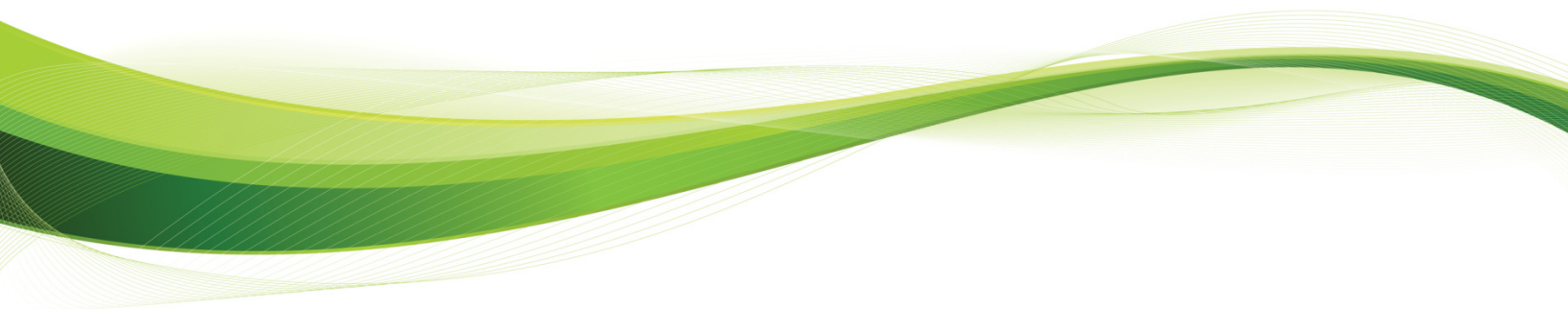




*Public Consultation on the Open Internet and Net Neutrality in  
Europe*



## Introduction

Sandvine appreciates the opportunity to provide comments in connection with the European Commission's (EC) *Questionnaire for the Public Consultation on the Open Internet and Net neutrality in Europe* (the Questionnaire). Headquartered in Waterloo, Ontario, Canada, Sandvine was established in 2001 and employs over 400 people globally. Sandvine's solutions are used by more than 200 Internet service provider customers in over 80 countries, including over 40 in Europe alone. Together, Sandvine's customers serve over 90 million fixed line broadband subscribers and more than 250 million mobile subscribers.

Sandvine is the global leader in network policy control solutions, which make the Internet better by protecting and improving the Internet experience for subscribers. The solutions comprise network equipment and software that help DSL, FTTx, cable, fixed wireless and mobile operators better understand network traffic, manage network congestion, create new services, mitigate traffic that is malicious or undesirable to subscribers, deliver QoS-prioritized multimedia services and increase subscriber satisfaction. A core part of Sandvine's technology is deep packet inspection, or DPI, one of the enabling technologies of the Internet. In January 2010, Infonetics Research named Sandvine as the market share leader in the "Standalone DPI Market."

Sandvine is very familiar with the Open Internet & Network Neutrality debate. One of Sandvine's major customers, Comcast Corporation, used a Sandvine solution to enable the traffic management technique that was at the centre of the Network Neutrality debate in the United States. In late 2008, Comcast switched to another Sandvine solution, Fairshare Traffic Management, and Comcast still uses that solution to manage traffic today. In 2009, Sandvine made submissions to the United States' Federal Communication Commission's (FCC) Notice of Proposed Rule Making on the Open Internet<sup>1</sup> and the FCC's Public Notice on broadband measurement and consumer transparency in fixed line networks<sup>2</sup> and a similar Public Notice for mobile networks<sup>3</sup>. In Canada, Sandvine made submissions to the Canadian Radio-television and Telecommunications Commission's (CRTC) Review of Internet Traffic Management Practices<sup>4</sup>.

In this document, Sandvine provides comments to certain questions in the EC's Questionnaire.

---

<sup>1</sup> Sandvine Incorporated. See <http://fjallfoss.fcc.gov/ecfs/document/view?id=7020370020>

<sup>2</sup> Sandvine Incorporated. See <http://fjallfoss.fcc.gov/ecfs/document/view?id=7020352787>

<sup>3</sup> Sandvine Incorporated. See <http://fjallfoss.fcc.gov/ecfs/document/view?id=7020514594>

<sup>4</sup> Sandvine Incorporated. See [http://www.crtc.gc.ca/public/partvii/2008/8646/c12\\_200815400/1029527.pdf](http://www.crtc.gc.ca/public/partvii/2008/8646/c12_200815400/1029527.pdf)

## Questions

**Question 1:** Is there currently a problem of net neutrality and the openness of the internet in Europe? If so, illustrate with concrete examples. Where are the bottlenecks, if any? Is the problem such that it cannot be solved by the existing degree of competition in fixed and mobile access markets?

To answer this question, there must be some understanding of what is meant by Net Neutrality and an open Internet. To do this there has to be agreement on what we mean by *Internet*. For the purposes of this discussion, Sandvine defines Internet as the group of networks that provide consumers with Internet access, as this is where the issues related to the debate arise. There are obviously many other networks that interconnect to comprise the Internet, but the central issues of openness and neutrality focus on consumer rights and so have not arisen in these other networks in the same way.

As another fundamental point, we have to identify who the Internet needs to be open and neutral for. As with any piece of critical social infrastructure such as highways or the electrical grid, Sandvine submits that the Internet exists to serve its end users - consumer Internet subscribers. A highway exists to serve traveler's transportation needs. While vehicle manufacturers, road construction companies, pavers and sign manufacturers all offer valuable products and services in the highway ecosystem, the traveler's needs define the needs of the infrastructure. Similarly, the Internet does not exist for the benefit of network providers, application providers, or content providers. These entities only exist to meet the demands of subscribers. The Internet exists to serve subscriber's communication, entertainment, information, and other ever-expanding needs. Subscribers' needs define what the Internet needs to be.

While there is no universally accepted definition, in Sandvine's experience with this debate a market for consumer Internet access can be said to be *open* and *neutral* if:

1. Openness: subscribers have access to all the lawful, non-harmful content, applications, and compatible devices of their choice;
2. Neutrality: there is no undue discrimination in access to devices, content or applications of the consumers' choice. Reasonable traffic management helps establish and maintain a neutral and open Internet;
3. Transparency: there is sufficient transparency with respect to the terms and conditions of network operators' service offerings such that consumers are making informed choices. Such transparency would include clear descriptions of the potential impact on users' rights under points 1 and 2 of:
  - a. Any traffic management practices;

- b. Any managed services included as part of the service plan;
  - c. Any limitations on access to content, applications and compatible devices included as part of the plan.
4. Competition: there is sufficient competition in the market for consumer access to the Internet. The absence of competition does not necessarily result in closed, non-neutral networks.

It is important to note that, with respect to the Openness requirement, not every individual service offering has to offer fully open access as long as, collectively, the market provides consumers with open access. In this way, individual operators and service offerings can offer service plans that better suit users' preferences and ability to pay, such as a discounted mobile service offering that restricts access to VoIP. It has been a presumption in the Network Neutrality debate globally that subscribers value openness to all content, applications and devices above all else, but this may not be the case. The presumption is likely founded on the basis that, largely, "fully-open" plans have been the only ones available to subscribers to date. Only recently have alternatives begun to appear and only experimentation with different types of service offerings will determine which plans subscribers value most. Further, with respect to the Neutrality requirement, the fact that a subscriber and network operator openly enter into a contract (with transparently disclosed terms) for a plan that is not "fully-open" defines the discrimination involved in the plan as reasonable, and not undue.

Regulation or legislation is only necessary where there is direct evidence of market failure to provide an open and neutral Internet.

