Intelligent Broadband Networks

Comments on Draft Reports

Prepared for:

Body of European Regulators for Electronic Communications

BEREC

Differentiation Practices and Related Competition Issues in the Scope of Net Neutrality

Guidelines for Quality of Service in the Scope of Net Neutrality

About Sandvine

Sandvine appreciates the opportunity to provide comments in connection with BEREC's draft reports: *Differentiation practices and related competition issues in the scope of Net Neutrality* and *Guidelines for quality of service in the scope of Net Neutrality*. Because the two reports are tightly linked, Sandvine has provided a single set of comments for both.

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 hundreds of millions of broadband and mobile data 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. Recent Infonetics Research reports have named Sandvine as the market share leader in the "Standalone DPI Market."

Sandvine is very familiar with the Network Neutrality debate. In 2009, Sandvine made submissions to the United States' Federal Communication Commission's (FCC) Notice of Proposed Rule Making on the Open Internet1 and the FCC's *Public Notice on broadband measurement and consumer transparency in fixed line networks*2 and a similar Public Notice for mobile networks3. In Canada, Sandvine made submissions to the Canadian Radio-television and Telecommunications Commission's (CRTC) *Review of Internet Traffic Management Practices*4. In 2010, Sandvine provided comments in connection with the European Commission's *Questionnaire for the Public Consultation on the Open Internet and Net neutrality in Europe*.

[1] Sandvine Incorporated. See http://fjallfoss.fcc.gov/ecfs/document/view?id=7020370020
[2] Sandvine Incorporated. See http://fjallfoss.fcc.gov/ecfs/document/view?id=7020352787
[3] Sandvine Incorporated. See http://fjallfoss.fcc.gov/ecfs/document/view?id=7020514594
[4] Sandvine Incorporated. See http://www.crtc.gc.ca/public/partvii/2008/8646/c12_200815400/1029527.pdf

A Reasonable Approach

Overall, Sandvine believes that BEREC has outlined a very reasonable approach for addressing differentiation and QoS in Internet Service Providers' (ISPs) networks. In particular, Sandvine concurs with these significant conclusions:

- There are many potentially legitimate goals for differentiation, including congestion management, the delivery of differentiated service tiers to end users, and network integrity
- The goal of differentiation should not be:
    - To execute an anti-competitive practice
    - To discriminate between individual Content and Application Providers.
- Reasonable traffic management is necessary to optimize network quality and efficiency. The cost savings resulting from such efficiency can be passed on to users through lower Internet Access Service prices.
- The reasonableness of a differentiation practice depends on a number of factors:
    - Its goal needs to be legitimate and driven by some underlying technical need
    - Its effect on end-users and Content and Application Providers, including notions of proportionality (such as effectiveness, and whether the practice is sufficiently narrowly-tailored to reasonably meet its stated objective, e.g., managing congestion)
    - Internet Access Service market conditions.
- Recognition that ISPs business models have changed as over-the-top services have challenged their traditional revenue sources, such as voice telephony, messaging and video. As a result, ISPs need to develop new revenue streams to provide sufficient incentives for investment in their networks, and differentiation practices could form one aspect of that.
- Differentiated services are commonly seen as welfare-enhancing to a market, as they tend to increase the diversity of offers available.
- It is inappropriate to foreclose on the idea of the Internet ecosystem developing into a two-sided market. While such business models have not yet achieved mass adoption, it is currently prudent to let market forces determine whether this evolution will occur or not.
- In all cases of differentiation, market contexts need to be taken into account, such as:
    - competitiveness and availability of alternative choices
    - transparency of offerings
    - switching costs.
- In all situations, case-by case analysis is necessary. Prescriptive, ex ante prohibitions or solutions are not appropriate.

