involved.

The European Commission, via its Directorate-General for Communications Networks, Content & Technology (DGConnect), launched the ambitious broadband mapping project in early 2016 (broadband mapping project). Its aim is to produce an online tool that centralises data from all public and private initiatives measuring the coverage and quality of fixed and mobile services from 31 European Union and European Economic Area countries. The main challenge lies in combining the different datasets into groups that use uniform and comparable methodologies. For this purpose, the Commission is working in close conjunction with BEREC.

As to BEREC, it continues its work on quality of service as part of its working group on Net neutrality, the two topics being closely intertwined. The quality of service workstream is divided into two sub-groups (see diagram page in section 3.4.2).

First of all, BEREC plans to publish a report that will help in developing a common methodology for measuring the quality of service and proposing methods for detecting possible traffic management practices within Internet access. In this context, as suggested by the open Internet regulation, BEREC offers avenues for certifying a performance monitoring mechanism for Internet access service (Art. 4.4 of the Regulation). It would enable any consumer to verify the actuality of the

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄八

Arcep State of Internet in France 2018: A high quality Internet to support innovation

A HIGH QUALITY INTERNET TO SUPPORT INNOVATION

Jonathan ARDOUIN,
Country Manager, France, **KRY**

Remote medical consultations are increasingly prevalent across Europe. They are helping to overcome permanent care issues and medical deserts that France also has to deal with. This new channel of care is still only nascent in France, as the 2018 Social Security Financing Act has only just enshrined the ability to be reimbursed for remote medical consultations.

KRY is the leading provider of video medical visits in France. Now in our third year of operation, we conduct close to 3% of all first aid consultations *via* video in Sweden, a country where video consultations are already a common practice and reimbursed by national health insurance. The hindsight we have gained from our experience in Sweden proves that video is the best channel – better than the phone or sending photos – for remote consultations, and can guarantee the same quality as an in-person visit. It allows the medical practitioner to establish ties with the patient and confidently make a diagnosis.

"TO BENEFIT FULLY FROM VIDEO CONSULTATIONS, A FAST AND STABLE INTERNET CONNECTION IS CRUCIAL FOR BOTH THE PATIENTS AND THE DOCTORS. TODAY IN FRANCE, IT IS COMMON FOR THIS TYPE OF CONSULT TO END ON THE PHONE BECAUSE THE INTERNET CONNECTION IS NOT FAST ENOUGH."

The one proviso, however, is that the video needs to be of high enough quality to allow the doctor to identify the patient's visible symptoms with certainty. To benefit fully from video consultations, a fast and stable Internet connection is crucial for both the patients and the doctors. Today in France, it is common for this type of consult to end on the phone because the Internet connection is not fast enough. And this even in major cities, and with users who have a "high-speed" connection.

The implication then is that a poor quality Internet service equals lost opportunities for patients: in areas where connections are too slow, patients will be deprived of rapid access to care. Common pathologies, which can easily be diagnosed *via* video, will need to be rerouted to already overtaxed physical channels (doctors' office, clinics, A&E). Having a high quality fixed Internet service is thus vital when it comes to telemedicine.

© KRY

2. AN INNOVATIVE CO-CONSTRUCTION APPROACH

On 19 January 2016, Arcep presented the conclusions of its strategic review, and announced the implementation of data-driven regulation, a greater push for co-constructed regulation, and Arcep's development around a role of neutral expert on digital issues.

The work devoted to Internet quality of service is fully in line with this new modus operandi.

In this data-driven approach to regulation, Arcep wants to use information regarding quality to stimulate competition that is not based solely on price, but also on the quality of the services being sold, with a view to monetising network investments.

To be both more efficient and more relevant, Arcep is thus seeking to co-construct this regulation:

- On the one hand, with "the crowd" by giving every citizen the power to become a mini-regulator. This was the impetus behind the launch of the "J'alerte l'Arcep" platform in October 2017, through which any consumer can report problems with their Internet access to Arcep (see inset). In addition to reporting, users can also input the results of QoS tests performed on their line using crowdsourced tools, and so contribute to the data used when publishing benchmarks of ISPs' performance;
- and, on the other, through a partner-centric approach with the ecosystem's stakeholders, for both the reporting and testing aspects referred to above. In the area of reporting, in addition to launching its own platform Arcep is examining the possibly of initiating a data-sharing scheme with consumer protection advocates[12]. This "unbundling" of the reporting process could help dismantle existing silos, and drive a better, collective understanding of the issues at hand. Arcep's partner-centric approach to crowdsourced testing is described below.

Alongside these co-construction efforts, Arcep is working to develop its own tools for collecting measurements that can enhance the data from its partners' third-party tools. These projects are detailed in Section 3 of this Chapter.

FYI

J'alerte l'Arcep

Launched in October 2017, the *"J'alerte l'Arcep"* platform is available to any citizen wanting to report an actual problem encountered with their mobile Internet, fixed Internet or postal services. The platform has logged 22,500 reports since it first launched. Of these reports, 68%* **concern a quality or availability issue with fixed or mobile services**. And, among them, two thirds concern the fixed market, and one third the mobile market.
This valuable feedback helps fuel the work that Arcep is doing on quantifying and identifying the problems that users are encountering, to then steer its actions towards the most appropriate solutions possible. It is on issues relating to Internet quality of service that the co-construction approach, the work being done on the BEREC tool and the monréseaumobile (my mobile network) scorecard described in Chapter 1, Section 3 come fully into play.

* Percentage obtained through reports logged between October 2017 and May 2018.

2.1. Bringing together stakeholders

Up until the end of 2016, Arcep's scorecard on the quality of fixed services was based on a system operating in a controlled environment. This type of testing was abandoned in early 2017 – largely because the real-life situations encountered by users were not being properly represented – and replaced by a system that would be based on crowdsourced testing tools.

The 2017 report on the state of the Internet in France presented the findings of the two studies that initiated the co-construction approach: the map of the ecosystem of the tools available in the marketplace, and a comparison of the results of different online testing tools. This report had stressed the need for a concerted community effort on several top priority issues. Arcep has since launched six courses of action as a direct result.

[12] In accordance with existing regulation, notably regarding data privacy.

PARTICIPANTS: A RICH AND VARIED ECOSYSTEM

Source: Arcep

To bring them to completion, **Arcep is acting as a neutral expert that brings the community together and fosters the work being done on matters of general interest.** These initiatives were carried out in collaboration with a wide spectrum of stakeholders from the crowdsourced metrology ecosystem[13] :

- testing tools: ASSIA, Case on IT (medUX), Cedexis, Directique, Ip-label, Gemalto, M-Lab, Ookla, nPerf, QoSi, SamKnows, V3D;
- ISPs: Bouygues Telecom, Free, Orange, SFR;
- academia and R&D: CNES, Inria;
- consumer protection organisations: INC, UFC Que-Choisir, which have also developed their own tools.

[13] Arcep invites any players who are not listed and who would like to take part in the co-construction efforts to get in touch.

Alongside the working groups, Arcep also consulted with other national regulatory authorities (notably AGCOM, BnetzA, COMREG, Ofcom and RTR) to pool their experience in measuring the quality of fixed services.

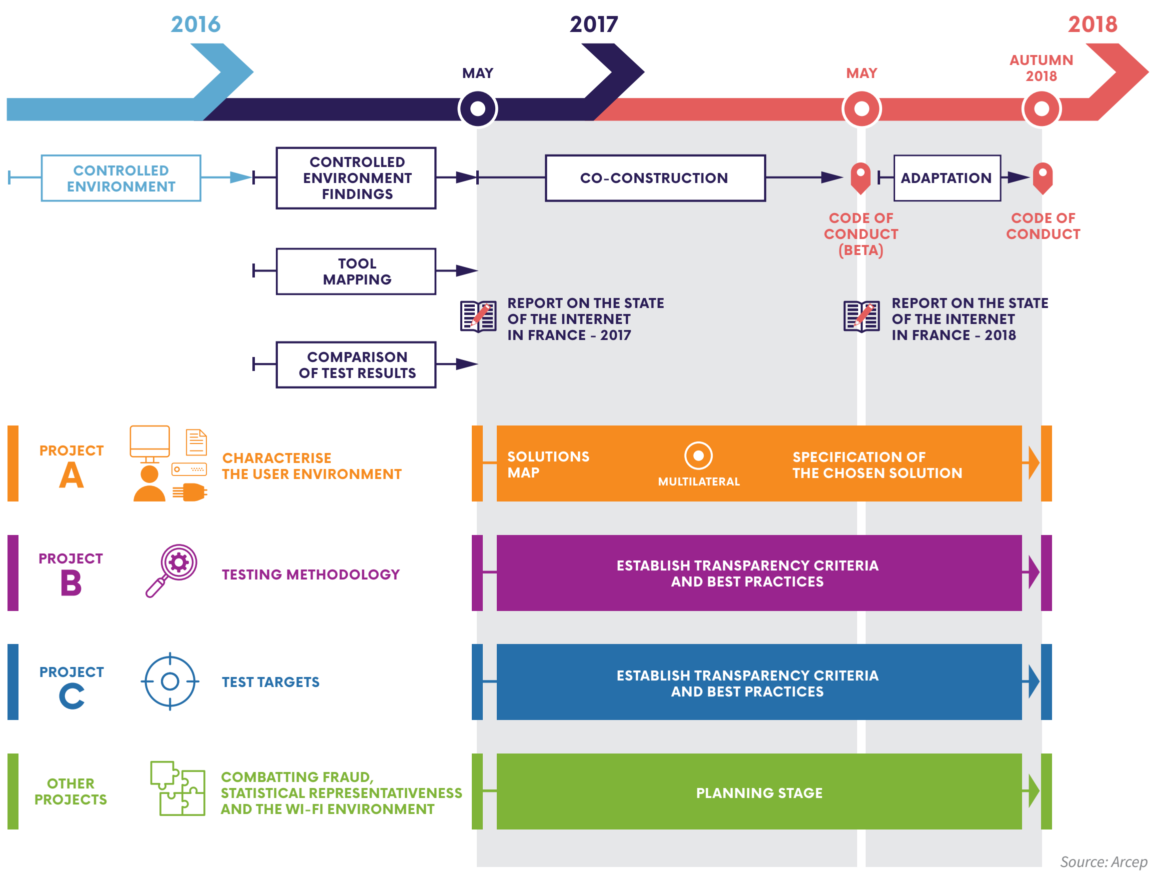
The goal that cuts across all of these initiatives is to enable the tools to meet consumers' and the Authority's needs as fully as possible, in terms of obtaining information on quality of service on the fixed and mobile Internet.