In Sandvine's experience with over 40 European network operators (including all access technologies: DSL, mobile, cable and FTTx) the markets of the European Union are open and neutral. Users have enjoyed unfettered access to the devices, content and applications of their choice. Mobile devices all work on the GPRS standard, which has made it easy to offer the same devices over multiple networks. The full breadth of content and applications has been available. According to Ofcom, there have been no European cases of undue discrimination resulting from inappropriate traffic management practices:

"Ofcom has not received any formal complaints about traffic management from industry, and this also seems to be the case in the majority of other EU countries. So far consumer and citizen groups have remained largely silent on the issues."

Sandvine's experience with traffic management echoes Ofcom's. None of Sandvine's European customers have deployed network policies that have been found to challenge the openness or neutral character of the Internet. In fact, amongst Sandvine's 200-plus network operator customers, only one has ever been found to have deployed a policy that challenged

the principles of an open and neutral Internet. That decision was against a U.S. operator, was highly controversial and was ultimately overturned<sup>5</sup>. The Federal Communications Commission in the United States<sup>6</sup>, the Google-Verizon proposal in the United States<sup>7</sup> and the Canadian Radio-television and Telecommunications Commission decision in Canada<sup>8</sup> have all enshrined (in slightly differing terms) the notion of *reasonable network management* in their definitions of an open and neutral Internet.

With respect to transparency, European network providers have improved the disclosure of the key aspects of managed services, access limitations, and traffic management practices related to their service offerings, though there is still room for improvement. Sandvine offers some suggestions for a transparency framework in its answer to Question #12. Finally, the extremely high level of competition for consumer Internet access in Europe is unmatched globally. As a result, Sandvine believes that Europe will be a launching pad for many innovative service offerings that will help to personalize the market for Internet access so that users pay just for the value they extract from their Internet experience. Such competition will also help to ensure that the European market remains open and neutral in the future.

---

<sup>5</sup> United States Court of Appeals. *Comcast Corporation v Federal Communications Commission and United States of America*. See <http://online.wsj.com/public/resources/documents/comcastfcc.pdf>

<sup>6</sup> Federal Communications Commission. *Notice of Proposed Rulemaking, October 22, 2009*. See <http://fjallfoss.fcc.gov/ecfs/document/view?id=7020348621>

<sup>7</sup> *Verizon-Google Legislative Framework Proposal*. See <http://www.scribd.com/doc/35599242/Verizon-Google-Legislative-Framework-Proposal>

<sup>8</sup> Canadian Radio-television and Telecommunications Commission. *Telecom Regulatory Policy CRTC 2009-657*. See <http://www.crtc.gc.ca/eng/archive/2009/2009-657.htm>

## Question 2: How might problems arise in future? Could these emerge in other parts of the internet value chain? What would the causes be?

The Internet is a commons, with each stakeholder able to affect the other. Even in a market where network operators offer full access to all *available* content, application and devices, the other stakeholders in the commons can restrict availability to their own products and services. Today, some hardware manufacturers are restricting application choice (e.g. Apple restricts Internet applications on its iPhone devices). Some application vendors are entering into exclusive arrangements (e.g. Skype with Verizon<sup>9</sup>) which prohibit the use of their application on other networks. Some content providers (e.g. ESPN360) are licensing their content only to specific Internet Service Providers. From the consumer standpoint, all equally infringe on the concept of an open Internet, regardless of where the restriction has been introduced.

Similarly, the economies of scale among certain content providers (e.g., Amazon or Apple iTunes) and content-delivery-networks (CDN) (e.g., Akamai) can lead rise to the same restrictions. As the large content sites and CDN's grow, their cost structure for delivering content gets smaller on a per byte (or per unit) basis, and that allows them to over-compete with new entrants. In such cases, they could discount their existing services to a level that would create inadequate or negative returns for new entrants that are considering a major investment but that don't enjoy the same scale. This in turn reduces the openness principle of the Internet.

As described in its answer to Question #1, Sandvine is in favour of network operators having the freedom to offer a full variety of service plans, including those that could restrict access to certain applications, as such variety benefits the consumer with additional choice. However, today, each party (consumer, access provider, transit provider, CDN provider, content provider, advertising provider) has an incentive to maximize their own gain, at the expense of all other players, and there may be no incremental cost to them for doing so. To be sustainable in the long term, Sandvine believes that the Internet value chain of consumer, access provider, transit provider, content-delivery-network, content, and advertiser, all need to have their economic interests aligned. Today, there is a modest increase in cost to a content network to double bandwidth, and an enormous cost increase to an access provider for the same increase. Aligning parties' economic interests end-to-end will be in the best interest of the consumer since it ensures that networks will be built to the maximum efficient utility. Sandvine has suggested some potential alternatives to achieve this goal in its answer to Question #8.

---

<sup>9</sup> Skype. *Verizon Wireless and Skype join forces to create a global mobile calling community*. See <http://about.skype.com/press/2010/02/verizon.html>

**Question 4:** To what extent is traffic management necessary from an operators' point of view? How is it carried out in practice? What technologies are used to carry out such traffic management?

### Traffic Management is Necessary

While traffic is *managed* for several purposes (for example, to remove malicious traffic from a network or to meet the terms of service agreements, such as with parental controls), as a first step, Sandvine believes that it is necessary to define traffic management for the purposes of responding to the Questionnaire. Broadly speaking, traffic management is the act of increasing the efficiency and/or quality of experience of a network given the envelope of technologies deployed in the network at a given time. In the context of the Network Neutrality debate the notion of “traffic management” has largely centred on the prioritization of traffic during times of congestion. Accordingly, Sandvine will focus its comments on this area.

*Traffic management is necessary because network congestion happens.* Consumer access networks are oversubscribed, so that there is access for all. It is the most efficient model to deliver a high speed broadband service at an economical price to consumers. However, as a result, congestion will occur when more people or traffic use an Internet location than there is capacity to support that use - *congestion is a product of peak usage, not average usage.* As an example, service providers in Japan and South Korea, which offer bandwidth to the home up to 20 times faster than typical residential broadband in Canada, still regularly experience significant periods of congestion and use traffic management techniques at those times.

Sandvine has observed that participants in the Network Neutrality debate often incorrectly refer to service plans that include monthly bandwidth limits, or quotas, as a means of traffic management. Such service plans do not address congestion, which occurs at times of peak network usage. Instead, they align the “average” capacity of the network with the “average demand” of the users. Monthly bandwidth limits represent a means for network operators to differentiate their service offerings, rather than representing a traffic management technique.

It is frequently unpredictable when and where congestion will occur. For example, a service provider cannot adequately provision its network for event or location driven surges in traffic, sudden changes in network demographics, or sudden losses of capacity that overwhelm the network. The network issues widely reported during the inauguration of US

President Barack Obama<sup>10</sup> or as news related to Michael Jackson's death circulated the Internet provide an excellent example.

### An Unmanaged Network is Not Neutral

When congestion occurs, if the network is left unmanaged certain applications and users will consistently dominate network resources to the detriment of the quality of experience of other applications and users. In short, *an unmanaged network is not a neutral network*. Some background on the different requirements of applications helps to illustrate.

Applications differ with respect to the amount of bandwidth (or throughput), latency, jitter, and packet loss that they require in order to be delivered at an expected quality of service level. Sandvine submits the following definitions for purposes of the Questionnaire.

- Bandwidth: traffic volume over time. It is usually measured over a short time, such as bits/second or megabits/second (Mbps), which is 1,000,000 bits/second.
- Latency: the delay for a message to get from one communications end point to the other, e.g., the time it takes for a VoIP data packet to leave the speaker's mouth and arrive at the listener's ear. It is typically measured in milliseconds.
- Jitter: the variation in the latency of one message to another, typically measured in milliseconds (e.g. if the first message takes 1ms and the second message takes 10ms, then there is 9ms of jitter).
- Packet loss: occurs when one or more packets of data traveling across a computer network fail to reach their destination. Packet loss can be caused by a number of factors, including signal degradation over the network medium, oversaturated network links, corrupted packets rejected in-transit, faulty networking hardware, faulty network drivers or normal routing routines.