While Sandvine does agree with BEREC’s approach in general, there are a few points that we believe are worth raising:

1. The “dirt road” Internet can occur (and in fact has already, in certain circumstances) even in the absence of ISPs traffic differentiation practices. Certain applications have been engineered to consume as much bandwidth as is available, leading to de facto prioritization of that traffic. An unmanaged network is not a neutral network. Because of the position of ISPs in the Internet ecosystem, they arguably have more natural incentives than Content and Application Providers to prioritize traffic in a way that would be in the best interests of most users most of the time. The ISP’s interest is in fairness user-to-user, the Content and Application Provider’s interest favours its own application or content over all others.

2. Despite recommending a case-by-case analysis in virtually all aspects of its analysis, BEREC’s reports have communicated an ex ante preference for application-agnostic traffic management techniques, even suggesting that NRAs may want to foreclose on application-specific differentiation in certain circumstances. One challenge of this approach is that application-specific techniques can produce much more proportional results by affecting only the traffic that needs to be affected to alleviate network congestion. Proportionality is one of the key aspects in determining the reasonableness of differentiation practices. Sandvine recommends that BEREC adopt a more neutral stance on application-specific differentiation and let the case-by-case analysis of “reasonableness” (including the notions of proportionality) pervade in all aspects. Such an approach would also better align with the recommendations of technical standards bodies such as the IETF, which have traditionally governed the network in a neutral way.

3. Sandvine has had great success helping a number of its customers to deploy application-specific differentiated service tiers, as bolt-ons to an unrestricted Internet Access Service. We provide an example for BEREC’s consideration.

4. BEREC suggested that minimum QoS targets could be set per application, but sees practical difficulties in doing so in a best efforts Internet. Sandvine believes that one of the key benefits of application-specific traffic differentiation is to achieve just this sort of goal – delivering the QoS necessary to provide a satisfactory user experience for most applications most of the time.

An Unmanaged Network is Not a Neutral Network

BEREC states that traffic differentiation, if widespread, could result in a “dirt road” best-efforts Internet on the basis that the bandwidth remaining for unprioritized traffic may become too limited to deliver satisfactory quality. While this may be one possibility, it ignores an important fact - without any traffic differentiation on the part of an ISP whatsoever, *traffic is prioritized anyway*, based on the design characteristics of the underlying applications. Accordingly, applications that are designed to consume a disproportionate share of the network resources can leave only a “dirt road” for other applications. In fact, this phenomena has happened to some extent in some networks in the past.

Certain applications quickly consume bandwidth. For example, typical peer-to-peer (P2P) file-sharing protocols, are specifically designed to maximize their use of available bandwidth – like a truck that expands in width to overtake adjacent lanes of the highway as soon as they become available. Sandvine recently reported that in European fixed access networks, two popular filesharing protocols (BitTorrent and eDonkey) together represented 50% of upstream bandwidth and 30% of aggregate upstream and downstream bandwidth[5]. These applications often run unattended on a 24x7 basis by highly active file-sharers and can introduce significant latency and jitter into the network.

Certain interactive, real-time applications, like voice over Internet protocol (VoIP) and online video gaming, are extremely sensitive to latency and jitter, and are closely associated with subscribers’ overall sense of network quality. Sandvine’s data shows that there is a surge in the use of these interactive, real-time entertainment and communication applications during peak hours, when the opportunity for congestion is highest. So, applications like P2P have the opportunity to create (and indeed have created in the past) a “dirt road” Internet for other applications, simply by virtue of their engineering. In short, an unmanaged network is not a neutral network.

Unlike ISPs, which have an inherent incentive to deliver a satisfactory performance for all applications (assuming adequate competition), Content and Application Providers are not at all subject to such discipline. In fact Content and Application Providers have a strong incentive to optimize the performance of only their application or content in the network. In its report on differentiation practices, BEREC demonstrates that it understands this point. What applications will be developed in the future? How might they impact the performance of other applications? How will that impact the QoE of end users? And would that impact the investment incentives of Content and Application Providers and ISPs?

---

[5] Sandvine Incorporated. *Global Internet Phenomena Report 1H 2012*. See http://www.sandvine.com/news/global_broadband_trends.asp.