To be more specific, Project A seeks to address the technical problem raised in the previous section, namely the lack of characterisation of the user environment when measuring the quality of fixed services. The other projects (B, C and those currently in the planning stage) address the need for greater transparency that was also identified in the previous section. In particular, they seek to establish a "code of conduct" for testing tools. This future code of conduct concerns two aspects: first, inviting the tools to back the publication of their results with a clear explanation of the methodological choices made, so that any outside party is able to understand the potential differences observed between tests performed with different tools. Second, to set out the best practices that are vital to obtaining reliable measurements. Although most of the choices that have been made have merit, some practices do seem more questionable, and warrant being modified.

The first version of the code of conduct will be published before the end of 2018. And it will evolve over time: every year, in theory, Arcep will publish successive, continually improved versions, which include not only changes to Projects A, B and C, but also the fruit of the projects that are currently in the planning stage.

A beta version of the maiden code of conduct can be found in Annex 1. Stakeholders are heartily encouraged to share any remaining feedback on the matter with Arcep before 15 July 2018.

CO-CONSTRUCTION APPROACH



Source: Arcep

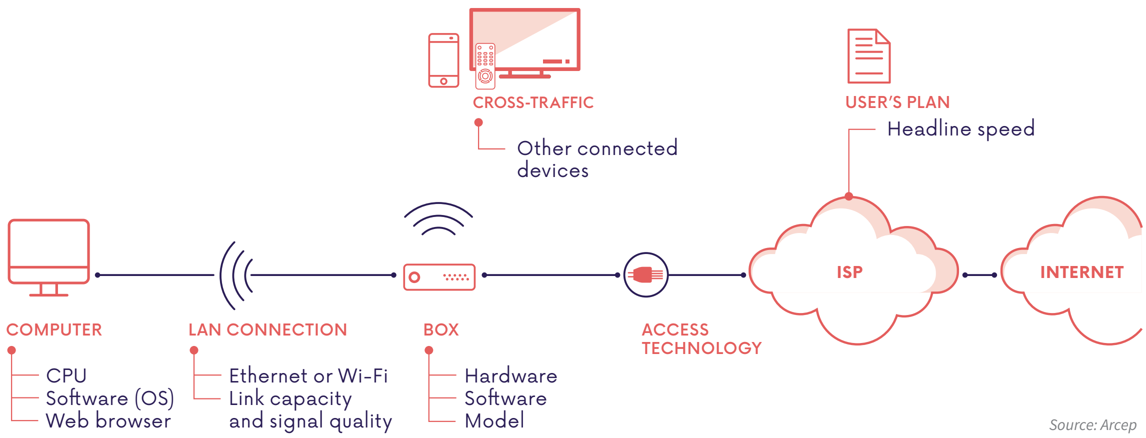
2.2. Project A: Characterising the user environment

The project dedicated to characterising the user environment on a fixed line, and notably the technology being used, has a dual purpose: first, it is vital to being able to create a truly relevant scorecard for consumers and, second, it is of significant value when establishing an accurate diagnosis of a quality of service issue. For instance, it is important to know whether a poor connection is due to the ISP's access network, the Wi-Fi network's quality or the simultaneous use of other connected devices on the local network when performing the test.

The following diagram recaps the main properties of the user environment that will influence the test results.

The current characterisation of the different elements varies depending on the type of testing tool being used. Some hardware probes[14] are, for instance, capable of testing a LAN[15] connection and even estimating cross-traffic[16] on the local network. On the flipside, while it is true that web testers[17] can be rapidly deployed on a large scale, they are only able to detail a small number of elements (web browser used, etc.).

THE USER ENVIRONMENT:
A SET OF CHARACTERISTICS WITH HIGH IMPACT ON MEASUREMENTS

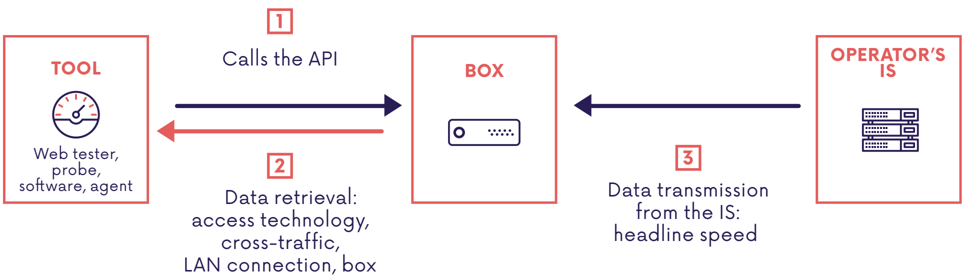


Source: Arcep

[14/15/16/17] See lexicon.

This project centred around the work coordinated by Arcep involving testing tools, ISPs and academia. The community began with an exploratory phase during which seven solutions attempting to satisfy the requirements were examined. One proposed path was to characterise the measurements through a questionnaire completed by the person performing the test, and more or less guided by the information given ahead of time by the ISPs (e.g. list of plans available for a given technology) or by an API[18] deployed between the tools and ISPs' information systems (IS). At this stage in the discussions, it appears that another solution may seem to offer the best compromise between exhaustiveness, reliability, security and development costs for most stakeholders. Arcep thanks them for their dynamic and constructive contributions.

**AN "ACCES ID CARD" API
FOR CHARACTERISING THE USER ENVIRONMENT**



Source : Arcep

A diagram of the solution is presented above. When a test is performed, the tool (whether a web tester, hardware probe, software agent on a box, software that can be installed on a device) simultaneously sends a request to the "access ID card" API located on the tester's box 1. If the tool queries this API, the box will send it the characteristics of the line at the time of testing 2. Most of the information is available natively on the box: access technology, information on the LAN connection and the box and — for most ISPs – a WAN[19] port traffic counter that makes it possible to detect cross-traffic. Other properties, such as headline speed, are not available locally on the box but on the operator's IS: through another API, if the ISP transmits them to the box often enough to ensure that the information is always up to date 3. It should be noted that operators' IS – the system at the heart of their internal processes' operation which may not be very reactive – never interacts directly with the tools.

Moreover, this solution is invisible to the person performing the test, and in no way diminishes the user experience. Further details on the solution's technical features can be found in Annex 2.

This ambitious project should thus enable the tools used to test fixed networks to achieve degrees of characterisation that are virtually equivalent to those obtained natively by mobile apps – which are already capable of identifying the access network (2G, 3G or 4G), for instance, and the strength of the signal since they are tied directly to the mobile operating system (OS) and there is no intermediary between the device and the network – contrary to a fixed network where the connection is supplied through a box.

By establishing API specifications and the list of tools authorised to access them, in concert with stakeholders, Arcep will continue to create the environment of trust needed for collaboration with the different players. Taking the State-as-a-platform approach to the fullest extent, Arcep will thus fulfil its mandate to inform consumers while leaving it up to its partners to develop innovations based on the information that has been collected.

[18/19] See lexicon.

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄九

IMDA Code of Practice for television broadcast standards

BROADCASTING ACT
(CHAPTER 28)

Code of Practice for Television Broadcast Standards

In exercise of the powers conferred by section 6 of the Broadcasting Act (Cap. 28), the Media Development Authority of Singapore hereby issues the Code of Practice for Television Broadcast Standards:

TABLE OF CONTENTS

PRELIMINARY

1.1 Citation and Commencement

This Code is issued pursuant to section 6 of the Broadcasting Act (Cap. 28). It may be cited as the Code of Practice for Television Broadcast Standards and shall come into force on 4 May 2015.