While bandwidth gets most of the attention, *adding bandwidth is not always (or even mostly) the answer to improving the user's quality of experience for an application*. The other factors can play a crucial role, so traffic management needs to be used in combination with adding network capacity.

---

<sup>10</sup> Macworld.com. *Inauguration taxes mobile phone networks*. See [http://www.macworld.com/article/138331/2009/01/cellphone\\_inauguration.html](http://www.macworld.com/article/138331/2009/01/cellphone_inauguration.html)

An application can be classified into one of three categories based on its requirements of a network across these four characteristics:

- **Bulk applications.** These applications include P2P file-sharing (e.g., BitTorrent, FastTrack, etc), web surfing, usenet news (NNTP), and file transfers over FTP or HTTP, for example, and will go as fast as the network will permit. TCP is designed to achieve the maximum communication rate possible, using all resources available. In practice bulk applications will go as fast as the thinnest part of the network between the client and server. In the case of the server co-located within the ISP network (e.g. a content-delivery network, a cache), this will be bound by the access equipment speed. In the case of a server which is located farther away, this may be bound by transit (connection to all worldwide public networks) or peering (connection to other nearby private networks) performance. Typically, servers of bulk applications (e.g. Speedtest.net, Rapidshare.com, Megaupload.com) will saturate the download speed of the consumer's modem, as they typically download-only. In the case of P2P, it is bi-directional so it can also have the same effect in the upstream direction.

Most bulk applications can run unattended by the user. File transfers are initiated by the user, who may then walk away - often for hours or even overnight - while the process completes. Bandwidth is the primary determinant of transfer speed of long-running connections and performance will generally improve linearly with increases in bandwidth. As a result, latency and jitter matter much less - users likely would not even notice their effect. Packet loss affects throughput: as packet losses occur, TCP reduces the number of packets sent per second and only increases throughput once packets stop being dropped.

Web surfing represents an exception in the Bulk category. "Web 2.0" sites have introduced interactive components to web surfing - the user typically attends the activity and data travels bi-directionally as users have started to be content providers in their own right. Increases in bandwidth do not translate linearly to increased performance as latency is a gating factor to the end-user experience. Beyond the time it takes to transfer the content, loading a website typically involves four "round trips" between a personal computer and the related web server. First, the Domain Name Server (DNS) must resolve the domain name (i.e, translate [www.sandvine.com](http://www.sandvine.com) to its numeric Internet address) then the three-way handshake established by TCP must be completed<sup>11</sup>.

---

<sup>11</sup> IETF RFC 793. See <http://www.ietf.org/rfc/rfc0793.txt>

Each of the four round trips is subject to the latency in the network, and when added together this delaying effect becomes the limiting factor in the transmission such that additional bandwidth does not dramatically improve loading times for a website. To illustrate with an example, if the latency in a single round trip is 500 milliseconds (0.5 seconds), a website would take at least two seconds (4 round trips x 0.5ms) to load even if the subscriber had an infinitely high bandwidth connection. Typically “Web 2.0” sites have more than one file (images, videos, ads, etc), so the time to load can be substantially worse.

- Interactive applications. These applications are paced by the consumer. In the case of VoIP, bandwidth largely depends on silence suppression and the codec bandwidth chosen, but it is typically 8-30Kbps. The bandwidth requirements of interactive applications are often modest (though in the case of video conferencing the rates are significantly higher: 200-500Kbps is common), but they typically require very low latency, jitter and packet loss to achieve a satisfactory quality of experience. For example, a VoIP user can perceive latency of 150 milliseconds on a call, and delays greater than 300 milliseconds render the call unusable<sup>12</sup>. As with web surfing, adding bandwidth will not necessarily address quality of service issues. In general, because of the sensitivity of Interactive applications to latency, jitter and packet loss it is particularly important to protect the quality of service for these applications.
- Paced/Burst-paced applications. Streaming video and audio applications such as YouTube and SHOUTcast fall into this category. The media involved has a natural bit rate, and the connection tries to achieve this rate on average over its lifetime, though for short durations the media will ‘burst’ to provide buffering on the client to allow for packet loss on the network (YouTube, because it uses TCP, will attempt to transmit at line rate initially). So, these applications can be modeled by the media they carry. For typical Internet streaming today, rates of approximately 300-400Kbps are common. Hulu, YouTube, and others are starting to shift to higher-definition video, for which the rate can increase to 1-6Mbps of bandwidth.

With paced/burst-paced applications it is important that a network sustain the minimum bandwidth requirements, but because of the buffering involved additional bandwidth only marginally improves performance, by making the applications less sensitive to latency, jitter and loss in the network.

---

<sup>12</sup> International Telecommunication Union. *ITU-T Recommendation. G.114*. See <http://www1.cs.columbia.edu/~andrea/new/documents/other/T-REC-G.114-200305.pdf>

The following table provides some representative benchmarks to achieve a minimum quality of service for certain popular applications. Such figures require significant assumptions, which Sandvine has included as Appendix 1:

Application Category	Application Class	Minimum Bandwidth	Maximum Latency	Maximum Jitter	Maximum Loss
<b>Bulk</b>	P2P	19Kbps	n/a		
	Web surfing	1Mbps (Web 2.0)	166ms (latency + jitter)		n/a
	Email	60Kbps	n/a		
	Usenet news	195Kbps	n/a		
	FTP file transfers	195Kbps	n/a		
<b>Interactive</b>	VoIP	16Kbps	300ms (latency + jitter)		< 0.5%
	Video gaming	50Kbps	75ms (latency + jitter)		< 0.5%
	Video Conferencing	250Kbps	300ms (latency + jitter)		< 0.05%
<b>Paced (and burst-paced)</b>	Video streaming streaming	300Kbps, to not have much of a wait time	< 1s for "channel change"	<50ms	<0.05%
	High def video	1-3Mbps depending on quality of HD.	< 1s for "channel change"	<50ms	<0.05%
	Audio streaming	Audio:160Kbps for CD quality.	< 1s for "channel change"	<50ms	<0.05%

Sandvine's studies have shown that latency-sensitive real-time, interactive applications are increasing dramatically in popularity, particularly during typical peak network hours when the opportunity for congestion is highest<sup>13</sup>. Again, traffic management, not just increases in network capacity is required to solve that problem.

### Regulators Recognize the Need for Traffic Management

These inherent differences in application traffic are starting to be recognized in "Network Neutrality" decisions around the world. In October 2009, the Canadian Radio-television and Telecommunications Commission (CRTC) concluded its Review of Internet Traffic Management Practices (ITMPs). As part of its decision, the CRTC stated:

"The Commission notes that the degree to which an application or service is delayed may have an impact on its performance. Furthermore, transmission delays may affect some types of applications or services more than others. For these reasons, it is important to identify which types of traffic and/or applications would be impacted by transmission delays.

In the case of time-sensitive audio or video traffic (i.e. real-time audio or video such as video conferencing and voice over Internet Protocol (VoIP) services), ITMPs that introduce delays or jitter are likely to cause degradation to the service. The Commission considers that when noticeable degradation occurs, it amounts to controlling the content and influencing the meaning and purpose of the telecommunications in question."

Accordingly, the CRTC required that the CRTC vet in advance any network management practice that affects time-sensitive traffic in this way.

