

Submission to the
Office of the Privacy Commissioner of Canada

Re: Public Consultation on the
Privacy Implications of Emerging Technologies



**Privacy is About Use Cases,
not About Technology**

March 15, 2010

About Sandvine

Sandvine is a Canadian company located in the heart of the Region of Waterloo technology cluster. The Company was established in 2001 and employs over 250 people in Canada. Sandvine has thrice been named to the Deloitte Technology Fast 50 list of fastest growing technology companies in Canada over a five-year period: in 2007, Sandvine was the top company. The Company was identified in the National Post as one of Canada's Top 100 Corporate R&D Spenders, based on fiscal 2007 spending. For the last three years Sandvine has been named one of the top 50 "Best Workplaces in Canada" in Canadian Business magazine.

Sandvine is focused on protecting and improving the quality of experience on the Internet with its Network Policy Control solutions. Our award-winning network equipment and solutions help cable, DSL, FTTx, fixed wireless and mobile operators better serve their subscribers; understand network trends; offer new services; mitigate malicious traffic; manage network congestion; and deliver QoS-prioritized multimedia services. With broadband service provider customers in over 70 countries serving hundreds of millions of fixed and mobile subscribers, Sandvine is enhancing the Internet experience worldwide.

Introduction

Communications technologies have changed people's views of personal privacy for hundreds of years. The invention of the printing press allowed wide-scale distribution of information about public figures that was previously impossible, funding an industry of paparazzi and tabloid reporters who appealed to the public's prurient interests as a means of selling advertising. The rise of the consumer Internet has given unprecedented ease of access to information which once may have been considered private and personal, including information from newsgroup postings, personal blogs, social networking sites, and, in some cases, from unintentional information leakage or even intentional information theft. The level of information available about individuals which is accessible through a simple search engine would be considered astonishing compared to as recently as even 10 years ago. Society has always adapted to changes in technology with a give and take, modifying guidelines and accepted practices on information usage, and realigning expectations with respect to information privacy. Will the Internet continue this trend, or are privacy concerns and progress destined to oppose each other?

Legislation and Technology

Legislation has often struggled to keep up with technology. Where a new technology has created a perceived need for legislation, legislators have often tended to focus on the technology itself, rather than the use cases involved. In essence, they focus on writing the letter of the law when they should focus on the spirit. Take for example the case of American jurist Robert Bork. Other than being famous for acquiescing to Richard Nixon's will and firing special prosecutor Archibald Cox, he is known for being a candidate to the US Supreme Court. During the debate of his nomination, Bork's video rental history was leaked to the press, which in turn led to the enactment of the Video Privacy Protection Act. In this case, the law clearly did not stay abreast of technology, and was enacted for the narrow purpose of preventing information about VHS tape rentals from reaching the public, anticipating neither DVD rentals 10 years later, nor video on demand over cable, nor Internet-based video distribution. If society had acted to place guidelines on 'dissemination of entertainment preference information', which was the actual

intent, we would have been better served, rather than having legislation narrowly targeted to a specific technology.

Societal Expectation

Privacy is all about the expectations of those involved. As a consumer, I expect the content of my email message to be private between me and my intended recipient, regardless of whether I send it via my residential Internet service provider's mail server, or via a web-based service such as Google's Gmail or Microsoft's Hotmail. If this email were to be used by anyone other than my intended recipient, my expectation of privacy would not be met, regardless of whether this unauthorised use was facilitated by the Internet service provider I am using or by a web-based service I am using. The societal expectation of privacy applies to the use of the information, not the method or point of interception. To allow a model where a web-based provider of email services can read my email, and use the contents to build a profile of me for advertising purposes, but not allow an Internet service provider (also with my consent) to do the same thing is to create an imbalance in my expectations that the majority of email users do not appreciate. Privacy use cases should be viewed through the expectations of the information originator, not through the specific narrow methods which are used to gather the information.

Demonising Technology

Throughout history, control of terminology has been used as a method of setting agendas and inciting preconceived conclusions on the basis of nomenclature alone. It appears that this is becoming true again in the current privacy debate regarding the term "deep packet inspection" and its acronym "DPI".

