Richard Bennett
661 Ruby Rd.
Livermore, CA 94550

April 21, 2009

The Honorable Chairman Rick Boucher
The Honorable Ranking Member Cliff Stearns
The Honorable Chairman Henry A. Waxman
The Honorable Ranking Member Joe Barton
The Honorable Members of the Subcommittee on Communications, Technology, and the Internet

Subject: Testimony before the House Committee on Energy and Commerce Subcommittee on Communications, Technology, and the Internet hearing on April 23, 2009

Dear Chairman Boucher, Ranking Member Stearns, and members of the Subcommittee,

Thank you for offering me the opportunity to address the subcommittee on the subject of technologies that monitor consumer use of communications networks. The topic is pertinent to the evolution of the networks, to the development of consumer awareness, as well as to potential new regulations if they're needed. I'd like to offer a few recommendations.

Network Monitors

I'm a network engineer, inventor, and writer. I've designed data transfer and Quality of Service protocols for some of our most widely used communications networks, switched Ethernet and Wi-Fi, as well as for some that haven't got off the ground yet, such as Ultrawideband. As a consequence, I've had occasion to use a variety of network monitoring and analysis equipment to observe traffic on networks.

Network monitors – often called "Sniffers" after a popular product produced by Network General in the 1980s – enable engineers to see every part of the packets that traverse the network segments to which the monitors are attached, including the various payloads present at the Ethernet, IP, TCP, HTTP, and Content layers, for example. These are vital tools that permit programmers and electrical engineers to accelerate systems and isolate and correct bugs that would otherwise limit network function. These systems pre-date the Internet by many years, and it's safe to say that we would have no working packet networks without them.

These devices have political uses as well – the controversy over Comcast's first-generation traffic shaping system was set off by a network technician who used an open source network monitor to discover suspicious TCP packets mingled in amongst the peer-to-peer file sharing packets he expected to see. Since the 1980s, these devices have had the ability to apply filters to network traffic based on sophisticated pattern matching, to produce logs of selected packets, and to perform a variety of statistical analyses to network traffic streams. They are frequently used by network administrators to troubleshoot problems in both local and wide-area networks, and are generally considered to be invaluable aids in maintaining the semblance of stability that users expect of their networks.

While these monitors have been used on occasion to steal passwords and other user information, these instances are rare and limited in scope simply because an Ethernet monitor can only be used to capture traffic on the particular part of the network to which it is attached. If I monitor the

traffic on my home network, as I frequently do, I can't see any of the traffic generated by my neighbors, even though we share a common coaxial cable to a shared CMTS; this is because my cable modem only passes traffic intended for my Internet access account. The only clue I have to my neighbors' usage is the delays that my traffic encounters on the way up and down the cable, and that only tells me how busy they are, not what sites they're visiting and which files they're downloading.

To obtain that level of information, I would have to use a Wi-Fi sniffer such as Air Pcap and hope their Wi-Fi networks are either completely unsecured or that they rely on an effectively useless cipher such as the deprecated *Wired Equivalent Privacy* standard known as WEP.

Anyone who uses a Wi-Fi network in a populated area without securing it with WPA or WPA2 is effectively sharing his personal web surfing and e-mail habits with any snoop who cares to hear them. This situation is intolerable to me, so I joined colleagues in the Wi-Fi Alliance in developing a system for quick and easy setup of secure Wi-Fi networks called Wireless Protected Setup or WPS. I hope all of you who use Wi-Fi understand that you're broadcasting your web surfing habits to anyone who cares to learn them if you haven't secured your networks. If you're forced to use an unsecured Wi-Fi network in exigent circumstances, you can provide yourself a measure of privacy by securing your e-mail connection with Transport Layer Security. Public Ethernet connections are also fundamentally insecure, as anyone connected to the same switch fabric you're connected to can easily capture your packets and examine them to his heart's content.