With respect to non-time-sensitive traffic, the CRTC further decided:

"With respect to non-time-sensitive traffic, the Commission considers that the use of ITMPs that delay such traffic does not require approval under section 36 of the Act. However, the Commission is of the view that non-time-sensitive traffic may be slowed down to such an extent that it amounts to blocking the content and therefore controlling the content and influencing the meaning and purpose. In such a case, section 36 of the Act would be engaged and prior Commission approval would be required."

---

<sup>13</sup> 2009 Global Broadband Phenomena Study, Sandvine Incorporated, <http://www.sandvine.com/downloads/documents/2009%20Global%20Broadband%20Phenomena%20-%20Executive%20Summary.pdf>

While the Network Neutrality debate in the United States continues, the FCC itself has recognized the need for “reasonable network management”, both in their most recent proposal as well as in their original policy statement. The recent Verizon-Google proposal<sup>14</sup> shows that major industry players on previously opposite sides of the Network Neutrality debate have also come to recognize the need for traffic management.

### How Traffic is Managed

The practice of managing congestion in telecommunications networks has been around for a long time. The traditional PSTN phone network used call admission control to manage congestion - a call was not admitted to the network unless end-to-end capacity existed to handle it. Traditional IETF standardisation for Quality of Service (QoS) calls for the communication endpoints to “mark” traffic<sup>15</sup> with the desired drop/latency/quality characteristics so that routers at each hop can weight their decisions. However, this marking has a few problems. It is not universally obeyed by current generation access devices in either DSL networks (ATM L2 backbone), or cable networks (DOCSIS layer), where congestion is highest, due to technical limitations in the devices. Also, the devices can’t trust the marks: over time applications have cheated the system by mischaracterizing their traffic in order to achieve higher priority. Accordingly, this solution no longer works.

A new class of intelligent networking equipment products, like Sandvine’s, emerged to allow service providers to accurately identify and characterize network activity and provide policy-based control of network congestion. Modern congestion management solutions aim to mitigate congestion in a way that achieves a neutral result - so that applications get the type of resources they need to satisfy users’ demands and so that users each get a fair share of the network.

Congestion Management Approaches. In Sandvine’s experience, there are two commonly used technological approaches to managing data packets when a broadband network is congested. The first is to limit the rate at which the packets of a specified class enter the network (through Traffic Policing or Traffic Shaping), which can delay session transmission. The second is to limit the number of concurrent sessions of a specified class on the network (Session Management), which can delay session initiation. These approaches are outlined below.

- Traffic Policing. A method of ensuring that packets of a specified class do not exceed a specified bit rate. For example, a 4Mbps service offered by a service provider is

---

<sup>14</sup> *Verizon-Google Legislative Framework Proposal.* See <http://www.scribd.com/doc/35599242/Verizon-Google-Legislative-Framework-Proposal>

<sup>15</sup> *IETF. An Architecture for Differentiated Services.* See <http://tools.ietf.org/html/rfc2475>

policed to 4Mbps per subscriber. Traffic policing can be applied per subscriber or per application. Bit rate limits are configurable.

- Traffic Shaping. Grooming a subset of traffic to a specific, configurable maximum bit rate. It is similar to traffic policing, but instead of dropping packets that exceed the bit rate limit, packets are placed in a queue and metered out so as not to exceed the bit rate limit. Traffic shaping can be applied per subscriber or per application. Again, bit rate limits are configurable.
- Session Management. Instead of limiting traffic to a specific bit rate, all transmissions occur at a full rate up to a threshold limit for a specified class. Any sessions above that threshold are not initiated at that network location so that they may occur instantly on another location of the same network (where the threshold has not been exceeded) or on other networks, or initiation of the session may be delayed. Session management is applied on an application basis. The threshold limit is fully configurable per application by the service provider.

Traffic Prioritization. With traffic management solutions such as those offered by Sandvine, service providers can expand the attributes on which policies are applied. For example, by default routers give all packets equal priority, and when a packet drop occurs it may be on a packet that is more sensitive to loss or is more valuable to the subscriber at that moment. Traffic policing and traffic shaping can be enhanced through prioritization techniques, which apply different classes of services to packets, giving each class a different priority. Prioritization can be strict. For example, if ever there is a packet of the top priority class queued for delivery, it will go first. This approach runs the risk of starvation of the lowest priority classes. Or, priority can be weighted to avoid the risk of starvation. Priority assignments and their weighting are entirely configurable. The classification of packets can be done on a per-flow, device, application and/or subscriber basis.

Application-centric and Subscriber-centric Policies. The ability to create traffic management policies on a joint per-application/per-subscriber basis greatly enhances the quality of experience to the end user and the “fairness” of the network. Such policies can allocate scarce network resources efficiently by taking into account the different characteristics of bulk, interactive and paced traffic, as well as different subscriber usage patterns for these applications.

- Application-centric. On a congested network link, prioritize interactive, real-time network applications that are latency- and jitter-sensitive (e.g., VoIP, online gaming) and that most affect the users’ perception of the current quality of Internet experience. Protocols that are not latency- and jitter-sensitive and that are typically

unattended by the user (such as email, FTP or P2P file transfers) can accept lower priority without any meaningful impact to the user's quality of experience.

- Subscriber-centric. On a congested network link, prioritize the traffic of subscribers who are not contributing disproportionately to congestion over a given time period, so that all users are free to consume their "fair share". A "fair share" based policy is designed to ensure fairness across users
- Subscriber- and Application-centric. Apply priority in a subscriber-centric manner while simultaneously prioritizing latency- and jitter-sensitive interactive applications - even for disproportionate users. This highly targeted approach isolates the root causes of congestion and preserves the quality of experience of affected subscribers to the maximum extent possible. In the future, Sandvine believes that service providers will be able to provide tools to their subscribers to let them directly select which applications should be prioritised in the event that they are identified as users contributing disproportionately to network congestion.

Upstream versus Downstream Policies. Congestion management can also be applied to upstream and/or downstream traffic. Networks have historically been designed with less upstream bandwidth than downstream bandwidth. As more applications and websites have encouraged the transmission of data upstream, the upstream path has frequently been the first point in the network where congestion is experienced. Consequently, separate congestion management policies for upstream and downstream traffic may need to be considered.

Depending on the approach and objective, a traffic management solution may need hardware or software technologies for:

- a) Collecting network usage data from network access equipment, such as a Cable Modem Termination Systems (CMTS) or a Broadband Residential Aggregation System (BRAS). Standardized techniques exist, such as polling via Simple Network Management Protocol (SNMP), or through IP Detail Records<sup>16</sup> (IPDR). Data collection can occur at configurable intervals.
- b) Collecting subscriber usage data from core or aggregation routers. Data collection can occur at configurable intervals.
- c) Identifying applications on an intelligent policy traffic switch like Sandvine's.

---

<sup>16</sup> See <http://en.wikipedia.org/wiki/IPDR>.

- d) Policy definition and signalling, from a separate policy controller, such as Sandvine's Service Delivery Engine.
- e) Policy enforcement (e.g., traffic shaping and traffic policing) can and has been effected by many access devices and core network equipment, such as a cable modem, CMTS, DSL modem, or BRAS. Sandvine's intelligent policy traffic switch also enforces policy.

### Sandvine's Principles for Reasonable Traffic Management

The notion of "reasonable network management" (or the absence of "undue discrimination") has become a cornerstone concept in the developing Network Neutrality debate. As with any principle related to the debate, the notion of reasonable network management must be framed in terms of the end user's Internet experience. Did the traffic management practice make the Internet experience better for most network users most of the time? Were users unreasonably limited in their access to content, applications, or devices of their choice as a result of the network management practice? In consultations with the FCC in the U.S., the CRTC in Canada, industry leaders such as the National Cable and Telecommunications Association, and network operators globally Sandvine has advocated the following criteria for "reasonable network management."