Application-specific Traffic Management Allows for Most Proportional Results

BEREC has indicated in both its “Differentiation Practices” and “Quality of Service” reports a strong preference for application-agnostic traffic management techniques. In certain parts of its reports, BEREC states that “In some cases an NRA may also consider it relevant to prohibit application-specific restrictions on a general basis.”

Application-agnostic techniques certainly can effectively reduce congestion in certain situations. However, such techniques are not necessarily the most proportional in all cases, and BEREC acknowledges that proportionality is one the key tests in determining the reasonability of a differentiation practice. We submit BEREC should strive to maximize the QoE of each application class, and focus on that as the desired outcome.

Consider an application-agnostic policy that is targeted at disproportionate users, perhaps defined as the top 10% of users of the network over a period of fifteen minutes prior to the network arriving at a congestion state. Such users will have been consuming a wide variety of traffic, some of which would be contributing significantly to their bandwidth consumption and others not. For example, by their nature, applications like VoIP, online video gaming and others do not contribute meaningfully to network congestion, but because they are time-sensitive applications, their usefulness to the consumer is greatly impacted by any delays in their delivery. By layering on an application-specific aspect, the congestion management policy could affect only:

- disproportionate users;
- applications that contribute disproportionately to bandwidth consumption; and
- applications that are not time-sensitive.

Such a policy could still achieve the congestion management goal and be significantly more narrowly-tailored, and therefore more proportional, than a policy that affected 100% of a disproportionate users’ traffic. In many cases, the most proportional result can only be achieved by focusing on both users and the traffic within the users’ profiles that are contributing disproportionately to network congestion.

Additionally, application-specific traffic management has enabled some very successful Internet offerings to date, and could enable some highly anticipated future opportunities. The Amazon Kindle is an example where only certain applications are allowed to be downloaded by a specific device. Similarly, machine-to-machine communications, while not in mass adoption today, will require that the Internet-connected “machines”, such as a stop light or traffic camera, only transmit certain application data. The Internet access services for eBook readers and traffic lights require application-specific traffic management to ensure that the connections aren't hacked for general use, which would in turn break the business model.

Throughout its reports BEREC emphasizes the importance of case-by-case analysis of differentiation practices and stays away from any ex ante judgments. We hope the examples above help to demonstrate the importance of maintaining that consistent approach when evaluating the reasonableness of application-specific traffic management.

Bolting-on ARPU and QoS through Differentiated Services

BEREC has acknowledged that ISPs may need to develop new revenue streams in order to provide ongoing incentives for investment, as ISPs traditional revenue streams, such as voice telephony and instant messaging, continue to be eroded by the over-the-top applications offered by Content and Application Providers. Differentiation of service offerings is one important solution to this challenge.

Sandvine has had significant experience implementing such plans for ISPs globally. One model that has resulted in success for our customers is with "bolt-on" offers that enhance unrestricted Internet Access Services. In this scenario, users all have unrestricted access to all Internet services, but can pay a fixed monthly charge for *unlimited* access to their favourite services.

A representative ISP that implemented such a plan realized a number of benefits, using this application-based strategy. Their subscriber base grew significantly and traffic from new subscribers, in combination with increased usage by existing users, resulted in an increase in network usage of over 200% during the quarter following the rollout. Subscribers were attracted to price certainty for the applications that they valued the most. It also eliminated the need for subscribers to track byte consumption and monitor quota thresholds – they no longer needed to calculate how many bytes a YouTube video consumed, or how many emails fell within the quota allowance.

For ISPs, tiered price plans of this nature have a natural positive impact on congestion management policies given their ability to align the operator's earned revenues with the cost structure necessary to support those services. In other words, achieved margins cover any necessary expansion costs. A full customer case study for one service provider that implemented this type of plan is available: http://www.sandvine.com/general/document.download.asp?docID=7

Interestingly, this differentiated service, while application-specific, does not involve differentiated treatment of the underlying application traffic. Instead, it involves differentiated *charging* in exchange for price certainty for the end user - conditions that should satisfy Network Neutrality goals. Over time, it is conceivable that such bolt-on services could subsidize a reduction in the base price of ISPs unrestricted Internet Access Services, or fund improved QoS for such services through network expansion.