DPI is necessary for the identification of traffic today because the historically-used "honour-based" port system of application classification (where applications identified themselves by the port number over which its data was communicated) no longer works. Essentially, application developers have either intentionally or unintentionally designed their applications to obfuscate the identity of the application. Today, DPI technology represents the only effective way to accurately identify different types of applications. DPI is, from a network engineering and architectural perspective, the act of any network equipment which is not an endpoint of a communication using any field other than the Layer 3 destination IP address for any purpose.

Each layer in the seven-layer Open Systems Interconnection (OSI) Reference Model has a header and payload, all the way up through Layer 7 and beyond, and networking equipment has always read Layer 7 "payload" data. For example, mail servers route mail based on the e-mail address, which is located in the Layer 7 payload data. Session Initiation Protocol (SIP) is a signaling protocol widely used for setting up and tearing down multimedia communication sessions such as VoIP. SIP needs to look in the Layer 7 payload data to find both phone numbers involved in a VoIP conversation, then set up the data (voice) flow. Routers/firewalls look at the Layer 7 SIP exchange to extract this flow information to let the data through. If they don't, the voice component is blocked. DPI is also a key part of the innovation in allowing a migration from IPv4 to IPv6, allowing a network operator to convert from one to the other using a carrier-grade network-address-translation (NAT), and keeping protocols such as VoIP operational.

In other words, DPI is a critical, ubiquitous and time-tested technology. Banning the use of DPI, as some have suggested is necessary based on privacy implications, would have far-reaching and damaging consequences across the Internet, where the technology is used extensively. Instead, when considering the privacy implications of DPI, as with any technology, the focus should be on the use case, not the technology itself.

DPI-enabled Network Policy Control Solutions Don't Inspect User Content

Sandvine submits that the true “content” of an Internet transmission is represented as the body of your e-mail message; the music or movie you are downloading; the video you are streaming; the words in your VoIP call, etc. DPI is most commonly used as part of Network Policy Control solutions, which establish policies in broadband networks to manage congestion, mitigate malicious traffic and deliver on the terms of subscriber service plans. None of these solutions inspect content as, quite simply, the content is not relevant to the decisions that the solutions need to make. To be clear, such solutions:

- Do not read your e-mail;
- Do not listen to your voice calls;
- Do not watch the video you are streaming, etc.

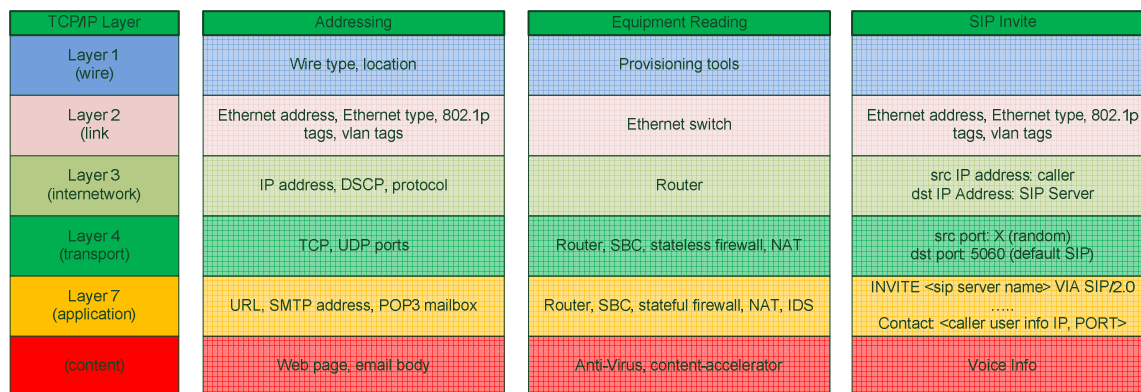
DPI: a Passive Inspection Technology

DPI is a passive inspection technology that uses two primary inspection techniques: (i) behavioural flow-based; and (ii) signature-based. The techniques can be used separately or together – whichever most efficiently and effectively achieves the appropriate level of accuracy in application identification.

With behavioural flow-based inspection, which Sandvine uses to identify malicious traffic and certain encrypted protocols, the packet itself is not inspected – at all. Instead, the *behaviours* of packets in a flow are analyzed to identify matches with known application behaviours. For example, a denial of service attack can be readily recognizable by the pattern of “hits” on its target server. As no inspection of the actual packets takes place, there is no opportunity to view any content.