1. Narrowly-tailored  
Traffic management is implemented only where congestion exists and when congestion is causing quality of experience issues for a large number of subscribers.
2. Proportional and reasonable effect  
A traffic management policy has an effect on subscribers or applications that is proportional to the effect the user or application is having on the network. Policy applies the smallest reasonable intervention to alleviate congestion in the network and improve quality of experience for the majority of subscribers.
3. Legitimate and demonstrable technical need  
Congestion and/or quality of experience issues can be demonstrated to exist in the network and management's technical remedies are effective in achieving its targeted goals.
4. Transparent disclosure  
Network operators need to disclose traffic management policies and changes thereto in a simple, useful and predictable manner.
5. Auditable  
Network operators can demonstrate that the above requirements are met through the auditing and reporting capabilities of its traffic management solution.

## Standards -based Approaches to Traffic Management Continue to Evolve

The Internet Engineering Task Force (IETF) is the open standards organization that works to develop and promote Internet standards, in particular those related to TCP/IP and the Internet protocol suite. The IETF's standards-based approaches, such as LEDBAT<sup>17</sup>, TCP/IP, ECN<sup>18</sup> (Explicit Congestion Notification), PCN<sup>19</sup> (Pre-Congestion Notification) use properties of the network to detect congestion and automatically reduce traffic throughput when congestion is evident. The algorithms rely on changes in latency or increased queuing (which causes latency to increase) and are specifically designed for bulk transfer applications. These initiatives are potentially effective but will take time to become adopted. For example, ECN was standardized in 2001 and is just appearing in network stacks in the latest operating systems. ECN was introduced in Windows Vista and, as in most other operating systems, it is disabled by default.

LEDBAT is an IETF initiative to standardize the implementation of a non-congestion-causing transport. Proprietary implementations of LEDBAT such as Microsoft CTCP, uTorrent uTP also exist, but there is some concern that they may interact badly with other standards, such as ECN, to create unfair results.

With the IETF's recent work on Congestion Exposure (ConEx)<sup>20</sup>, it is beginning to investigate how to make all stakeholders accountable for their impacts on the Internet commons. ConEx was discussed at the most recent meeting of the IETF, on November 10, 2009 in Hiroshima, Japan<sup>21</sup> and is described this way:

“Congestion Exposure (ConEx) is a proposed new IETF activity to enable congestion to be exposed along the forwarding path of the Internet. By revealing expected congestion in the IP header of every packet, congestion exposure provides a generic network capability which allows greater freedom over how capacity is shared. Such information could be used for many purposes, including congestion policing, accountability and inter-domain SLAs. It may also open new approaches to QoS and traffic engineering.”

“The Internet is, in essence, about pooling resources. The ability to share capacity has been paramount to its success and has traditionally been managed through the voluntary use of TCP congestion control. However, TCP alone is unable to prevent bandwidth intensive applications, such as peer-to-peer or streaming video, from

---

<sup>17</sup> IETF. *Low Extra Delay Background Transport (ledbat)*. See <http://datatracker.ietf.org/wg/ledbat/charter/>

<sup>18</sup> IETF. *RFC 3168, The Addition of Explicit Congestion Notification (ECN) to IP*. See <http://tools.ietf.org/html/rfc3168>.

<sup>19</sup> IETF. *Congestion and Pre-Congestion Notification (pcn)*. See <http://www.ietf.org/html.charters/pcn-charter.html>.

<sup>20</sup> IETF. *Congestion Exposure (conex)*. See <https://datatracker.ietf.org/wg/conex/charter/>

<sup>21</sup> IETF. See <http://www.ietf.org/mail-archive/web/int-area/current/msg02041.html>

causing enough congestion to severely limit the user-experience of many other end-hosts.”

“We believe these problems stem from the lack of a network-layer system for accountability -- among all parties -- for sending traffic which causes congestion. We propose a metric where IP packets carry information about the expected rest-of-path congestion, so that any network node may estimate how much congestion it is likely to cause by forwarding traffic. A network operator can then count the volume of congestion about to be caused by an aggregate of traffic as easily as it can count the volume of bytes entering its network today. Once ISPs can see rest-of-path congestion, they can actively discourage users from causing large volumes of congestion, discourage other networks from allowing their users to cause congestion, and more meaningfully differentiate between the qualities of services offered from potential connectivity partners. Meanwhile end-hosts may be freed from rate restrictions where their traffic causes little congestion.”

**Question 5:** To what extent will net neutrality concerns be allayed by the provision of transparent information to end users, which distinguishes between managed services on the one hand and services offering access to the public internet on a 'best efforts' basis, on the other?

In an open and neutral market, it is highly desirable for network operators to offer a full variety of service tiers, including “managed services” that offer enhanced performance for certain applications, and other service tiers that limit access to certain applications. These service tiers can be offered in the market alongside unmanaged, “best-efforts” Internet services, with the market setting the appropriate prices for each. Such a “menu” of service offerings gives consumers more choice and segments the market so that consumers pay only for what they want, rather than everyone being obliged to pay for the full “buffet”, as is often the only option available today with traditional all-you-can-eat Internet access packages.

For example, it would be desirable for a mobile network operator to offer only discounted broadband Internet services that excluded VoIP and peer-to-peer file sharing applications. Equally, it would be desirable for a fixed line operator to offer a premium-priced online gaming service that delivered all gaming traffic in a prioritized fashion. Ultimately the market will decide the attractiveness and pricing of each service and each will prosper (or not) accordingly.

As Sandvine discussed in its answer to Question #1, for Network Neutrality to exist, a market needs to have openness, neutrality, competition and transparency, as defined in that answer. Openness does not have to apply to each service offering, but to the market in general. From a consumer’s perspective it is highly desirable to have the choice of managed service offerings that discriminate between types of traffic. With respect to neutrality, if a subscriber and network operator contract for a service that discriminates between traffic (such as the examples above), they are defining the discrimination as *reasonable*. Beyond that, providing that there is sufficient competition in the market to allow for consumer choice between network operators and service offerings, transparency of service terms is sufficient to make the market work efficiently and allay Network Neutrality concerns.

## Question 6: Should the principles governing traffic management be the same for fixed and mobile networks?

Sandvine believes that it is desirable to have a common set of rules for all stakeholders in the Internet commons, including mobile network operators. As long as traffic management principles (such as those suggested by Sandvine in the answer to Question #4) are framed in terms of the subscriber's Internet experience then they can immediately apply to all access networks. Such a framework inherently allows for the differences between access network characteristics to drive different "reasonable" traffic management techniques.

Unlike fixed line networks, the bandwidth available in a wireless network is fixed and defined by its associated radio spectrum. With limited bandwidth, issues related to latency, jitter and packet loss also become exacerbated in the mobile environment. So, mobile networks are particularly susceptible to congestion and quality-of-service issues, and such limitations are already being noticed by users of some of the world's largest mobile networks<sup>22</sup> despite still-modest data usage.

A few bulk file-sharing or file-transfer sessions occurring on a particular node of a fixed line network are unlikely to cripple the user experience for all other applications. The same may not be true for a given mobile network. A few similar sessions on a cell site could seriously impair the web surfing, voice call and gaming experience of all users connected to that site, for example. Consequently, to protect the user experience for these popular applications, it would be reasonable to create a policy that began to manage bulk application traffic at a lower threshold (and/or manage it in a different way) than for a fixed line network. In fact, these applications, and others like Slingbox (which "slings" bandwidth-intensive television signals to Internet devices, such as a Smartphone), may have such a detrimental effect on network performance for all applications in a given location that blocking them could be deemed a reasonable practice in a given situation at a given time. A case-by-case analysis would have to be performed to know.