Measuring QoS

BEREC suggests that quantitative QoS requirements can be set for different applications and Sandvine concurs. Yet, BEREC has concerns about whether such requirements would be feasible in a best efforts Internet. Application-specific traffic management can help alleviate that concern. Network traffic can be managed to optimize the performance of applications according to their (and end-users) unique needs, e.g., prioritizing time-sensitive traffic.

Sandvine has suggested such application-specific quality thresholds in connection with a U.S. Federal Communications Commission public notice on broadband measurement, and repeats them here.

For the purposes of quality measurement, Sandvine defines three application categrories: Bulk, Interactive and Paced/Burst-paced.

Bulk. These applications include P2P filesharing (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, accelerating until packet loss occurs. TCP is designed to achieve the maximum communication rate possible. 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 collocated 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) will saturate the download speed of the consumer's modem, as they typically download-only. In the case of P2P filesharing, it is bi-directional so it can also have the same affect 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. The content is typically for offline consumption. Bandwidth is the primary determinant of transfer speed 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 is used by the network to control the maximum achieved speed.

Web surfing represents an exception in the Bulk category. "Web 2.0" sites have introduced interactive components to web surfing – typically the user interacts with the website and expects a near-immediate response. Data is traveling bidirectionally as users have become content providers in their own right, by posting videos to YouTube, for example. Increases in bandwidth do not translate linearly to increased performance because it takes several "round trips" between a personal computer and the related web servers to load a website – typically at least four: the Domain Name Server (DNS) lookup[6] two for the three-way handshake established by TCP[7] and one to retrieve the content. 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.

Interactive. 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[8]. 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. Streaming applications such as YouTube and SHOUTcast fall into this category. The media involved has a natural bit rate based on the content's encoding, 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 when possible to build the buffer then reduce to the natural rate). 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 1-7Mbps 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.

---

[6] Based on IETF RFC 1035. See http://www.ietf.org/rfc/rfc1035.txt

[7] Based on IETF RFC 793. See http://www.faqs.org/rfcs/rfc793.html

[8] See T. Blajic, D. Nogulic, M, Druzijanic, *Latency Improvements in 3G Long Term Evolution*, p. 1-2, available at http://www.ericsson.com/hr/about/events/archieve/2007/mipro_2007/mipro_1137.pdf, or http://www.telephonyworld.com/training/brooktrout/iptel_latency_wp.html.

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 measurements can done by any network equipment reliable for this purpose and should take place as close to subscriber's premises as is practical in order to see all subscriber traffic.

The network measurements could be taken monthly over a one-minute interval in peak and off-peak times. Sandvine's own research has shown that peak hours are from approximately 7:30 pm to 10:00 pm in Europe. These hours could be reliably used to define "peak" and "off-peak" for the purpose of these measurements. Alternatively you could use the true peaks (as measured by available Business Intelligence solutions, like Sandvine's Network Analytics), which vary day by day.

The following table provides some representative benchmarks to achieve a minimum quality of service for certain popular applications.

Application Category	Application Class	Minimum Bandwidth	Maximum Latency	Maximum Jitter	Maximum Loss
Bulk	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		
Interactive	VoIP	16Kbps	300ms (latency + jitter)		< 0.5%
	Video gaming	50Kbps	75ms (latency + jitter)		< 0.5%
	Video Conferencing	250Kbps	300ms (latency + jitter)		< 0.05%
Paced (and burst-paced)	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: 128Kbps for CD quality. 56Kbps for radio	< 1s for "channel change"	<50ms	<0.05%

For transparency, ISPs could report on their performance for each application class using a stoplight model: green for meeting or exceeding the required performance, yellow for variably achieving that performance and red for underachieving.