Behavioural flow-based inspection has limitations. On its own it is only sufficiently accurate in identifying classes of applications (e.g., bulk applications, like P2P or FTP for file-sharing, or interactive applications, like video gaming or VoIP) rather than down to the individual protocol level (such as SIP). Accordingly, any network policies that rely on the technique may be overly broad and affect more users, applications and protocols than absolutely necessary to achieve the policy goal. Additionally, any network reporting that relied solely on behavioural flow-based inspection would be too vague to allow for useful network planning.

With signature-based inspection, a library of known application “signatures” is compared to a packet to identify matches. By way of example, the diagram below shows the breakdown of a SIP VoIP packet against the OSI Reference Model.



In most cases for SIP, the caller will contact the SIP server over port 5060. However, a SIP server can, and often is, configured to work on different ports. Thus, to accurately identify this traffic, a signature based on the IETF RFC standard for the SIP protocol is applied to the packet. In this example, and as shown in the fourth column of the diagram, at Layer 7 the solution looks for the presence of “INVITE” followed by some server address then “VIA SIP/2.0”. By examining this protocol’s header, the solution is able to determine whether the flow is SIP. The solution does not look at the flow’s contents, i.e., the voice information, as it is not required to make the protocol identification.

If there is no match, then the solution immediately forgets the inspected data and compares the next signature definition in its library to the packets being inspected. The entire packet is not scanned, as if browsing through a magazine. Instead, only those locations that hold identifying signature characteristics are inspected and only to the extent necessary to see if there is a match with the signature profile in the library.

In either case, the DPI device never records any of the information past the life of the detection, other than the identity of the protocol, and it only uses this information as an input to decide whether it is relevant for the application of a network policy. The process is similar to a mail-sorting machine: the address is matched, the decision is made and the address is then forgotten.

For both behavioural flow-based and signature-based inspection, once identification has occurred further inspection not only stops, but the attributes examined in the process of arriving at that identification are discarded. For signature-based inspection, identification can typically happen in the first couple of data packets in a stream. More often than not, those first few data packets don’t contain data that would typically be considered the content of a transmission, such as the text in an e-mail or the voice in a VoIP call, etc. For example, for a SIP-based VoIP call the first two data packets would be part of the “control flow”, which is used to establish call permissions and locations, etc., to initiate the call. Data from the actual conversation would only appear in subsequent packets.

Network Policy Control Solutions don't Keep Personal Data

Because typical Network Policy Control solutions do not inspect the actual content of users’ Internet traffic, they also cannot record, report on, or store such personal information. The most “personal” information that these solutions record for a given Internet account (i.e, not a particular individual, but the IP address attached to an Internet account, which may include

access for many individuals) is aggregate volume usage data, by application or protocol. For example, Network Policy Control solutions could report the number of bytes of a VoIP protocol sent and/or received by a given Internet account over a fixed period.

Applications of Technologies may Raise Privacy Issues, not Technologies Themselves

Despite the fact, that DPI does not inspect or retain subscriber “content” suggestions have been made that somehow the mere presence of DPI-based technology raises privacy issues. Some have called for an outright ban. Imagine if this approach were applied to other technologies, such as those supporting cameras. Single Lens Reflex (SLR) technology underlies cameras that can be used to take photos at family birthday parties or applied for surveillance of individuals and public spaces. One use of the technology raises privacy issues, the other does not. Nobody questions the value or validity of SLR technology. DPI technology should not be questioned either. Privacy concerns properly attach to applications or uses of technologies, not to the technologies themselves.

Privacy-sensitive Solutions are in Demand

Sandvine recognizes that certain solutions that are unrelated to Network Policy Control (such as lawful intercept, copyright enforcement, and targeted advertising) may raise personal privacy considerations and are in high demand from consumers, governments and society. Such solutions are achieved through a variety of technologies, not just — or even predominantly — DPI.

Targeted advertising provides a good example. This type of solution can enhance the Internet experience for subscribers by presenting them with more relevant advertising information. Typically, targeted advertising solutions monitor user Internet activities, such as detailed analysis of website visits, to create a more complete user profile for enhanced marketing. The collection and storage of that type of profile information has privacy implications for users.