Similarly, if managing mobile data traffic on a subscriber-specific basis, it might be necessary to start managing "disproportionate users" consumption at a threshold level that would be much lower (or managing it in a different way) than for fixed line networks.

Since the standard for each rule would be based on what the user experiences on his network, Sandvine's framework would automatically allow for different network management practices based on different network characteristics.

---

<sup>22</sup>ArsTechnica. *AT&T CTO downplays role of iPhone in network's issues*. See [http://arstechnica.com/apple/news/2009/10/att-cto-downplays-role-of-iphone-in-networks-issues.ars?utm\\_source=rss&utm\\_medium=rss&utm\\_campaign=rss](http://arstechnica.com/apple/news/2009/10/att-cto-downplays-role-of-iphone-in-networks-issues.ars?utm_source=rss&utm_medium=rss&utm_campaign=rss)

The test of whether a traffic management policy is reasonable or not would fall to the analysis of individual cases. Did the traffic management practice make the Internet experience better for most network users most of the time? Were users unreasonably limited in their access to content, applications, or devices of their choice as a result of the network management practice?

Mobile networks are the newest entrants in the market for broadband access so user behaviour is rapidly evolving. The nature of data traffic traversing the mobile network is more dynamic than for any other access network class. Time and experimentation will be required for individual network operators to devise the most effective traffic management policies.

**Question 7:** What other forms of prioritisation are taking place? Do content and application providers also try to prioritise their services? If so, how - and how does this prioritisation affect other players in the value chain?

The business model of network operators creates an inherent incentive for them to ensure that *all* content, applications, devices and users enjoy a good quality of experience. A poor experience results in subscriber churn. In other words, the incentives of network operators are generally aligned with most subscribers' interests. In this regard, they stand alone in the Internet value chain. Application and content developers are *uniquely* interested in maximizing the quality of experience *for just their own offerings and users*, even if doing so harms the experience for other content and applications. Not surprisingly then, application and content developers have found ways to prioritize their own Internet offerings (which often requires network operators to deploy traffic management policies to create a level playing field).

#### Prioritization by application developers

Certain applications have been developed to take a disproportionate share of network resources, thereby harming the user experience for other applications at times of network contention. For example, applications or services that do not use congestion control<sup>23</sup> (a mechanism to naturally slow output as increases in packet loss or buffering are detected) automatically gain priority in the network and contribute disproportionately to congestion. TCP congestion control is largely studied and designed around a single or small number of connections from a given host. Some applications, such as typical P2P file-sharing applications, commonly use a high number of TCP sessions per host, which can frustrate congestion control to take up the maximum amount of bandwidth available. If you imagine a four-lane highway system, and each data packet as a vehicle on that highway, a typical file-sharing packet will immediately expand to become two-lanes-wide as soon as the lane next to it is empty, and so on. Again, *an unmanaged network is not a neutral network*.

Also, applications which drive a proprietary congestion control algorithm, that is not based on the transport layer (TCP, UDP, etc.) and is instead based on delay may be at risk of causing oscillation on networks which have variable bandwidth (e.g. WiMax, mobile). An example would be the IETF LEDBAT standard, a “low Extra Delay Background Transport”, which is used in popular BitTorrent peer-to-peer clients such as uTorrent.

---

<sup>23</sup> Implemented in TCP in 1987, see [http://en.wikipedia.org/wiki/Congestion\\_control](http://en.wikipedia.org/wiki/Congestion_control).

## Prioritization by content developers

Content developers prioritize their traffic by moving it closer to the subscriber. For example, content owners will strike commercial arrangements with a content delivery network, such as offered by companies like Akamai Technologies, who have servers co-located within consumer Internet access networks - i.e., near the subscriber instead of at some remote location on the Internet. The proximity of the hosted content to subscribers results in a prioritized service levels compared to that of content providers that can't afford such commercial arrangements.

Some extremely popular content providers co-locate dedicated servers in access networks to deliver *just their own content*. Examples would include Google and Amazon.com. Again, the result is prioritized service levels for the hosted content compared to that of other content providers without similar resources.

**Question 8:** In the case of managed services, should the same quality of service conditions and parameters be available to all content/application/online service providers which are in the same situation? May exclusive agreements between network operators and content/application/online service providers create problems for achieving that objective?

The Internet is comprised of a set of connected private networks. In the diagram in **Error! Reference source not found.**, the consumer buys connectivity for a set time from the access provider. The access provider purchases bandwidth on a peak utilization basis from the transit provider. The aggregation provider also purchases bandwidth on peak utilization from the transit provider. The content owner purchases bandwidth from the hosting provider by peak utilization, as does the hosting provider from the aggregation provider.

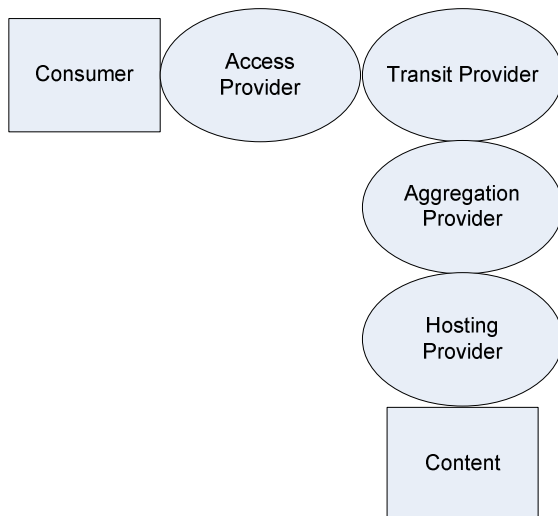


Figure 1: Stylised network topology

So, currently there is a disconnect. The Internet is treated as a set of two-sided contracts, and there is no alignment of the cost model or interests. Consumer Internet access is the only link in the chain based on time (e.g., monthly contracts). All other contracts, either between the network operators or between the content owners and network operators are based on volume: ninety-fifth percentile bandwidth usage. This disconnect causes inefficient use of the network.

The concept of “wholesale QoS”, competitively priced, would align the economic interests of all the parties. Wholesale QoS is the concept of network operators selling parameters of network usage related to quality of service rather than time. Wholesale QoS aligns economic interests, because an access provider would make available a tariffed service with certain

peak usage or volume, latency, jitter or packet loss characteristics that are aligned with the quality of service required for the application being used. With a wholesale QoS model, at each Internet exchange the content flow could request the network behaviour required to guarantee the necessary quality of service, and either be allowed it or denied it, based on technical (capacity does not exist) or economical (nobody is willing to pay) bases. A similar example is used in the public phone network: if there is insufficient network bandwidth, the caller receives a 'fast-busy' signal (this is called network admission control).

In wholesale QoS, the consumer could pay the tariff directly or, like a toll-free telephone model, the tariff can be paid by the content provider. Since the content owner requires sufficient access bandwidth, and quality of service, in order to deliver their product, they would have an incentive to purchase end-to-end quality of service if it was available and economical. Various business models emerge:

- YouTube may not be willing to pay for a QoS guarantee but an advertiser linked to the video might.
- A subscription streaming service such as Netflix may require a guaranteed bandwidth sufficient for high-definition video. They have a monetary relationship with the consumer. If the access network cannot provide this quality, neither the consumer nor the content provider will be satisfied. Allowing the content provider to request a guaranteed QoS, as part of that subscription fee, allows all business models to work. Similarly, quality of service need not be bandwidth based, but could be volume based.
- Today, the Amazon Kindle contains a 3G modem. The consumer does not see this as they have a monetary relationship only with Amazon. For each book purchased, Amazon purchases network bandwidth to transport the content. This allows a simplified business model for all parties.