1.2 Purpose of this Code

(a) The Broadcasting Act and the Media Development Authority of Singapore Act (Cap. 172) ("MDA Act") make it the duty of the Media Development Authority of Singapore ("MDA") to exercise licensing and regulatory functions in respect of television ("TV") broadcasting services and to develop codes of practice relating to broadcast and technical standards relating to media and TV broadcasting services.

(b) The purpose of this Code is to ensure that licensed nationwide TV Licensees in Singapore meet the requirements of high standards of technical quality and reliability of licensable TV broadcasting services.

1.3 Legal Effect of this Code

(a) In accordance with section 1.4, every nationwide TV Licensee to which MDA grants a broadcasting licence under section 8 of the Broadcasting Act ("Licensee") must comply with the applicable provisions of this Code.

(b) The obligations contained in this Code are in addition to those contained in the MDA Act, the Broadcasting Act, as well as other regulations, broadcasting licences or codes of practice issued by MDA. To the extent that any provision of this Code is inconsistent with the terms of MDA Act or the Broadcasting Act or any regulations under the Broadcasting Act, the provisions of the MDA Act or the Broadcasting Act and any regulations under the Broadcasting Act shall prevail. To the extent that this Code is inconsistent with the provisions of any broadcasting licences or codes of practice issued by MDA, the terms of this Code shall prevail. If any provision of this Code is held to be unlawful, all other provisions will remain in full force and effect.

(c) Where the applicable broadcast standards specified in this Code are not met by any Licensee, MDA will assess each case individually and may, pursuant to section 16 of the Broadcasting Act, issue directions in writing to the Licensee requiring the Licensee to make necessary improvements to achieve the required standards or to take any other action with regard to broadcast standards necessary in order to comply with the provisions of this Code.

(d) Pursuant to section 12(1) of the Broadcasting Act, the MDA may cancel or suspend a licence for such period as MDA thinks fit and/ or impose financial penalties on a Licensee that contravenes any provision of this Code.

1.4 Application of this Code to Licensees

(a) Unless otherwise stated, the provisions of this Code shall apply to all nationwide TV Licensees. For avoidance of doubt, this Code shall not apply to niche TV Licensees.

1.5 Definitions

(a) In this Code, unless the context otherwise requires:-

"Analogue cable TV service" means a licensable TV broadcasting service comprising analogue television signals delivered using coaxial cable transmission technology;

“Analogue terrestrial TV service” means a licensable TV broadcasting service comprising analogue TV signals delivered using over-the-air broadcast transmission technology;

“Cable TV service” means a licensable TV broadcasting service comprising analogue or digital TV signals delivered using coaxial cable transmission technology. It comprises Analogue cable TV service and Digital cable TV service;

"Digital cable TV service" means a licensable TV broadcasting service comprising digital TV signals delivered using coaxial cable transmission technology;

"Digital terrestrial TV service" or “DTV service” means a licensable TV broadcasting service comprising digital TV signals delivered using over-the-air broadcast transmission technology;

“Free-to-air Terrestrial broadcast TV service” means an unencrypted terrestrial licensable TV broadcasting service comprising analogue or digital TV signals delivered using over-the-air broadcast transmission technology that viewers can receive without having to pay a Subscription fee;

“Free-to-air TV service” means an unencrypted terrestrial licensable TV broadcasting service that viewers can receive without having to pay a Subscription fee;

“Indoor reception” means reception of over-the-air broadcast TV signals within a building using a portable antenna;

"Internet Protocol TV Service" or “IPTV service” means a licensable TV broadcasting service comprising digital TV signals delivered using internet protocol (“IP”) based broadband technology. The service is delivered over a closed network using infrastructure that is specifically configured to receive an IPTV channel, or channels, from a particular broadband network service provider;

“Licence” means a licence granted under Section 8 of the Broadcasting Act, and “Licensee” shall be construed accordingly;

"Managed transmission TV service" means a licensable TV broadcasting service (comprising (i) Terrestrial broadcast TV service; (ii) Cable TV service; and (iii) IPTV service) delivered using a transmission network which the Licensee has control over the quality of service delivered to the viewer because the network is owned, maintained and/or operated by the Licensee, or by third parties hired and/or contracted by the Licensee;

“Must carry channels” mean the free-to-air nationwide terrestrial TV channels provided on subscription nationwide TV services as directed by the Authority;

“Person” refers to any individual, any company, partnership or association, and any body of persons, corporate or unincorporated;

“Subscriber” mean any person who has requested the Licensee for the reception and/or display of any programme carried on the Service and has agreed to pay the fees and charges which may be levied by the Licensee;

“Subscription fee” means any form of consideration;

“Subscription TV service” means a licensable TV broadcasting service made available to viewers only upon the payment of a Subscription fee, and “subscription nationwide TV service” shall be construed accordingly;

“Terrestrial broadcast TV service” means a licensable TV broadcasting service comprising analogue or digital TV signals delivered using over-the-air broadcast transmission technology. It comprises Analogue terrestrial TV service and DTV service; and

“Viewer” means any person who receives any licensable TV broadcasting service provided by a Licensee.

2 SERVICE COVERAGE REQUIREMENTS

2.1 Introduction

(a) Section 2 "Service Coverage Requirements" sets out the broadcast standards in relation to service coverage performance that all Licensees must comply with where applicable.

2.2 Terrestrial broadcast TV services

2.2.1 Application

(a) The provisions set out in this sub-section shall apply to Licensees providing free-to-air Terrestrial broadcast TV services.

2.2.2 Obligations for Licensees to meet specified requirements

(a) A Licensee providing free-to-air Terrestrial broadcast TV service shall ensure that such service is simultaneously receivable in at least 98% of Singapore's geographical area (including outlying islands).

(b) A Licensee providing free-to-air DTV service shall ensure that the indoor reception of such service is enabled for at least 98% of all residential properties in Singapore.

(c) A Licensee shall use its best efforts to provide solutions for any Housing and Development Board (HDB) residential property, or recommend solutions to private residential properties, that are unable to receive such services.

2.2.3 Compliance with obligations

(a) A Licensee shall provide MDA with a description of its procedures for ensuring that the required service coverage is achieved. It shall also carry out routine assessments of the coverage of its service(s) and undertake appropriate measures to address any viewer complaints[1] or feedback on coverage issues.

[1] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

(b) A Licensee shall submit to MDA quarterly reports, in the format as specified by the Authority, on viewer complaints[2] relating to the coverage of its service(s) that were received over the past three (3) calendar months, within one (1) month from the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.[3]

2.3 Cable TV services

2.3.1 Application

(a) The provisions set out in this sub-section shall apply to Licensees providing Digital cable TV services.

2.3.2 Obligations for Licensees to meet specified requirements

(a) A Licensee providing Digital cable TV service shall ensure that such service will be made available to any person in Singapore who makes a request to the Licensee for the connection to the Licensee's telecommunication system for the reception of such services where the Licensee has rolled out its network. For the purposes of this sub-section, "roll out" means the installation of the Licensee's telecommunication system, whether in, on, under or otherwise through any existing or future public road, lane or street.

2.3.3 Compliance with obligations

(a) A Licensee shall submit to MDA quarterly reports, in the format as specified by the Authority, on viewer complaints[4] relating to service coverage that were received over the past three (3) calendar months, one (1) month after the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.[5]

2.4 IPTV services

2.4.1 Application

(a) The provisions set out in this sub-section shall apply to Licensees providing IPTV services.

2.4.2 Obligations for Licensees to meet specified requirements

---

[2] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

[3] Complaints which arise from issues related to viewer premise equipment may be excluded from this report.

[4] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

[5] Complaints which arise from issues related to viewer premise equipment may be excluded from this report.

(a) A Licensee providing IPTV service shall ensure that such service will be made available to any person in Singapore who makes a request to the Licensee where the Next Generation Nationwide Broadband Network (Next Gen NBN) has been rolled out.[6]

2.4.3 Compliance with obligations

(a) A Licensee shall submit to MDA quarterly reports, in the format as specified by the Authority, on viewer complaints[7] relating to service coverage that were received over the past three (3) calendar months, within one (1) month from the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.[8]

[6] The Licensee may provide its IPTV service solely on the network owned, maintained, and/or operated by the said Licensee or by the third parties that the said Licensee may hire and/or contract.

[7] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

[8] Complaints which arise from issues related to viewer premise equipment may be excluded from this report.

3 TV SIGNAL STRENGTH REQUIREMENTS

3.1 Introduction

(a) Section 3 “TV Signal Strength Requirements” sets out the broadcast standards in relation to the transmission requirements for TV signal strength that Licensees must comply with where applicable.

3.2 Terrestrial broadcast TV services

3.2.1 Application

(a) The provisions set out in this sub-section shall apply to Licensees providing free-to-air Terrestrial broadcast TV services.

3.2.2 Obligations for Licensees to meet specified requirements

(a) A Licensee providing free-to-air Analogue terrestrial broadcast TV service shall ensure that the outdoor TV signal strength within the required coverage area for such services shall not fall below the minimum signal strength of 65 dBµV/m for Band IV and 70 dBµV/m for Band V, as specified in Recommendation ITU-R BT.417[9]. These requirements apply to the median field strength at a height of 10m above ground level.