DPI technology *can* comprise a component of targeted advertising solutions, but it has been *very rarely used this way*. Instead, other technologies have dominated. Google is one of the leaders in targeted advertising, but to Sandvine's knowledge Google's targeted advertising solutions do not use DPI.

According to Google's own Advertising and Privacy notice in connection with its enormously popular Gmail e-mail application, Google reads your mail to make decisions on targeted advertising.

“The Gmail filtering system also scans for keywords in users' emails which are then used to match and serve ads. When a user opens an email message, computers scan the text and then instantaneously display relevant information that is matched to the text of the message.¹”

Other Google applications have privacy implications. According to Google's Web History Privacy Notice:

¹ Google, “Google Privacy Center, Advertising and Privacy,” see http://www.google.vg/intl/en/privacy_ads.html

“Web History records information about the web pages you visit and your activity on Google, including your search queries, the results you click on, and the date and time of your searches in order to improve your search experience and display your web activity. Over time, the service may also use additional information about your activity on Google or other information you provide us in order to deliver a more personalized experience².”

Google's PageRank service operates by “sending Google the addresses and other information about sites at the time you visit them.”³ According to Google's Privacy FAQ, Google stores certain information about your searches for as much as 18 months prior to anonymizing it⁴. Again, to Sandvine's knowledge, none of these solutions use DPI.

Lawful intercept provides another example of how privacy-sensitive solutions can be enabled by a wide variety of technologies. In the United States under the Communications Assistance for Law Enforcement Act (CALEA), service providers are required to identify and intercept criminal data traffic under a lawful warrant provided by law enforcement agencies. DPI technology could be used in a solution designed to support the collection of that data, but so too could a home computer “tapped” into the communications of the individual that is the subject of the warrant.

Again, applications of technologies may raise privacy concerns, not technologies themselves. Sandvine urges the Office of the Privacy Commissioner to adopt a technology-neutral view when considering privacy on the Internet.

DPI Enables Adequate Consent for Privacy-sensitive Applications

In many cases, questions around privacy-sensitive Internet applications will ultimately come down to the ability to secure sufficient user consent. To date, vendors of privacy-sensitive applications like targeted advertising have struggled with providing reliable mechanisms for managing user consent. The mechanisms, whether designed as opt-in (where the user must proactively consent to being subject to the solution) or opt-out (where the user must proactively demand NOT to be subject to the solution) have typically been cookies-based. Cookies are pieces of text, stored by a user's web browser that contain the user's settings or other data used by websites.

There are significant problems with a cookies-based system. Cookies can be cleared by the user (purposely or inadvertently), which then erases the “opt-in” or “opt-out” permissions related to a privacy-sensitive application. Also, cookies are associated with a particular computer's Internet browser, not the user's Internet account. So, if a subscriber uses his Internet account from multiple computers the targeted advertising permissions stored in the cookie do not follow the user between computers. Similarly, if the same user has multiple browsers on the same computer (e.g., Internet Explorer and Firefox), the targeted advertising permissions stored in the cookie do not follow the user between browsers.

² Google, “Web History, Web History Privacy Notice,” see <http://www.google.vg/searchhistory/privacy.html>

³ Google, “Install or uninstall: Toolbar Privacy Notice,” see <http://www.google.com/support/toolbar/bin/static.py?page=privacy.html&hl=&v=>

⁴ Google, “Privacy FAQ,” see http://www.google.com/privacy_faq.html

Fortunately, a better solution to the consent problem is available: through a network-level association between the subscriber's account and his permission settings related to the privacy-sensitive applications. Regardless of the computer he uses to access his Internet account or the browser that he uses on those computers, the permissions follow the user. Only if the user intentionally changes his account-level privacy permissions could a previously opted-out user be opted-in. Such a solution could be implemented through a solution that incorporates DPI technology.

DPI-supported Network Policy Control Solutions Offer Consumers More Choices and Create Competition

Network providers are just beginning to explore the use of Network Policy Control to help them create service offerings that are more attractive to consumers in an increasingly competitive Internet access market. In North America, high-speed Internet services have traditionally been offered in the form of flat-rate, monthly plans. Consumers may be interested in other types of service plans that better reflect the unique ways that they use their Internet connections. Such plans would necessitate the ability to differentiate between the traffic of individual subscribers, and between applications – they would require DPI.