Wholesale QoS would ensure that as content providers increase the fidelity of their content, the consumers could enjoy it correctly. It would also give network operators an incentive to invest in their networks, as they are being compensated for the quality of service they deliver by collection of the tariffs. Finally, consumers would pay a fair rate for the services they receive, rather than the current "all-you-can-eat" model which favours disproportionate users. The disconnect is repaired.

Standards bodies have long recognized this need for quality of service. Examples are the IETF RFC 2474 (Differentiated Services), IETF RFC 2478 (Common open policy service), the DSL Forum Resource Admission Control (RAC-F) function, the ETSI 3GPP Policy Control And Charging architecture (3GPP PCC), etc. The one thing these standards do not address is a business model to align the interests - they solely focus on the technical abilities to do so. Today, there are some technical challenges associated with end-to-end QoS spanning

multiple networks. If a commercial driver for it were to exist, these challenges would be solved.

A harder to resolve problem would be the capacity model needed in the access network for the unknown and unforecasted demand of application providers who might suddenly over-tax a network. This would be no worse than the 'non QoS' use case of today, but the consumer expectation might not be met.

Today there are some pre-standards activities underway to provide content-neutral methods of aligning the economic interests and providing the effect of wholesale QoS. An example is the IETF working group 'ConEx', as discussed in Sandvine's answer to Question #4 to the Questionnaire.

Network operators who also provide managed content service may have minimum service quality level agreements with the content owners, relating to fidelity of video, availability, etc. These should not be exclusive of providing wholesale QoS on underlying network properties to other applications of similar kind. Quality of experience should be monitored at an application level to ensure like applications receive similar experience to the consumer.

**Question 9:** If the objective referred to in Question 8 is retained, are additional measures needed to achieve it? If so, should such measures have a voluntary nature (such as, for example, an industry code of conduct) or a regulatory one?

If the regulatory environment favoured or allowed wholesale QoS, operators would be inclined to commercially provide it. This economic incentive would drive the innovation required to solve the inherent technical challenges in achieving end-to-end QoS and capacity planning, across multiple networks. Economic interests alone should be sufficient. Normally in this environment an Internet standards body such as the IETF would draft the appropriate standards and 'Best Current Practices (BCP)' documents which would outline the expectations. Although voluntary, such standards tend to be well adopted and neutral.

**Question 10:** Are the commercial arrangements that currently govern the provision of access to the internet adequate, in order to ensure that the internet remains open and that infrastructure investment is maintained? If not, how should they change?

See Sandvine's responses to Questions #2 and #8.

## **Question 12: How should quality of service requirements be determined, and how could they be monitored?**

For network performance disclosures to be meaningful, subscribers need to know whether a network can reliably deliver the expected quality of experience for their favourite applications - based on the service level they have contracted for with their network operator. Network performance disclosure should cover peak and off-peak times and times when traffic management is in effect. To achieve these goals, network operators need to report the capabilities of their network at the subscribers' location by application and compare that to the minimum performance requirements for each application class.

### **Application Requirements**

In its answer to Question #4, Sandvine supplied some representative benchmarks to achieve a minimum quality of service for certain popular applications, supported by significant assumptions included as Appendix 1.

Sandvine submits that each Internet application provider should report the minimum bandwidth and maximum latency, jitter and packet loss required for satisfactory delivery of their applications. This way, network operators could have access to averages for an application class for their own network reporting purposes (as described below) and consumers could make better decisions about how to use their own network connections.

### **Network performance measurements**

Sandvine submits that network performance should be measured on a per-application class basis because satisfactory application performance is of central importance to the user's experience. For an individual user, the measurement of his network performance will be in part determined by the applications he uses. For example, if a subscriber only uses his Internet connection for online video gaming, which typically demands bandwidth of approximately 50Kbps, then his measured bandwidth performance over a given period will be 50Kbps, even though his service tier may promise and could in fact deliver much more. The shortfall would be as a result of the subscriber's preferred usage of the connection, not necessarily any limitation in the network connection itself to deliver speeds up to the promised throughput. The user's experience for gaming is, in fact, better defined by the latency, loss and jitter.

Sandvine submits that network providers should measure their network's performance for each subscriber, by application class, on the following metrics:

- Average achieved peak bandwidth at peak hours and off-peak hours.

- Average latency during peak and off-peak hours.
- Average jitter during peak and off-peak hours.
- Average loss during peak and off-peak hours.

The network measurements could be taken monthly over a one-minute interval in peak and off-peak times. Sandvine’s own study has shown that peak hours are from approximately 7:00 pm to 10:00 pm globally<sup>24</sup>. These hours could be reliably used to define “peak” and “off-peak” for the purpose of these measurements.

### **Method of measurement**

Ofcom has done some pioneering work in this area for fixed line networks by using equipment installed in the users’ home to measure network performance. While Sandvine applauds the efforts, we believe that in-home measurements have some important limitations:

- The technology in the home introduces some variables that are not under the control of the service provider. For example, users with badly shielded wires, or with a CB antenna, or running HomePNA<sup>25</sup> networking can create a significant noise problem that affects network performance.
- The use of the in-home monitoring is voluntary, which will not guarantee an appropriate distribution of measurements across all of the access networks in a market. All networks should be properly represented in order to give consumers adequate information for an informed choice.
- In a mobile environment, Doing measurements on the device themselves leads to several problems:
  - Inconsistent implementations by manufacturer;
  - Additional data load on the network generated by the measurement and reporting of results;
  - Cost to consumer for additional data load;

---

<sup>24</sup> 2009 Global Broadband Phenomena Study, Sandvine Incorporated, <http://www.sandvine.com/downloads/documents/2009%20Global%20Broadband%20Phenomena%20-%20Executive%20Summary.pdf>

<sup>25</sup> HomePNA is a home-networking technology that re-uses the existing coaxial wire in your home to form a local area wired network. It is fundamentally incompatible with DOCSIS since it uses the same upstream spectrum. A single user running HomePNA without having disconnected from the public cable plant can affect the upstream bandwidth of all shared users.

- Device battery life and memory are limited.

Accordingly, Sandvine recommends that each access network provider perform the measurements through appropriate network equipment and report the results directly to both regulators and the public.

**Reporting**

The network measurements should be reported to consumers and regulators as averages for a service area, by access-type, and service plan for both peak and off-peak hours. Here is one potential presentation of the data:

**Service Area X  
DSL Platinum Service  
Average Network Performance for January 2010  
Peak Hours**

Application Class	Minimum Bandwidth	Maximum Latency	Maximum Jitter	Maximum Loss
<b>P2P</b>	<i>Required</i>	195Kbps	n/a	n/a
	<i>Delivered</i>	250Kbps	n/a	n/a
<b>Web surfing</b>	<i>Required</i>	1Mbps (Web 2.0)	166ms (latency + jitter)	n/a
	<i>Delivered</i>	2Mbps	160ms (latency + jitter)	
<b>Email</b>	<i>Required</i>	60Kbps	n/a	n/a
	<i>Delivered</i>	120Kbps	n/a	n/a
<b>Usenet news</b>	<i>Required</i>	195Kbps	n/a	n/a
	<i>Delivered</i>	250Kbps	n/a	n/a
<b>FTP file transfers</b>	<i>Required</i>	195Kbps	n/a	n/a
	<i>Delivered</i>	250Kbps	n/a	n/a
<b>VoIP</b>	<i>Required</i>	16Kbps	300ms (latency + jitter)	< 0.5%