(b) A Licensee providing free-to-air DTV service shall ensure that the minimum indoor TV signal strength within the required coverage area for such services shall be in accordance with Recommendation ITU-R BT.2254 based on Singapore transmission parameters.[10]

3.2.3 Compliance with obligations

(a) A Licensee shall provide MDA with a written description of its procedures for ensuring that the required standards of TV signal strength is achieved within the required coverage area.

(b) A Licensee shall also carry out and report annual field measurements at sample locations in the areas between transmitter locations or as directed by MDA and provide such measurement reports to MDA as and when required by MDA.

---

[9] ITU-R Recommendation BT.417.

[10] ITU-R Recommendation BT.2254: Frequency and network planning aspects of DVB-T2.

(c) A Licensee shall submit to MDA quarterly reports, in the format as specified by the Authority, on viewer complaints[11] relating to TV signal strength that were received over the past three (3) calendar months, within one (1) month from the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.[12]

3.3 Cable TV services

3.3.1 <u>Application</u>

(a) The provisions set out in this sub-section shall apply to all Licensees providing Cable TV services.

3.3.2 <u>Obligations for Licensees to meet specified requirements</u>

(a) A Licensee providing Digital cable TV service shall ensure that the TV signal strength for such service shall not fall below the minimum of 47 dBµV, as specified in the European Standard EN 50083-7[13].

(b) A Licensee providing Analogue cable TV service shall ensure that where the frequency range and service are 54-824 MHz TV, the signal strength shall not fall below the minimum of 60 dBµV (as specified in the European Standard EN 50083-7[14]).

(c) For the purpose of this section, the TV signal strength for Cable TV service refers to the minimum voltage level that must be present at each viewer premise's cable outlet.

3.3.3 <u>Compliance with obligations</u>

(a) A Licensee shall provide MDA with a written description of its procedures for ensuring that the required standards of TV signal strength is achieved.

---

[11] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

[12] Complaints which arise from issues related to viewer premise equipment may be excluded from this report.

[13] European Standard Series EN 50083-7. Cable networks for television signals, sound signals and interactive services. Part 7: System performance.

[14] European Standard Series EN 50083-7. Cable networks for television signals, sound signals and interactive services. Part 7: System performance.

(b) A Licensee shall submit to MDA quarterly reports, in the format as specified by the Authority, on viewer complaints[15] relating to TV signal strength that were received over the past three (3) calendar months, within one (1) month from the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.

[15] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

4 PICTURE AND AUDIO QUALITY REQUIREMENTS

4.1 Introduction

(a) Section 4 “Picture and Audio Quality Requirements” sets out the broadcast standards in relation to picture and audio quality that Licensees must comply with where applicable.

4.2 Terrestrial broadcast TV, Cable TV, IPTV services

4.2.1 Application

(a) The provisions set out in this sub-section shall apply to Licensees providing Terrestrial broadcast TV services, Cable TV services and/ or IPTV services.

4.2.2 Obligations for Licensees to meet specified requirements

(a) A Licensee providing Terrestrial broadcast TV service, Cable TV service or IPTV service shall ensure that

(i) ”live” programmes on such service shall achieve a picture and audio grade of 5 on the ITU-R 5-Point Quality Grading Scale as described in ITU–R BT.500[16] ; and

(ii) recorded programmes on such service shall achieve a picture and audio grade of 4 on the ITU-R 5-Point Quality Grading Scale as described in ITU–R BT.500[17].

(b) A Licensee shall ensure that the picture and audio of the transmitted programmes are accurately synchronised.

4.2.3 Compliance with obligations

(a) A Licensee shall provide MDA with a written description of its procedures for ensuring that the programmes on its service achieve the required standards of picture and audio quality.

(b) A Licensee shall also carry out routine assessments of the technical quality of its TV service.

(c) A Licensee shall, without any undue delay, attend to and handle, as well as provide proper avenues for the speedy resolution of viewer complaints or feedback relating to picture and audio quality.

[16] ITU-R Recommendation BT.500: Methodology for the subjective assessment of the quality of television pictures.

[17] ITU-R Recommendation BT.500: Methodology for the subjective assessment of the quality of television pictures.

(d) A Licensee shall submit to MDA quarterly reports, in the format as specified by the Authority, on viewer complaints[18] relating to picture and audio quality that were received over the past three (3) calendar months, within one (1) month from the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.[19]

(e) The Licensee shall retain a continuous recording of the broadcast version of all programmes transmitted on its service for a period of four (4) weeks from the date on which the programmes was broadcast.

(f) In the event of a pattern or trend of viewer complaints related to picture and audio quality, MDA may, in its discretion, launch an investigation into the service provided by the Licensee. Where an investigation is undertaken in this regard, the Licensee shall provide at MDA's request and without charge, recordings of the programme(s) or channel(s) in question. Such recordings are to be made of the transport stream after all encoding and multiplexing have taken place.[20]

4.2.4 Guidance notes

(a) A lower picture and audio quality grade may be justified for news inserts, actuality or historical material where it is not practicable to improve further the technical quality, or where low quality clearly forms part of the editorial intent of the programme.

[18] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

[19] Complaints which arise from issues related to viewer premise equipment may be excluded from this report.

[20] The purpose of this requirement is to address cases where Licensees persistently fail to meet the picture and audio quality standards, rather than short term failures due to outages. In general, short term impairment or degradation of picture and audio quality will be considered an outage rather than a breach of the picture and audio quality standards.

5 RELIABILITY REQUIREMENTS

5.1 Introduction

(a) Section 5 "Reliability Requirements" sets out the broadcast standards in relation to the reliability of services (measured in terms of channel and service availability to viewers) that Licensees shall comply with where applicable.

(b) For the purposes of Outage Reliability Index ("ORI") reporting, a distinction is made between Channel Availability and Network Availability:

   (i) "Channel Availability" refers to the time during which an individual programme channel is available as measured at the point of delivery into the transmission network.

   (ii) "Network Availability" refers to the time during which the transmission network is operable and not in a state of failure or outage.

(c) An outage is considered to have occurred when:

   (i) there is an absence of channel or service;

   (ii) there is an intermittent or persistent loss of audio or video for one or more channels, or

   (iii) there is significant degradation[21] of service to below a normal or acceptable level of quality.

(d) In relation to free-to-air DTV services, an outage is also considered to have occurred when:

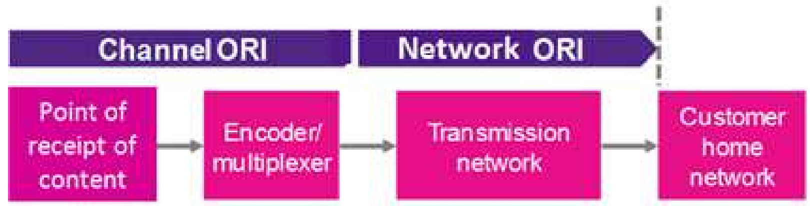
   (i) there is a breakdown of the main transmitter;

   (ii) there is a breakdown of a repeater; or

   (iii) there is a breakdown of a transposer/ gap filler.

[21] Non-exhaustive examples of "degradation of service" will include, without limitation, any or all of the following:
(a) Pixelation of pictures;
(b) Picture freezing;
(c) Audio synchronisation issues.

(e) Channel and Network availabilities for managed transmission TV services are measured by separate Channel ORI and Network ORI as shown in the diagram below:



(f) In this Code, "Channel ORI" means the measure of the reliability of a Licensee's playout system from the point of receipt of content or headend to the point at which the content is encoded and multiplexed into a transport stream for delivery over the transmission network.

(g) In this Code, "Network ORI" means the measure of the reliability of the transmission network deployed by the Licensee to deliver its service. The computation of Network ORI depends on transmission technology and network architecture deployed by the Licensee (See Guidance Notes in sub-section 5.2.4 / 5.3.4 / 5.4.4).

5.2 Terrestrial broadcast TV services

5.2.1 <u>Application</u>

(a) The provisions set out in this sub-section shall apply to Licensees providing Terrestrial broadcast TV services, unless otherwise stated.

5.2.2 <u>Obligations for Licensees to meet specified requirements</u>

(a) A Licensee providing Terrestrial broadcast TV service (other than free-to-air Terrestrial broadcast TV service) shall maintain a minimum monthly Channel ORI of 99.80% for each individual channel on such service.

(b) A Licensee providing free-to-air Terrestrial broadcast TV service shall maintain a minimum monthly Channel ORI of 99.90% for each individual channel on such service.

(c) A Licensee providing Terrestrial broadcast TV service (other than free-to-air DTV service) shall ensure that the minimum monthly Network ORI for the transmitter is maintained at 99.80%.

(d) A Licensee providing free-to-air DTV service shall ensure that:

(i) the minimum monthly Network ORI for the DTV network is maintained at 99.80%;

(ii) the minimum monthly Network ORI for the main transmitter is maintained at 99.80%; and

(iii) the minimum Network ORI for each repeater in its DTV network is maintained at 99.50%, averaged over the preceding six (6) months.