For example, “light” Internet users may be interested in a service package that ties their fees to the bytes they consume on the network. But would these consumers want to pay for malicious traffic that affected their usage in a month, or visits to the service provider's web service portal to address service issues? A user- and application-specific policy would be required to manage the plan. By contrast, disproportionately heavy users likely don't want to pay “by the byte”, but they may be interested in a service plan that provided a financial incentive to shift their activity to non-peak network hours. Such a plan would help all users by freeing up more capacity at peak times, when network congestion and application degradation is most likely to occur. DPI can help here.

Other consumers may value their Internet connection by the quality of experience they receive for their favourite applications, like latency-sensitive Internet video gaming and over-the-top VoIP. Network providers could offer a Premium Video Gaming or Premium VoIP service plan that delivers exactly the type of Internet experience these consumers want. Such plans would need to be supported by application-specific and user-specific policies enabled by DPI.

New service plans like these would represent significant innovations and require significant investments in network engineering (and marketing) by network providers. They would also offer consumers new choices and in so doing create new grounds for competition among network providers.

DPI-supported Network Policy Control Solutions use IETF-approved Techniques

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. DPI is an inspection technology, and while there are no IETF standards for inspection of Internet traffic many IETF standards implicitly require the use of DPI, such as RFC 3489, “Simple Traversal of User Datagram Protocol (UDP) Through Network Address

Translators (NATs)”⁵, and RFC 2766, “Network Address Translation - Protocol Translation (NAT-PT)”⁶.

Also, IETF's RFC 4594⁷, “Configuration Guidelines for DiffServ Service Classes,” suggests that there should be different prioritization for different applications depending on their sensitivity to delay, loss and jitter. They suggest the following categories of services: telephony, telephony/video signalling, multimedia conferencing, real time interactive, broadcast video, low latency data, high throughput data, and low priority data. The RFC continues to discuss the sensitivity of each of the application categories to network conditions. Only through the use of DPI could these applications be reliably identified and therefore receive appropriate treatment in the network.

Canada Can Continue to Lead

Before the Federal Communications Commission (FCC) in the United States launched their “Open Internet” Notice of Proposed Rulemaking in October 2009, the Canadian Radio-television and Telecommunications Commission (CRTC) initiated and concluded a very similar Review of Internet Traffic Management Practices. While the CRTC created no new rules, they provided some guidelines to help service providers craft their network policies. The FCC has an opportunity to follow this sensible lead. Canada has the same opportunity to lead on the issue of Internet privacy.

In the summer of 2008 a special committee of the US congress invited testimony on the subject of behavioural targeting in Internet-based advertising. The specific technique investigated was, perhaps inadvertently, labelled as DPI. Rather than focus on the use cases (e.g. whether it was acceptable to build a profile of a user for the purpose of targeted advertising), the technology itself became the focus of the examination. It appears that as a result of this inquiry and the press coverage and commentary arising from it, in the public’s mind, all uses of DPI now somehow, by definition, involve privacy invasion, rather than just those that go into specific content and use the information.

The public would be best served by guidelines for *online information usage*, rather than for the means of information collection. From an end-user perspective, it does not matter whether someone builds my profile by looking at packets on the wire, or by placing cookies on the web sites I visit. Both yield the same result: a third-party has a model of my interests and behaviours.

Humans are adaptable. Society will evolve. Our concept of privacy in the information age will change over time, and our expectations of what uses are private will become clearer. If we focus on the use cases in a user-centric fashion, rather than the techniques or technologies, any future guidelines will be easier to convey and enforce. Society is not well served by a narrow focus like protecting the privacy of video cassette tape rentals, nor is it well served by trying to prevent a technology because it has concerns with one of the use cases the technology enables. DPI is needed for our continued innovation of the Internet. Let us focus on making our future spirit of privacy expectations clear rather than limiting our attention to one particular means.

5 See <http://www.faqs.org/rfcs/rfc3489.html>

6 See <http://www.faqs.org/rfcs/rfc2766.html>

7 See <http://www.ietf.org/rfc/rfc4594.txt>