As a purely technical matter, there's no difference between the means that Wi-Fi engineers use to diagnose network problems and those used by snoopers on public Wi-Fi networks to steal passwords: the same packet capture tool can do both. But one activity is legitimate (and even necessary to the proper functioning of networks) and the other is not.

So my first recommendation to the committee is to **emphasize intent and behavior rather than technology** in its continuing efforts to protect communication privacy. Technologies are neither good nor bad, it's the uses we put them to that matter.

The Culture of Over-Sharing

Another threat to consumer privacy, and in my mind a much greater one, is what I'll call the Culture of Over-Sharing. With the advent of personal web sites, blogs, social networks, and Twitter, people are sharing information about themselves that would certainly make their grandparents blush. I follow a number of tech journalists on Twitter, and I can now tell you more details of their personal health, diets, and dating habits than about the stories they cover or the conferences they attend. I don't particularly care for this personal information, but it's a part of the package.

Stories abound about young people who've posted drunken party pictures of themselves while they were in college finding embarrassment, often costly, when they apply for jobs and have to explain their antics to Google-savvy recruiters. The Internet is a harsh mistress, and much of what happens there stays there, seemingly forever.

I've been operating a series of blogs on technology and politics since 1995, and recently have received a number of requests from past commenters to remove missives they posted to a blog a few years ago. One recent correspondent said his roommate had posted radical sentiment under his name (I have no way to verify one way or another,) and another admitted frankly to being

young, reckless, and grammatically challenged when posting comments that he now feels make him less employable. So I've adopted a policy of removing older comments for any reason at all. The lesson that I draw from this is that **retention policies are critically important to privacy**. It's the nature of networks to disseminate information, public and otherwise, but the game doesn't change radically until past, present, and future are combined into large, searchable archives that holds us captive to our pasts forever. People, especially those who were young once, need to have the ability to reinvent themselves, and our culture of over-sharing combined with our massive Internet archives, is eroding it.

Consumer Education

I've alluded to consumer awareness, or the lack thereof already, but I'd like to emphasize it as there have been recent instances of inadvertent sharing. CNet News reports[1] that the Committee on Oversight has heard testimony on the following events:

- On February 28, 2009, a television station in Pittsburgh reported that the blueprints and avionics package for "Marine One," the President's helicopter, was made available on a P2P network by a defense contractor in Maryland.

- On February 26, 2009, the Today Show broadcast a segment on inadvertent P2P file sharing, reporting that social security numbers, more than 150,000 tax returns,25,800 student loan applications, and nearly 626,000 credit reports were easily accessible on a P2P network.

- On February 23, 2009, a Dartmouth College professor published a paper reporting that over a two-week period he was able to search a P2P network and uncover tens of thousands of medical files containing names, addresses, and Social Security numbers for patients seeking treatment for conditions such as AIDS, cancer, and mental health problems

- On July 9, 2008, the Washington Post reported that an employee of an investment firm who allegedly used Lime Wire to trade music or movies inadvertently exposed the names, dates of birth, and social security numbers of about 2,000 of the firm's clients, including Supreme Court Justice Stephen Breyer. There have been reports alleging file sharing programs have been used for illegal purposes, such as to steal others' identities.

Technology always moves faster than regulation, and we want to keep it that way, but consumers need to be aware that some of the applications they run, particularly peer-to-peer file sharing applications, expose more information than they may want. It's unlikely that producers of Peer-to-Peer applications will be responsive to Congressional mandates of full disclosure; theirs is a quirky community with little regard for authority, but steps can be taken to make consumers aware of the dangers of inadvertent over-sharing.