5.2.3 Compliance with obligations

(a) A Licensee shall monitor and submit to MDA monthly ORI reports on the Channel ORI and Network ORI results.

(b) The formula for computing Channel ORI and Network ORI for Terrestrial broadcast TV services is:

$$Outage\ Reliability\ Index\ (ORI) = (1 - R) \times 100\%\ ,$$

where $R = \frac{Total\ lost\ time\ for\ all\ outages\ in\ a\ calendar\ month}{Total\ number\ of\ broadcast\ hours\ in\ a\ calendar\ month}$.

(c) The formula for computing Network ORI for each repeater stipulated in 5.2.2(d)(iii) of this Code is:

$$Outage\ Reliability\ Index\ (ORI) = (1 - R) \times 100\%\ ,$$

where $R = \frac{Total\ lost\ time\ for\ all\ outages\ in\ the\ past\ 6\ months}{Total\ number\ of\ broadcast\ hours\ in\ the\ past\ 6\ months}$.

(d) In computing the Network ORI for the DTV network stipulated in clause 5.2.2(d)(i) of this Code, for outages that affect a localised area or a subset of viewers, the effective lost time shall be calculated using a normalisation factor (N) based on the Effective Radiated Power of affected transmitter(s) as a proportion of the total Effective Radiated Power of all transmitters in the DTV network. The formula is:

$$Effective\ lost\ time\ for\ outage = \ Normalisation\ factor\ (N) * Time\ lost\ due\ to\ outage\ ,$$

where $N = \frac{Effective\ Radiated\ Power\ of\ affected\ transmitter(s)}{Total\ Effective\ Radiated\ Power\ of\ all\ transmitters}$, and

Total Effective Radiated Power refers to the total Effective Radiated Power of all transmitters in the DTV network used for delivery of the free-to-air DTV service.

For outages that result in complete loss of service, $N = 1$. If an outage affects only certain transmitter(s) in the DTV network, then $N < 1$. For cases where $N < 1$, broadcasters shall provide description of how the normalisation factor is determined in the monthly ORI reports.

(e) A Licensee shall also record and report all outage incidents in the monthly ORI reports. The basic details required are date and time of outage incidents; description of incidents in terms of the network elements affected, the number of viewers or subscribers affected, the affected programmes or services and time taken to restore the service. The report on outage incidents shall be in the format as specified by the Authority.

(f) The Channel ORI and Network ORI results should take into account loss of video or sound or control data essential to view the services due to any cause under the control, either directly or through contract arrangements, of the Licensee. Outages that occur due to factors not under the control of the Licensee may be exempted from the computation of ORI results, although they should be recorded in the monthly ORI reports as stipulated in clause 5.2.3(e).[22]

(g) The loss of ancillary data and services such as subtitles is exempted from the computation of ORI results. For such cases, Licensees shall display an apology message on the affected programme(s) as soon as possible when the fault occurs.

(h) A Licensee shall submit to MDA quarterly reports, in the format as specified by the Authority, on viewer complaints[23] relating to outages and poor reception quality that were received over the past three (3) calendar months, within one (1) month from the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.[24]

5.2.4 Guidance notes

(a) Licensees shall submit to MDA monthly Channel ORI results for each free-to-air TV channel.

(b) The Network ORI shall be measured at the transmitter.

5.3 Cable TV services

5.3.1 Application

(a) The provisions set out in this sub-section shall apply to Licensees providing Cable TV services.

[22] Exemptions may include impairment due to external content source, equipment managed by viewers, planned maintenance, unscheduled interruptions to power supply, extreme or unforeseen weather conditions and sun outage(s).

[23] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

[24] Complaints which arise from issues related to viewer premise equipment may be excluded from this report.

5.3.2 Obligations for Licensees to meet specified requirements

(a) A Licensee providing Cable TV service shall maintain a minimum monthly Channel ORI of 99.80% for each individual channel on such service.

(b) A Licensee providing Cable TV service shall maintain a minimum monthly Channel ORI of 99.90% for each individual "must carry" channel[25] on such service.

(c) A Licensee providing Digital cable TV service shall ensure that the minimum monthly Network ORI for such service is maintained at 99.80%.[26]

(d) A Licensee providing Analogue cable TV service shall ensure that the minimum monthly Network ORI for such service is maintained at 99.80%.[27]

5.3.3 Compliance with obligations

(a) A Licensee shall monitor and submit to MDA monthly ORI reports on the Channel ORI and Network ORI results.

(b) The formula for computing Channel ORI and Network ORI for Cable TV service is:

$$Outage\ Reliability\ Index\ (ORI) = (1 - R) \times 100\%\ ,$$

where $R = \frac{Total\ lost\ time\ for\ all\ outages\ in\ a\ calendar\ month}{Total\ number\ of\ broadcast\ hours\ in\ a\ calendar\ month}$.

(c) In computing the Network ORI, for outages that affect a localised area or a subset of viewers, the effective lost time shall be calculated using a normalisation factor (N) based on the proportion of homes or viewers affected by the outage. The formula is:

$$Effective\ lost\ time\ for\ outage = \ Normalisation\ factor\ (N) * Time\ lost\ due\ to\ outage\ ,$$

where $N = \frac{Homes\ affected\ by\ outage}{Total\ connected\ homes}$, and

Total connected homes refer to number of homes receiving the TV service.

For outages which affect the entire service, $N = 1$. If an outage affects only certain homes or viewers, then $N < 1$. For cases where $N < 1$, broadcasters shall provide description of how the number of homes affected is determined in the monthly ORI reports.

[25] The "must-carry" channels are Channel 5, Channel 8, Suria, Vasantham, okto, Channel U and Channel NewsAsia.

[26] For the purposes of normalisation, "total connected homes" refers to the total number of subscribers to the Licensee's digital cable TV services.

[27] For the purposes of normalisation, "total connected homes" refers to the total number of homes connected to the Licensee's cable TV network which do not subscribe to the Licensee's digital cable TV services.

(d) A Licensee shall also record and report all outage incidents in the monthly ORI reports. The basic details required are date and time of outage incidents; description of incidents in terms of the network elements affected, the number of viewers or subscribers affected, the affected programmes or services and time taken to restore the service. The report on outage incidents shall be in the format as specified by the Authority.

(e) The Channel ORI and Network ORI results should take into account loss of video or sound or control data essential to view the services due to any cause under the control, either directly or through contract arrangements, of the Licensee. Outages that occur due to factors not under the control of the Licensee may be exempted from the computation of ORI results, although they should be recorded in the monthly ORI reports as stipulated in clause 5.3.3(d).[28]

(f) The loss of ancillary data and services such as subtitles is exempted from the computation of ORI results. For such cases, Licensees shall display an apology message on the affected programme(s) as soon as possible when the fault occurs.

(g) A Licensee shall submit to MDA quarterly reports, in the format as specified by the Authority, on viewer complaints[29] relating to outages and poor reception quality that were received over the past three (3) calendar months, within (1) month from the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.[30]

5.3.4 Guidance notes

(a) Licensees are not required to report the Channel ORI result for an individual channel if no outage is suffered in a calendar month. Outage incidents, if they occur, shall be reported to the MDA on a calendar month basis.

(b) The Network ORI shall be measured at the network nodes connected to the premises receiving Cable TV services. In the case where a nationwide outage had occurred for the entire service or channel(s) within the service, the total number of connected homes or subscribers of the service or channel(s) within the service shall be deemed to have experienced an outage.

---

[28] Exemptions may include impairment due to external content source, equipment managed by viewers, planned maintenance, unscheduled interruptions to power supply, extreme or unforeseen weather conditions and sun outage(s).

[29] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

[30] Complaints which arise from issues related to viewer premise equipment may be excluded from this report.

5.4 IPTV services

5.4.1 Application

(a) The provisions set out in this sub-section shall apply to Licensees providing IPTV services.

5.4.2 Obligations for Licensees to meet specified requirements

(a) A Licensee providing IPTV service shall maintain a minimum monthly Channel ORI of 99.80% for each individual channel on such service.

(b) A Licensee providing IPTV service shall maintain a minimum monthly Channel ORI of 99.90% for each individual "must carry" channel[31] on such service.

(c) A Licensee providing IPTV service shall ensure that the minimum monthly Network ORI for such service is maintained at 99.80%.

5.4.3 Compliance with obligations

(a) A Licensee shall monitor and submit to MDA monthly ORI reports on the Channel ORI and Network ORI results.

(b) The formula for computing Channel ORI and Network ORI for IPTV service is:

$$Outage\ Reliability\ Index\ (ORI) = (1 - R) \times 100\% ,$$

where $R = \frac{Total\ lost\ time\ for\ all\ outages\ in\ a\ calendar\ month}{Total\ number\ of\ broadcast\ hours\ in\ a\ calendar\ month}$.

(c) In computing the Network ORI, for outages that affect a localised area or a subset of viewers, the effective lost time shall be calculated using a normalisation factor (N) based on the proportion of homes or viewers affected by the outage. The formula is:

$$Effective\ lost\ time\ for\ outage = \ Normalisation\ factor\ (N) * Time\ lost\ due\ to\ outage ,$$

where $N = \frac{Homes\ affected\ by\ outage}{Total\ connected\ homes}$, and

Total connected homes refer to number of homes receiving the TV service.