Application Class	Minimum Bandwidth	Maximum Latency	Maximum Jitter	Maximum Loss
<i>Delivered</i>	50Kbps	185ms		0.1%
<b>Video gaming</b>				
<i>Required</i>	50Kbps	75ms (latency + jitter)		< 0.5%
<i>Delivered</i>	75Kbps	45ms		0.1%
<b>Video Conferencing</b>				
<i>Required</i>	250Kbps	300ms (latency + jitter)		< 0.05%
<i>Delivered</i>	265Kbps	185ms		0.3%
<b>Video streaming</b>				
<i>Required</i>	300Kbps	< 1s	<50ms	<0.05%
<i>Delivered</i>	425Kbps	0.5s	22ms	0.25
<b>High def video streaming</b>				
<i>Required</i>	2Mbps	< 1s	<50ms	<0.05%
<i>Delivered</i>	1.8Mbps	0.5s	22ms	0.25
<b>Audio streaming</b>				
<i>Required</i>	128Kbps	< 1s	<50ms	<0.05%
<i>Delivered</i>	200Kbps	0.5s	22ms	0.25

**Green:** Network **can** reliably deliver required application performance

**Orange:** Network **may** reliably deliver required application performance (average performance better than required performance by 10% or less)

**Red:** Network **cannot** reliably deliver required application performance (average performance worse than required performance)

The reported averages can be based on sampling of the full set of subscriber data measurements as long as an appropriate level of statistical reliability is achieved. In order to provide some idea of network trends, the data could also be presented in a time series (e.g., for the last six months) to graph changes in application performance on the network over time.

Consumer reporting of network performance measurements should be easy to find, understand and compare between network providers. Most users don't understand the

network demands (bandwidth, latency, jitter, loss) of their favourite applications. So, any disclosure by the network provider about its network's ability to deliver an application has to be made in simple terms, i.e., "yes, we can deliver the application reliably", "no we can't deliver the application reliably", or "maybe can deliver the application reliably". Additionally, simple definitions of each performance characteristic could be provided for interested users.

The information should be made available on the network provider's public website and (to the extent it is practical) identified on the website in a standard way across network providers (e.g., "Network performance, by application") and to the extent feasible in a standard location (e.g., as a link from the operator's Terms of Service). Such standardization would assist consumers' ready access to the data.

With respect to the effect of traffic management practices on network performance, a similar table to that illustrated above could be presented for the network when traffic management policies are in effect.

### Mobile Broadband Services Measurement and Disclosure

Conceptually the same principles that apply to measurement and reporting of application requirements versus network performance apply to mobile networks as well. However there are some significant technical differences between fixed and mobile networks that require special consideration. Additional or different measurements may need to be taken and reported on. For Sandvine's full discussion on this topic, please refer to <http://fjallfoss.fcc.gov/ecfs/document/view?id=7020514597>

**Question 13:** In the case where NRAs find it necessary to intervene to impose minimum quality of service requirements, what form should they take, and to what extent should there be co-operation between NRAs to arrive at a common approach?

Sandvine recommends against national regulatory authorities (NRAs) attempting to regulate an international set of networks as there may be unintended consequences. Access networks are joined to other access networks via complex multi-national links, and national regulation may cause inter-network operators to reroute traffic in less optimal, less efficient ways to avoid regulatory cost.

Sandvine believes that, in a competitive market, transparent disclosure of network operators' performance capabilities, as described in Sandvine's answer to Question #12, would encourage network operators to far exceed any minimum standard that a NRA could reasonably impose.

To the extent that NRAs impose minimum quality of service requirements, Sandvine urges that these be done with respect to Internet applications consumers use, rather than inherent properties of the network (e.g. a typical user can stream standard definition video for more than 1 hour uninterrupted vs. an operator must provide 1Mbps of minimum bandwidth at all times). Quality of experience is framed with respect to the user, and the applications they use, and may be a moving target as the Internet, applications, and devices evolve. An acceptable user experience on a handheld mobile device is a different experience than on a home PC on a wireline connection.

Additionally, to the extent that NRAs impose minimum quality of service requirements, Sandvine urges that these be done in a coordinated fashion across participating countries to avoid balkanization of the Internet experience and to avoid unduly burdensome requirements for network operators that offer their services in multiple countries.

**Question 14:** What should transparency for consumers consist of? Should the standards currently applied be further improved?

Sandvine believes that it has addressed this question in its response to Question #12.

## Appendix 1 - Assumptions for Application Requirements

P2P, Usenet, FTP: Bandwidth is the most important network characteristic as it affects the time required for these applications to transfer the data. A typical movie is about 700 MB. If a typical user expects, at a minimum, to download a movie overnight (i.e., 8 hours), the minimum bandwidth required would be 195 Kbps.

Web Surfing: A typical “Web 2.0” website requires approximately 10 to 20 connections to download the approximately 0.5MB to 2MB of data needed to display the page. Studies have shown that to maintain a good user experience this must be done within 2-4 seconds<sup>26</sup>.

To reach a website, its name must first be translated into its numeric IP address via the DNS. This happens for each server that the webpage references. Many webpages have images, videos and advertisements on different servers and thus the Internet browser must resolve each DNS name. Each time DNS is used, 2x the latency (for the round trip) is added to the total time to load the page. Also, to compensate for jitter in most PC environments, the PC buffers the data to the extent of the jitter in the network so that the total latency time is actually (latency + jitter).

For each connection or file that needs to be downloaded, the (latency + jitter) is added multiple times as the browser initiates a TCP connection to the server to retrieve the file. This multiple is usually three, in accordance with TCP’s three-way handshake for initiating connections.

Once the connection is established, the time to download the file is a function of the bandwidth available. However, given that websites often have many small files (images, text), TCP is not always able to achieve the full throughput rate due to its “slow-start<sup>27</sup>” algorithm.

The above argument does not take into account many of the complex algorithms or tools in place such as parallel connections and HTTP pipelining, but does show that bandwidth is not the only determining factor for measuring HTTP quality of experience. In fact, latency and jitter will likely be the gating factors on the user’s quality of experience with Web 2.0 websites.

For a 1.5MB webpage with 20 connections to load with a satisfactory user experience, available bandwidth must be at least 1 Mbps and latency+jitter must not exceed 166ms.

Email: A normal text email is between a few kilobytes and a few hundred kilobytes. Email is not instantaneous, however, there is a perception that it is near real-time. To send or receive an email with a large attachment in under a minute, the bandwidth required is approximately 60Kbps.

---

<sup>26</sup> See [http://www.akamai.com/html/about/press/releases/2009/press\\_091409.html](http://www.akamai.com/html/about/press/releases/2009/press_091409.html)

<sup>27</sup> IETF RFC 2001. See <http://tools.ietf.org/html/rfc2001#ref-2>

VoIP: The most basic audio codecs require bandwidth of approximately 16 Kbps (allowing for overhead of the Internet). VoIP is a real-time application that is very sensitive to latency and jitter. ITU-T G.114 suggests that the maximum one-way latency + jitter be 150 ms<sup>28</sup> (or round-trip 300 ms), above which it becomes noticeable to the end user. Most VoIP protocols use stateless connections (UDP) and have no built in retransmit. Loss must not be over 0.5% for calls to be audible.

Video Conferencing: Similar to VoIP, the application is bi-directional and is highly susceptible to latency, jitter and loss, however the bandwidth requirements are higher due to the addition of the video.

Video & Audio Streaming. These applications are primarily uni-directional. The average normal-definition video on YouTube requires approximately 300Kbps. High-definition videos (depending on the quality, i.e., different encodings) require bandwidth between 1-3Mbps. Because most streaming done on websites like YouTube use HTTP, latency, jitter and loss are not a major concern. Traditional streaming video (RTSP, RTP, etc) are done over UDP and are affected more by loss. Streaming of compressed CD quality audio requires approximately 160Kbps of bandwidth.

---

<sup>28</sup> International Telecommunication Union. ITU-T Recommendation. G.114. See <http://www1.cs.columbia.edu/~andrea/new/documents/other/T-REC-G.114-200305.pdf>