Malware and Botnets

Perhaps the most significant threat to consumer privacy is deliberate identity theft. By now, this threat is well-understood: millions of computers worldwide are infected with viruses that put them under the effective control of the virus' creators. Infected computers, tied together in a huge *botnet*, are used to send Spam and to run key-loggers that steal personal information. The end

[1] Greg Sandoval, "Congress to Probe P2P Sites over Inadvertent Sharing," *CNet News*, April 21, 2009: http://news.cnet.com/8301-1023_3-10224080-93.html?part=rss&subj=news&tag=2547-1_3-0-20

produce is sent back to the controller where it's used for criminal purposes. It's suspected that some botnets may be controlled by foreign intelligence services as they're shown up in interesting places, such as the Dalai Lama of Tibet's offices in India. While Spam is in integral part of the Internet's e-mail system today, and will remain so as long as we don't adopt a system of user authentication as part of normal e-mail practice, efforts to mitigate its effects are impressive.

Spam fighters maintain a set of DNS Blacklists which squelch, by their estimation, some 81% of attempted Spam at the source, simply by checking the Internet Domain Name of the source networks against a list of known Spam networks. This is a very important function, but it raises the shackles of some privacy advocates, who see it as discriminatory and non-transparent. DNS Blacklists certainly do contain false positives from time to time, but they incorporate procedures for the removal of domains unfairly listed. The value of this kind of Spam mitigation is enormous, and it goes beyond the protection of consumer privacy: Spam has a considerable carbon footprint and contributes to global warming. According to a recent report by McAfee, Inc[2]:

- *An estimated worldwide total of 62 trillion spam emails were sent in 2008*
- *The average spam email causes emissions equivalent to 0.3 grams of carbon dioxide(CO2) per message*
- *Globally, annual spam energy use totals 33 billion kilowatt-hours (kWh), or 33 terawatt hours (TWh). That's equivalent to the electricity used in 2.4 million homes, with the same GHG emissions as 3.1 million passenger cars using two billion U.S. gallons of gasoline.*
- *Spam filtering saves 135 TWh of electricity per year. That's equivalent to 13 million cars off the road*
- *Much of the energy consumption associated with spam (nearly 80 percent) comes from end users deleting spam and searching for legitimate email (false positives). Spam filtering accounts for just 16 percent of spam-related energy use.*

Clearly, Spam mitigation is a social good. The Blacklist method isn't sufficient on its own; it's driven by intelligence about which e-mail messages are Spam and which aren't. This determination is made by a number of means, one of them human intelligence, but machines are part of the process as well. Mechanical recognition of Spam depends on a process of pattern matching e-mail against known contents of Spam currently in circulation in the Internet. Like anti-virus software, Spam detectors search for Spam signatures in ordinary e-mail, flagging or deleting suspect messages. This is in fact a very invasive process, one that can often cause legitimate messages to end up in user's spam folder or worse. But it's a system that Internet users embrace because its benefits far outweigh its drawbacks.

The lesson I suggest we should learn from Spam mitigation is to **examine mechanical processes for their practical benefits as well as their theoretical harm** to abstract notions of privacy, and to consider what our networking experience would be like without them. The damage to personal privacy inflicted by Spam signature searches has to be balanced against the greater harm that unchecked Spam inflicts. Similarly, an e-mail ethos based on personal identity rather than semi-anonymous access has benefits that are not lost on the architects of Internet e-mail. Future systems will surely be designed in more robust manner.

---

[2] *The Carbon Footprint of Email Spam Report*, McAfee Inc. and ICF International, http://img.en25.com/Web/McAfee/CarbonFootprint_28pg_web_REV.PDF, retrieved April 21, 2009.

Traffic Engineering

Contrary to popular belief, the physical networks that carry Internet Protocol packets are not "stupid" networks. Most IP networks of significant size carry a combination of generic Internet traffic and private IP traffic that has to be delivered according to Service Level Agreements (SLAs) between end-user organizations and network carriers. Some SLAs are very stringent, allowing for as little as 2 milliseconds of latency (delay) between transmitter and receiver. In order to satisfy the needs of customers with varying SLAs, network operators buy network equipment that's capable of prioritizing packets. These systems depend on the ability to classify network flows[3] and to count packets per second over long periods of time. Their prioritization function interacts with accounting and policy functions to promote or demote specific flows depending on the customers standing in terms of volume and rate and his contract. Traffic engineering of this sort, generally using the MPLS[4] protocol which reduces the overhead of repetitive route lookup as the packet moves from one router to another, is at the heart of the modern Internet.