For outages which affect the entire service, $N = 1$. If an outage affects only certain homes or viewers, then $N < 1$. For cases where $N < 1$, broadcasters shall provide description of how the number of homes affected is determined in the monthly ORI reports.

[31] The "must-carry" channels are Channel 5, Channel 8, Suria, Vasantham, okto, Channel U and Channel NewsAsia.

(d) A Licensee shall also record and report all outage incidents in the monthly ORI reports. The basic details required are date and time of outage incidents; description of incidents in terms of the network elements affected, the number of viewers or subscribers affected, the affected programmes or services and time taken to restore the service. The report on outage incidents shall be in the format as specified by the Authority.

(e) The Channel ORI and Network ORI results should take into account loss of video or sound or control data essential to view the services due to any cause under the control, either directly or through contract arrangements, of the Licensee. Outages that occur due to factors not under the control of the Licensee may be exempted from the computation of ORI results, although they should be recorded in the monthly ORI reports as stipulated in clause 5.3.3(d).[32]

(f) The loss of ancillary data and services such as subtitles is exempted from the computation of ORI results. For such cases, Licensees shall display an apology message on the affected programme(s) as soon as possible when the fault occurs.

(g) A Licensee shall submit to MDA quarterly reports, in the format as specified by the Authority, on viewer complaints[33] relating to outages and poor reception quality that were received over the past three (3) calendar months, within one (1) month from the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.[34]

5.4.4 Guidance notes

(a) Licensees are not required to report the Channel ORI result for an individual channel if no outage is suffered in a calendar month. Outage incidents, if they occur, shall be reported to the MDA on a calendar month basis.

(b) The Network ORI shall be measured at the access network level. For purposes of normalisation, Licensees should provide details on how the number of homes affected is determined in the monthly ORI reports. In cases where the number of homes affected by an outage cannot be accurately determined, Licensees should provide an explanatory note in the monthly ORI reports with details on the nature of the outage and the difficulties in determining affected homes. MDA will consider these outages on a case-by-case basis. In the case where a nationwide outage had occurred for the entire service or channel(s) within the service, the total number of subscribers of the service or channel(s) within the service shall be deemed to have experienced an outage.

[32] Exemptions may include impairment due to external content source, equipment managed by viewers, planned maintenance, unscheduled interruptions to power supply, extreme or unforeseen weather conditions and sun outage(s).

[33] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

[34] Complaints which arise from issues related to viewer premise equipment may be excluded from this report.

6 LOUDNESS REQUIREMENTS

6.1 Introduction

(a) Section 6 “Loudness Requirements” sets out the broadcast standards relating to loudness and must be complied with by Licensees where applicable. The loudness standards must be maintained at the viewers’ premises. Compliance with the required standards will minimise large variations in loudness during transitions between different types of content and between channels.

6.2 Terrestrial broadcast TV, Cable TV, IPTV services

6.2.1 Application

(a) The provisions set out in this sub-section shall apply to Licensees providing Terrestrial broadcast TV services, Cable TV services or IPTV services.

6.2.2 Obligations for Licensees to meet specified requirements

(a) A Licensee shall maintain consistency in the loudness of all audio broadcasts on its Terrestrial broadcast TV services, Cable TV services or IPTV services and shall comply with either the Advanced Television Systems Committee (“ATSC”) or European Broadcasting Union (“EBU”) standards.

(b) Under the ATSC standards, a Licensee shall ensure that all programmes, including commercials, shall comply with the loudness level specified in ATSC RP A/85.[35]

(c) Under the EBU standards, a Licensee shall ensure that all programmes, including commercials, shall comply with the loudness level specified in EBU R128[36].

6.2.3 Compliance with obligations

(a) A Licensee shall provide MDA with a written description of their procedures for ensuring that the required loudness standards are achieved.

[35] ATSC Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television. LKFS is defined as loudness, K-weighted and measured relative to full scale. LU is defined as a loudness unit equivalent to a decibel.

[36] EBU Recommendation R 128: Loudness normalisation and permitted maximum level of audio signals. LUFS is defined as loudness unit measured relative to full scale and LU is defined as a loudness unit equivalent to a decibel.

(b) A Licensee shall carry out internal assessment and monitoring using loudness measurement equipment to verify that loudness levels are in line with the required standards. Both the ATSC RP A/85 and EBU R128 standards refer to a standard ITU measurement algorithm for loudness (ITU-R BS.1770).

(c) A Licensee providing free-to-air Terrestrial TV services shall conduct annual loudness spot checks on programme transmissions at a date and time determined by the Authority. The Licensee shall also prepare and submit to MDA a loudness report for each annual loudness spot check conducted 14 days after the spot check.

(d) A Licensee providing Cable TV or IPTV services shall conduct annual loudness spot checks on programme transmissions at a date and time determined by the Authority. The Licensee shall also prepare and submit to MDA a loudness report for each annual loudness spot check conducted 14 days after the spot check.

(e) The first loudness report is to be submitted one (1) year from the issuance of the Code of Practice for TV Broadcast Standards for existing Licensees, or one (1) year from the commencement of service for new Licensees.

(f) MDA may also direct any Licensee to conduct spot check(s) as and when required by MDA, usually in response to viewer complaints.

(g) A Licensee shall submit to MDA quarterly reports, in the format specified by the Authority, on viewer complaints[37] relating to loudness that were received over the past three (3) calendar months, within one (1) month from the end of the quarter. In the event where no complaints were received for the quarter, the Licensee shall still submit the quarterly report and indicate accordingly.[38]

[37] Viewer complaint refers to an expression of dissatisfaction with the service providers' service in relation to broadcast standards via oral or written communication that requires some action by the service provider beyond the initial contact.

[38] Complaints which arise from issues related to viewer premise equipment may be excluded from this report.

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄十

座談會會議記錄

附錄十 座談會會議記錄

時間	107 年 9 月 4 日（星期二）下午 14:00-16:00
地點	財團法人電信技術中心 高雄本部 1F 國際會議廳
與會單位與人次	與會單位：電信技術中心、凱擘大寬頻、台灣大哥大、國立高雄科技大學、遠傳電信、中華電信等共 6 家，共計 21 人次
內容	一、 數位匯流影音平臺服務品質量測之政策法規： **線上影音平臺服務品質量測的政策意涵** 目前線上影音無論是國內外皆無法律強制的量測標準，主要原因是線上影音服務類型眾多，各家廠商所實施或注重的服務也不盡相同，若需量測，須由廠商發布其 API 或串流源供測試所需，但現階段市場生態如要執行量測勢必遭遇困難。 ● 雖線上影音服務量測準則仍處在模糊地帶，但可針對市售產品列舉較常使用 KPI 作為參考，如： ● 資源上下載速率、資源使用延遲時間、系統穩定度、影音幀速率、影音幀分辨率、影音壓縮率、網路頻寬速率、網路延遲時間、封包遺失率、DNS 解析時間。 ● 為研析針對國內有線電視系統、電信事業固網與行動寬頻等業者了解其影音服務架構與產業現況，並對國際間主要國家之線上影音服務法規、監理政策，研析國際監理機關對線上影音服務之立場與看法。 **美歐網路中立性政策立場** 歐盟網路中立性法規執行準則（2016） ● 準則一：應保障之使用者權利 （法規第 3 條之 1） ● 準則二：合理之商業行為 （第 3 條之 2） ● ISP 與使用者間所達成之協議或契約中，唯侵害使用者權利之項目，否則不受本法之限制。例如吃到飽方案、達流量上限之額外加購流量 ● 準則三：合理之商業行為-零費率爭議 （第 3 條之 2） ● 準則四：合理之流量管理機制 （第 3 條之 3） ● 符合資訊透明（transparent）、無歧視（non-discriminatory）、比例原則（proportionate）等原則 ● 準則五：網路中立性允許提供專業服務（第 3 條之 3） 二、 串流影音平臺服務品質量測方法：

	服務品質簡介： 串流影音服務：透過開放式網際網路直接對用戶提供各種視音訊內容的服務。視音訊內容以串流方式經由網際網路，再透過用戶端的行網或固網傳送至使用者的電視、電腦、智慧型手機或平板電腦等各種終端設備。用戶端可以邊載邊看（只需等待相對短暫的初始片段下載時間，就可以持續收看完整的影音內容）。 串流影音平臺：串流影音服務的提供者 用戶端需求：支援的瀏覽器或特定的播放器/應用程式（App） 營運模式：廣告、贊助、付費訂閱、授權、週邊商品與大數據運用等 **服務品質量化：** 定性而言，影音服務的「高品質」意謂著「低延遲、流暢穩定、高畫質與高傳真」，具體的量化數據才有助於資料蒐集、處理、統計與分析。基於網路性能指標的量測，包含傳輸延遲（DELAYORLATENCY）、傳輸延遲變異（IPDV）、丟包率（PACKETLOSSRATE）、上/下行吞吐量。 客觀 QOE 模型範例：預估QoE得分=（初始緩衝時間得分×$\theta 1$）+（PPI得分×$\theta 2$）+（卡頓率得分×$\theta 3$）

活動簽到表

數位匯流影音平臺服務品質量測座談會

公司	姓名	簽名
財團法人電信技術中心	施嘉興	施嘉興
財團法人電信技術中心	呂維曦	呂維曦
財團法人電信技術中心	廖志峰	廖志峰
財團法人電信技術中心	鄭吉清	鄭吉清
財團法人電信技術中心	楊智鈞	楊智鈞
財團法人電信技術中心	張彥威	張彥威
凱擘大寬頻	林益曄	林益曄
台灣大哥大股份有限公司	陳秉隆	陳秉隆
台灣大哥大	陳煒琳	陳煒琳
南臺科技大學	許子衡	
國立高雄科技大學	陳慶永	陳慶永
遠傳電信	吳文嘉	吳文嘉
遠傳電信	莊政崴	莊政崴
台哥大	徐吉昌	徐吉昌
台哥大	邱思迪	邱思迪
台哥大	龔俊融	龔俊融
高科大	盧韋辰	盧韋辰

數位匯流影音平臺服務品質量測座談會

公司	姓名	簽名
高科 大	張冠中	張冠中
台哥大	楊駿傑	楊駿傑
中 華	王壽斌	王壽斌
中 華	李泰志	李[illegible]
TTC	郭作麟	郭作麟

活動狀況	 

	 
問與答	問：針對網站管理權則區分，國外的影音網站該如何規管？ 答：目前 NCC 的權責僅能規管國內的影音服務平臺業者，若產生糾紛，則以消費者爭議的民事方式處理。

	問：以網路中立行來說，消費者要求低價吃到飽，又要求網路品質須達到某個程度的要求，對行動網路業者來說所需要進行的基礎建設負擔相當龐大與吃重，教授的看法為何？ 答：低價吃到飽確實會影響台灣電信業者的發展，網路中立與網路費率有連帶關聯，政府管越多，資費就不會往上爬，對網路建設期會有很大的負面影響。台灣的電信法規範要求以比其他國家嚴謹與嚴格，因此無須再討論網路中立的問題。 問：影音服務品質的量測標準為何? 答：QoS 我們採用國際間通用的量測方法，QoE 參考國際間（只有）學術與業界使用的方法進行測試。 問：會使用何種量測方式？ 答：會使用客觀的量測方式進行。 問：量測的參考點為何？ 答：會在不同縣市進行單一用戶測試 15 天。

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄十一

說明會會議記錄

附錄十一 說明會會議記錄

場次	台北場
時間	107 年 9 月 7 日（星期五）上午 10:00-12:00
地點	財團法人電信技術中心 台北辦公室 大會議室
與會單位與人次	與會單位：國家通訊傳播委員會、電信技術中心、凱擘大寬頻、中華電信、台灣有線寬頻產業協會、國立高雄科技大學、台灣大哥大、中嘉網路、遠傳電信等共 9 家，共計 27 人次
場次	台中場
時間	107 年 9 月 13 日（星期四）下午 14:00-16:00
地點	國立台中教育大學 求真樓 K401 會議室
與會單位與人次	與會單位：電信技術中心、台灣基礎開發科技、台灣寬頻通訊顧問股份有限公司、群健有線電視、台灣寬頻、台灣大哥大、國立台中教育大學、中華電信、台基科等共 9 家，共計 23 人次。
內容	**串流影音服務簡介** 串流影音服務：透過開放式網際網路直接對用戶提供各種視音訊內容的服務 串流影音平臺：串流影音服務的提供者 用戶端需求：支援的瀏覽器或特定的播放器/應用程式(App) 營運模式：廣告、贊助、付費訂閱、授權、週邊商品與大數據運用等 **服務品質量化：** 定性而言，影音服務的「高品質」意謂著「低延遲、流暢穩定、高畫質與高傳真」，具體的量化數據才有助於資料蒐集、處理、統計與分析。基於網路性能指標的量測，包含傳輸延遲（DELAYORLATENCY）、傳輸延遲變異（IPDV）、 丟包率（PACKETLOSSRATE）、上/下行吞吐量。

	客觀 QOE 模型範例：預估QoE得分=（初始緩衝時間得分×$\theta 1$）+（PPI得分×$\theta 2$）+（卡頓率得分×$\theta 3$） **量測方法概覽** 基於網路性能指標的量測：傳輸延遲(Delay or Latency)、傳輸延遲變異(IPDV)、往返時間延遲(RTT)、丟包率(Packet Loss Rate)、上/下行吞吐量(DL/UL Throughput) 客觀 QOE 模型範例:預估 QoE 得分=(初始緩衝時間得分 × $\theta 1$)+(PPI 得分 ×$\theta 2$)+(卡頓率得分 ×$\theta 3$) **量測工具說明** 提出可行的 QoE 模型：vMOS=f(視訊解析度,初始緩衝時間,卡頓率) 開發量測工具：以 YouTube 服務為量測對象開發 Android App，含網路性能指標（QoS）量測、網路問題診斷、影音服務體驗品質（QoE）量測等功能 **佈建規劃** 進行多個採樣用戶（多家固網與行網業者）的長期測試：共進行三梯測試，每一梯測試 40~50 支行動裝置，每一梯進行 15 天量測，期間每隔 2 小時即自動執行一次測試（一部業者影片與一部共同影片），每單次自動測試約耗時 7~15 分鐘（實際與網路傳輸條件有關）。依此取樣頻率設計，測試預計將可蒐集到 14,400 至 18,000 筆總量測數據，之後再依各分析項目取其中適用的子集合數據進行分析。

活動簽到表(台北場)

數位匯流影音平臺服務品質量測方法說明台北場

公司	姓名	簽名
財團法人電信技術中心	施嘉興	
財團法人電信技術中心	呂維曦	呂維曦
財團法人電信技術中心	廖志峰	廖志峰
財團法人電信技術中心	鄭吉清	鄭吉清
財團法人電信技術中心	楊智鈞	楊智鈞
財團法人電信技術中心	張彥威	張彥威
凱擘股份有限公司	林建吉	林建吉
凱擘股份有限公司	鄭迪	鄭迪
中華電信	蔡傑宇	蔡傑宇
中華電信	陳宥名	陳宥名
中華電信	張期淳	張期淳
中華電信行通分公司	張志遠	張志遠
台灣有線寬頻產業協會	彭淑芬	彭淑芬
台灣有線寬頻產業協會	盧怡靜	盧怡靜
台灣有線寬頻產業協會	陳孟絹	陳孟絹
國立高雄科技大學	陳慶永	陳慶永

數位匯流影音平臺服務品質量測方法說明台北場

公司	姓名	簽名
台灣大哥大	陳立坤	陳立坤
台灣大哥大	楊誠	
台灣大哥大	曾志強	
中嘉網路	盧信儒	盧信儒
台灣寬頻通訊顧問股份有限公司	周譓仁	
遠傳電信股份有限公司	黃漢臣	黃漢臣
遠傳電信股份有限公司	簡宗賢	簡宗賢
遠傳電信股份有限公司	廖上全	廖上全

數位匯流影音平臺服務品質量測方法說明會

公司	姓名	簽名
遠傳電信	顏欽賢	顏欽賢
台灣大哥大	黃平欣	黃平欣
中華電信	蘇昱璋	蘇昱璋
中華電信	吳榮英	吳榮英
〃	陳一之	陳一之
國家通訊傳播委員會	林佳珍	林佳珍

活動狀況 (台北場)	  

活動簽到表(台中場)	

數位匯流影音平臺服務品質量測方法說明台中場

公司	姓名	簽名
財團法人電信技術中心	施嘉興	施嘉興
財團法人電信技術中心	呂維曦	呂維曦
財團法人電信技術中心	廖志峰	廖志峰
財團法人電信技術中心	鄭吉清	鄭吉清
財團法人電信技術中心	楊智鈞	楊智鈞
財團法人電信技術中心	林政諺	
台灣基礎開發科技	黃冠偉	黃冠偉
台灣寬頻通訊顧問股份有限公司	盧廷訓	盧廷訓
國立高雄科技大學	陳慶永	
群健有線電視	何志鴻	何志鴻
台灣寬頻	黃植麟	黃植麟
台灣大哥大	吳韋慶	吳韋慶
台灣大哥大	施明宗	施明宗
台灣大哥大	詹家彰	
台中教育大學	李宗翰	李宗翰
TTC	郭俊麟	郭俊麟
台灣大哥大	許[illegible]加	許[illegible]加

數位匯流影音平臺服務品質量測方法說明台中場

公司	姓名	簽名
中華電信	胡時中	胡時中
中華电信	柯[illegible]	柯[illegible]
中華電信	曾鉉[illegible]	曾鉉[illegible]
中華电信	胡[illegible]玲	胡[illegible]玲
台中教育大學	鄭穎謙	鄭穎謙
TBC	盧廷[illegible]	
台基科	黃冠偉	黃冠偉
〃	彭云瑋	彭云瑋
中華电信	林[illegible]	林[illegible]