A simplified form of traffic engineering is now employed by Comcast on its residential broadband network to protect IP service from overload. When a link has been congested for a meaningful period of time, the system identifies heavy users of network resources (bandwidth.) Any of these users who've exceeded a meaningful threshold are placed in a lower priority category until the load they offer the network declines. This system is notably "protocol agnostic" as it treats all Internet applications the same: if a user is engaged in a large file transfer for a significant period of time (15 minutes or more) that places him in the low priority category, and if the user is also using Skype or some other VoIP service, his VoIP performance will suffer until he takes steps to curtail his downloading.

This system addresses one of the fundamental architectural shortcomings of the Internet, the absence of a per-user fairness system. This problem has been addressed in numerous forms, and perhaps most clearly by Dr. Bob Briscoe, Chief Scientist at British Telecom Research[5]:

> *Resource allocation and accountability keep reappearing on every list of requirements for the Internet architecture. The reason we never resolve these issues is a broken idea of what the problem is. The applied research and standards communities are using completely unrealistic and impractical fairness criteria. The resulting mechanisms don't even allocate the right thing and they don't allocate it between the right entities. We explain as bluntly as we can that thinking about fairness mechanisms like TCP in terms of sharing out flow rates has no intellectual heritage from any concept of fairness in philosophy or social science, or indeed real life. Comparing flow rates should never again be used for claims of fairness in production networks. Instead, we should judge fairness mechanisms on how they share out the 'cost' of each user's actions on others.*

The Internet is a system built on the dynamic sharing of network bandwidth, but it lacks a general-purpose mechanism of allocating it across user accounts fairly. Because the Internet lacks this vital mechanism, it's necessary for network operators to supply it themselves, as they have since the first deployment of Internet Protocol in a wide-area network by Ford Aerospace in 1981.

---

[3] A "flow" is a series of packets between a common source and destination.
[4] E. Rosen, A. Viswanathan, and R. Callon, *Multiprotocol Label Switching Architecture*, January 2001, IETF RFC 3031, http://tools.ietf.org/html/rfc3031
[5] Bob Briscoe, *Flow-Rate Fairness: Dismantling a Religion*, http://www.cs.ucl.ac.uk/staff/B.Briscoe/projects/2020comms/refb/fair_ccr.pdf

While the network engineering community is acutely aware of the limitations of the Internet's architecture and protocol design, advocates for open access and related causes often gloss over this issue in their search for the perfect network. The network technician who discovered the original Comcast system for managing P2P by injecting TCP Reset packets complained that the user-volume-based system amounted to "discrimination based on user-history [sic][6]." If that's the case, it's a brief history, no more than 15 minutes long.

The lesson to be learned about traffic engineering is that the realities of business and the shortcomings of the Internet as a global system for multiple uses often collide with utopian desires for the more perfect network. In very real sense, the TCP/IP Internet remains a work in progress 35 years after it was proposed as a research network for the exclusive use of highly-trained network engineers, professors, and graduate students. It was somewhat unfortunate that it was pressed into service in a completely different role in the early 90s when the plethora of personal computers demanded interconnection. Compromises against ideals of network function are inevitable in this scenario, and should not automatically be judged failures simply because they violate abstract notions of network design that have never been more than pipe dreams.

Standards bodies continue to address the Internet's need for continual improvement, and researchers are hard at work on several projects that would replace the current Internet with an improved network that reflects some of the knowledge we've gained in the last 35 years. In the meantime, I would urge the Committee **not to hold Internet operators to unrealistic standards**. Keeping the Internet running smoothly is a difficult task in the best of times, and any practice that has a plausible connection with this goal should be seen as constructive and responsible, even if it requires accounting for usage and acting accordingly. In the long run, traffic management systems that rely on accounting and prioritization are much friendlier to innovation than those that simply charge for usage.

Deep Packet Inspection

Discussions like this one inevitably come to Deep-Packet Inspection, that poorly-defined term that seems to portend something ominous ("it was a dark and stormy night for IP packets.") [7] As I've endeavored to show, there are legitimate and illegitimate uses for most aspects of network technology, and this is no exception. If we recognize that much of the traffic on the Internet is digital piracy (it doesn't matter how much as long as we agree that it's significant,) we have to accept that some means of mitigation is appropriate, just as it is for Spam, viruses, and overload. The most effective means of piracy mitigation – other than jail time for the operators of piracy-enabling web sites and tracker services like The Pirate Bay – is a system in which piracy cops enter swarms of users downloading and sharing digital material in a manner contrary to law. This system doesn't rely on DPI, as it simply uses non-encrypted information made available by the

[6] Robb Topolski, *Re: [p2pi] Follow-Up from Comcast Presentation*, e-mail to IETF P2PI, June 6, 2008. http://www.ietf.org/mail-archive/web/p2pi/current/msg00072.html

[7] A great deal of the animosity against DPI seems to stem from the belief that a functional layering approach to communications regulation should replace the current model, which FCC Commissioner McDowell has described as "technology silos." The silo model is defective because it focuses on technologies rather than services, and breaks down in the face of the similarity between video, IP transport, and voice services delivered across multiple technologies. Unfortunately, the functional layering model simply rotates the silo 90 degrees, and retains multiple ambiguities due to the fact that networks often perform similar functions – such as retransmission and error detection – at multiple layers. The service and disclosure model currently used by the UK telecom regulator, Ofcom, is far superior to either.

pirates, perhaps inadvertently. Encryption doesn't make this any harder to do, as the fabric of P2P piracy is sharing known content with random partners across a network. These exchanges revolve around a content identifier know as a file "hash" which is computed across the entire range of a file. File hashes can be extracted from certain P2P transactions automatically, and these transactions can point the piracy cop toward trackers who may not have been known at the outset. Hence, DPI has a role, albeit a limited one, in piracy mitigation. As long as digital piracy is against the law, there has to be some accepted means of finding it and stopping it. This needn't involve door-to-door searches or trips to Guantanamo Bay, but it's not simply a matter of sitting on our hands and saying, as the founders of The Pirate Bay said after their conviction in a Swedish court, that anything easy to do should be legal[8].

DPI can also be useful as a means of relaxing per-user quotas imposed by a fairness system, to better tune network service to application requirements. In a more perfect Internet – the one envisioned by the architect of the IP datagram's Type of Service field, the architects of Integrated Services, and the designers of Differentiated Services – applications should be able to communicate requirements to the network, and the network should do its best to meet them according to the service level that a user has purchased. It's for this reason that the Internet Protocol's header structure includes a field for signaling such requirements to the network. Unfortunately, in the transition from research to production network, this signaling was overlooked. Moving the Internet off the NSF Backbone and onto a mesh of private networks required the invention of a new protocol for service providers to communicate routing information with each other. This protocol – Border Gateway Protocol (BGP) – did not include a mechanism for attaching Quality of Service levels to routes. Private IP networks overcome this problem by adopting MPLS and using Ethernet VLANs, but the problem of communicating QoS levels in the public Internet remains unresolved. There is hope that a draft pending before the IETF's Inter-Domain Routing Working Group provides the solution[9]. The creator, a professor at the Chemnitz University of Technology, has tested his solution in number of public Internet exchanges in Europe and reserved the necessary numbers from ICANN.