<table>
<tr>
<td>活動狀況
(台中場)</td>
<td>


</td>
</tr>
</table>

	 
問與答	問：國外影音平台快速崛起，嚴重影響國內眾多業者，此案所進行的測試是否如同在為國外的影音平台做服務品質背書？ 答：透過本案所提出的量測方法，能有助於消費者釐清影音服務平台服務品質不佳的問題點，但期望勿成為替非規管之影音平台背書工具而造成更大衝擊。 問：此案規劃的測試方法是否會納入電信業者網路服務品質的固定測試項目？ 答：就現階段而言，主管機關並不會納入定期評量的評測項目之中。

數位匯流影音平臺服務品質量測方法之委託研究採購案

期末報告初稿

附錄十二

教育訓練會議記錄

附錄十二 教育訓練會議記錄

場次	北部場
時間	107 年 11 月 22 日(星期四) 上午 10 點至 11 點
地點	交通通訊傳播大樓 2003 會議室(台北市中正區仁愛路 1 段 50 號 20 樓)
與會單位與人次	與會單位：國家通訊傳播委員會、電信技術中心，共計 11 人次
場次	中部場
時間	107 年 11 月 27 日(星期二) 上午 10 點至 11 點
地點	國家通訊傳播委員會中區監理處 1 樓研習室
與會單位與人次	與會單位：國家通訊傳播委員會、電信技術中心，共計 24 人次。
場次	南部場
時間	107 年 11 月 30 日(星期五) 上午 10 點至 11 點
地點	國家通訊傳播委員會南區監理處 402 會議室

與會單位與人次	與會單位：國家通訊傳播委員會、電信技術中心，共計 20 人次。
內容	**串流影音服務簡介** 串流影音服務：透過開放式網際網路直接對用戶提供各種視音訊內容的服務 串流影音平臺：串流影音服務的提供者 用戶端需求：支援的瀏覽器或特定的播放器/應用程式(App) 營運模式：廣告、贊助、付費訂閱、授權、週邊商品與大數據運用等 **服務品質量化：** 定性而言，影音服務的「高品質」意謂著「低延遲、流暢穩定、高畫質與高傳真」，具體的量化數據才有助於資料蒐集、處理、統計與分析。基於網路性能指標的量測，包含傳輸延遲（DELAYORLATENCY）、傳輸延遲變異（IPDV）、 丟包率（PACKETLOSSRATE）、上/下行吞吐量。 客觀 QOE 模型範例：預估QoE得分=（初始緩衝時間得分×$\theta 1$）+（PPI得分×$\theta 2$）+（卡頓率得分×$\theta 3$） **量測方法概覽** 基於網路性能指標的量測：傳輸延遲(Delay or Latency)、傳輸延遲變異(IPDV)、往返時間延遲(RTT)、丟包率(Packet Loss Rate)、上/下行吞吐量(DL/UL Throughput) 客觀 QOE 模型範例:預估 QoE 得分=(初始緩衝時間得分 × $\theta 1$)+(PPI 得分 ×$\theta 2$)+(卡頓率得分 ×$\theta 3$) **量測工具說明** 提出可行的 QoE 模型：vMOS=f(視訊解析度,初始緩衝時間,卡頓率) 開發量測工具：以 YouTube 服務為量測對象開發 Android App，含網路性能指標（QoS）量測、網路問題診斷、影音服務體驗品質（QoE）量測等功能 **佈建規劃** 進行多個採樣用戶（多家固網與行網業者）的長期測試：共進行三梯測試，每一梯測試 40~50 支行動裝置，每一梯進行 15 天量測，期間每隔 2 小時即自動執行一次測試（一部業者影片與一部共同影片），每單次自動測試約耗時 7~15 分鐘（實際與網路傳輸條件有關）。依此取樣頻率設計，測試預計將可蒐集到 14,400 至 18,000 筆總量測數據，之後再依各分析項目取其中適用的子集合數據進行分析。

活動簽到表(北部場)

國家通訊傳播委員會
107年度數位匯流影音平臺服務品質量測方法與應用教育訓練
北區場次/簽到單（學習時數登錄）

時間：107年11月22日上午10時至11時

地點：會本部交通通訊傳播大樓2003會議室

序號	單　位	姓　名(簽到)
1	平臺處	王柳縣
2	綜規處	花[illegible]中
3	平台處	林佳玲
4	〃	龐秀芬
5	〃	廖文予
6	〃	羅婉瑜
7	內容處	楊慧娟
8	平台處	李建榮
9	平台處	陳翰霖
10	平台處	薛鈴琦
11	平台處	林祐仲
12		
13		
14		
15		
16		
17		

第1頁

數位匯流影音平臺服務品質量測教育訓練北部場

公司	姓名	簽名
通傳會	劉姿馥	
通傳會	王慧瓊	
通傳會	龐秀芬	龐秀芬
通傳會	林雅燕	
通傳會秘書室	張新民	
通傳會平臺事業管理處	林祐仲	
通傳會平臺事業管理處	林佳玲	林佳玲
通傳會平臺事業管理處	蘇鈴琇	
通傳會平臺事業管理處	羅婉瑜	
通傳會平臺事業管理處	王柳霖	
通傳會平臺事業管理處	伍麗芬	
通傳會基礎設施事務處	王嘉鵬	
通傳會電臺與內容事務處	楊慧娟	
通傳會電臺與內容事務處	林怡萱	
通傳會北區監理處	蔡國棟	
	陳翰霖	陳翰霖

數位匯流影音平臺服務品質量測教育訓練北部場

公司	姓名	簽名
財團法人電信技術中心	施嘉興	
財團法人電信技術中心	楊智鈞	楊智鈞

活動狀況 (北部場)	  

活動簽到表(中部場)

數位匯流影音平臺服務品質量測教育訓練中部場

公司	姓名	簽名
通傳會平臺事業管理處	林佳玲	林佳玲
通傳會中區監理處	王紹柏	
通傳會中區監理處	黃東賢	黃東賢
國家通訊傳播委員會	洪澔璟	洪澔璟
〃	周宇行	周宇行
〃	王錫奎	王錫奎
〃	吳孟芳	吳孟芳
〃	劉孟忠	劉孟忠
〃	林銘鋒	林銘鋒
〃	楊明松	楊明松
〃	林岡章	林岡章
〃	王麗貞	王麗貞
〃	巫靜宜	巫靜宜
〃	舒宸祥	舒宸祥
〃	鄭智元	鄭智元
〃	林坤成	林坤成

數位匯流影音平臺服務品質量測教育訓練中部場

公司	姓名	簽名
NCC 中區監理處	朱汶鈺	朱汶鈺
〃	張紹庭	張紹庭
NCC 〃	許紹楨	許紹楨
〃	黃惠梓	黃惠梓
〃	王嬡貞	王嬡貞
〃	黃朝麟	黃朝麟
〃	李彥賢	李彥賢
〃	藍美華	藍美華
〃	陳佩均	陳佩均

數位匯流影音平臺服務品質量測教育訓練中部場

公司	姓名	簽名
財團法人電信技術中心	施嘉興	
財團法人電信技術中心	楊智鈞	楊智鈞

活動狀況 (中部場)	  

活動簽到表(南部場)

數位匯流影音平臺服務品質量測教育訓練南部場

公司	姓名	簽名
通傳會平臺事業管理處	林佳玲	
通傳會南區監理處	王秀蕙	王秀蕙
通傳會南區監理處	祝淑蘭	祝淑蘭
通傳會南區監理處	洪惠珠	
〃	丁佩瑋	丁佩瑋
〃	王成賢	王成賢
〃	顏志隆	顏志隆
〃	侯青秀	侯青秀
〃	許榮進	許榮進
〃	蕭麗貞	蕭麗貞
〃	林芳蓮	林芳蓮
〃	蔡宜瑋	蔡宜瑋
〃	張仁發	張仁發
〃	林瓊章	林瓊章
〃	蘇鳳鈴 ~~蘇鳳[illegible]~~	蘇鳳鈴
〃	葉瑞岑	葉瑞岑

數位匯流影音平臺服務品質量測教育訓練~~中~~南部場

公司	姓名	簽名
南區處	高毓喬	高毓喬
南區處	蘇麗琴	蘇麗琴
南區處	李信玲	李信玲
南區處	方子安	方子安
〃	袁書蘋	袁書蘋
〃	蕭[illegible]	蕭[illegible]

數位匯流影音平臺服務品質量測教育訓練~~中~~南部場

公司	姓名	簽名
財團法人電信技術中心	楊智鈞	楊智鈞

活動狀況
(南部場)

大綱
串流影音服務產業現況
國際間影音服務品質量測方法研析
串流影音服務簡介與典型的傳輸系統架構
量化品質量測，Why, Who, and How?
量測方法概覽
基於網路性能指標的量測
基於使用者體驗/感受的量測
影音服務品質量測APP
操作與Q&A
電信技術中心

	 
問與答	問：為什麼測試工具無法測試線上影音服務所有的影片？ 答：測試需要蒐集影音服務執行時的相關數據，這些數據必須由業者有提供，現階段測試工具只能使用配合的廠商提供測試源進行測試。