In the meantime, the most effective way of determining application requirements is to examine streams and map them to QoS categories by their evident properties. Generally speaking, networks can provide the greatest utility if they can expend their most scarce resource, low-latency delivery, on the packets most in need of it. In the consumer scenario, these are VoIP packets. VoIP is a low-bandwidth application, generally requiring no more than 128 kilobits/second, and often much less. It's a stringent application in terms of delay, however, as it can't tolerate latencies greater than 150 milliseconds (thousands of seconds) from end to end. VoIP is generally recognized as a candidate for a network boost. P2P file sharing, on the other hand, is a candidate for demotion because it tends to use as much bandwidth as is available, and to do so for a very long time, often in the range of hours. Once an ISP has determined stream requirements, it can adjust its handling so as to provide rapid delivery for VoIP and economical delivery for P2P. This is simply a matter of assigning packets to appropriate SLAs within and without the ISP's network. If every interconnected part of the Internet doesn't immediately support such an extension of past functionality, there's no cause for alarm as some day most will. The intersection of technology, economics, and marketing is too compelling for any other outcome.

---

[8] Owen Thomas, "Jail Time Shuts Down The Pirate Bay Joke Machine," *Valleywag*, Apr 17 2009. http://gawker.com/5216499/jail-time-shuts-down-the-pirate-bay-joke-machine

[9] Thomas Martin Knoll, Simple Inter-AS CoS, March 9, 2009. http://www.ietf.org/proceedings/09mar/slides/idr-5.pdf

So there's no reason to fear the use of DPI for traffic engineering. There is no loss of personal privacy from such behavior, nor would its adoption drive the Internet into a posture that's less friendly to competition. If anything, **the ability of applications to select a transport service appropriate to their needs would be an enormous boon to developers** of either time-sensitive or volume-sensitive applications. They only suffer if all traffic has to be treated as if it were the same when it's clearly different.

Tracking Cookies

One development that concerns me is the expanded use of tracking cookies to build dossiers of user behavior across the Internet. The most notorious current example is the Double-Click DART cookie[10] used by Google's AdSense program. The DART is a unique identifier placed in a user's computer by Google to track his or her movements around web sites that participate in the AdSense contextual advertising program. DART cookies as currently conceived are not especially evil – they simply allow advertisers to know how many times users have seen their ads on average, and which web sites are frequented by the same people - but there's something creepy about writing a blog post critical of Google and knowing that everyone who reads it essentially reports as much to the mother ship. Although the DART identifier is simply a random number with no particular connection to a discernable human being, the portion of the Internet's population who have both Gmail accounts and DART cookies certainly are potentially identifiable to anyone with sufficient access to Google's data base.

The prospect of ever-increasing dossiers of Internet users with information about who they are, where live, who their friends are, what blogs they read, and what trips they take is simply disturbing. While there is no evidence that this tracking data has yet been abused, it's simply a matter of time until a deranged Googler tracks an ex-girlfriend or an over-ambitious product manager applies some artificial intelligence to predict what we will buy that we didn't even know we wanted.

I have no particular recommendation regarding tracking cookies and the related dossiers but for the Committee to keep an eye on the way they're used and on the lookout for feature creep. All collectors of information seem to share the attitude that if a little bit of information is good, a lot is better, and all information tends to leak over time.

Conclusion

The most effective means of monitoring consumer behavior is a well-placed virus, and failing that it's a system of web tracking with a persistent cookie linked to a personal account. A number of technologies with primarily beneficial uses have been demonized for eroding privacy, often unfairly. The greatest threats to consumer privacy are not technologies – we're awash in technology – but business models that depend on the bartering of personal information. The Internet is unfortunately surrounded and permeated by an "information wants to be free" ethos in which advertising is the key source of revenue for the providers of application and content-level services. This business model inevitably collides with personal privacy concerns, and needs to be constantly monitored. I fear the only way to ensure robust protection for personal privacy in the long run is to replace the open-access, advertising-supported business model with one in which we pay for content and services. Given the strength of the Internet's now well-established tradition of pushing ads into and alongside practically everything that we see, this is not going to

[10] http://www.doubleclick.com/privacy/faq.aspx

be an easy transition, if it's to happen at all. But as long as personal information is the coin of the realm, it will be harvested, archived, and bartered.

Thank you for your kind attention,

Richard Bennett
BroadbandPolitics.com