Before the

FEDERAL COMMUNICATIONS COMMISSION

Washington, DC

WC Docket No. 16-106

In the Matter of

Protecting the Privacy of Customers of Broadband and other Telecommunications Services

Encryption Cannot Protect Consumer Privacy From ISPs

Jon M. Peha

Professor, Carnegie Mellon University

Address: Carnegie Mellon University
Department of EPP
Pittsburgh, PA 15213-3890
peha@cmu.edu
www.ece.cmu.edu/~peha/bio.html

July 6, 2016

Overview

This reply comment is a response to misinformation about encryption that has appeared in several comments in this proceeding. These comments promote the fallacy that use of encryption in the Internet could somehow prevent providers of a broadband Internet access service (BIAS) from obtaining extensive sensitive information about Internet users. For example, the T-Mobile comment says "developments such as consumers' use of multiple ISPs and devices and the increasing prevalence of encryption and proxy services have eroded whatever expansive access that BIAS providers once might have had to their users' traffic."[1] Comcast agrees, saying that "ISPs' visibility into the Internet behavior of their customers is also limited because more and more of the traffic that they do carry is encrypted by a third party."[2] Verizon claims that it only has access to a customer's Internet use when the customer is not encrypting: "Verizon would have access to only one small slice of that subscriber's overall Internet use (i.e., unencrypted use on Verizon's mobile network when not connected to her home, or a public, WiFi network)."[3] Other ISPs have made similar assertions. In reality, BIAS providers still have this "expansive access" into their users' traffic even when people use encryption, and even when they use "multiple ISPs and devices" as well, as this comment will demonstrate.

As an aside, it is worth noting that a tremendous amount of sensitive information flows unencrypted over the Internet every day, and this will continue. The percentage of Internet traffic that is unencrypted is falling, in no small part because an edge provider that accounts for roughly a third of Internet traffic in North America (Netflix) is adopting encryption.[4] Although the percentage may fall, with total Internet traffic growing rapidly and exponentially, the total volume of unencrypted traffic is likely to remain large for years to come.

The remainder of this comment will make the unrealistic assumption that 100% of all data transported over the Internet is encrypted, and will explain why it is still technically possible for a BIAS provider to gain extensive access to sensitive customer information. Although the focus is technical, this comment will then discuss the policy implications of these technical observations with respect to the *definition of CPNI*, *opt-out vs. opt-in*, and *parity with edge providers*.

What a BIAS Provider Can Observe

In the Internet, information is transmitted in chunks of data called packets. Most packets contain what network engineers would call application-layer data, and others sometimes call user content, although the term "content" is somewhat imprecise. Packets carrying this application-layer data or content are called data packets. When we say Internet traffic is encrypted, we mean this portion of this type of packet is encrypted. Data packets contain additional information at the beginning (header) and end (trailer) of the packet. Header and trailer information is not encrypted, because network devices need

---

[1] Comments of T-Mobile USA, Comments in the Matter of Protecting the Privacy of Customers of Broadband and other Telecommunications Services, Federal Communications Commission WC Docket No. 16-106, May 27, 2016.
[2] Comments of Comcast Corporation, Comments in the Matter of Protecting the Privacy of Customers of Broadband and other Telecommunications Services, Federal Communications Commission WC Docket No. 16-106, May 27, 2016.
[3] Comments of Verizon, Comments in the Matter of Protecting the Privacy of Customers of Broadband and other Telecommunications Services, Federal Communications Commission WC Docket No. 16-106, May 27, 2016.
[4] Sandvine, "Global Internet Phenomena Spotlight: Encrypted Internet Traffic," 2016.

access to this information. Moreover, there are packets that do not carry application-layer data, and these packets are often entirely unencrypted. Thus, even if 100% of user content is encrypted, a great deal of information is still easily visible to ISPs. This section provides an overview of what can be learned from this information. For more detailed discussion, see Section 2 of a previous paper.[5]

We first consider information that can be found in the header of a single data packet. In nearly all cases, packet headers necessarily reveal the source of content and services, and as the next section will discuss, often something about content in the process.[6] This is because the header includes the IP address of the sender and the IP address of the intended recipient. A packet can also reveal information about the device that sent the packet. In addition to an IP address, packets include the media access control (MAC) address. For example, a device that connects to a Wi-Fi hotspot will reveal its Wi-Fi MAC address. It is usually possible to determine who manufactured a device by looking at its MAC address. Through cooperation with that manufacturer, it may be possible to learn more about the type of device, such as model number, and roughly when and where it was sold. Finally, a single IP packet reveals some information about the application that generated it, through fields such as port number, packet length, and choice of transport protocol (UDP or TCP).

An ISP can learn much more by looking not at a single packet, but at a stream of packets, as occurs using flow classification.[7] For example, voice over IP (VOIP) streams have obvious characteristics; they generate bidirectional traffic at a small but steady data rate for extended periods. By looking at the duration of calls, the size of packets, and the time between packets, a VOIP call can be identified with high probability. Even more is possible by observing multiple packet streams to and from a given device. For example, at Carnegie Mellon we have developed algorithms that can determine which devices are using BitTorrent file sharing from the nature of flows and information in packet headers.[8] These techniques work well regardless of whether content is encrypted, because they do not use application-layer data anyway.

Additional information is revealed from other types of packets, such as packets exchanged during interactions with the domain name system (DNS). The domain name system converts human-friendly domain names such as fcc.gov (for the Federal Communications Commission) and epp.cmu.edu (for the Carnegie Mellon University Dept. of Engineering & Public Policy) to machine-friendly IP addresses that can go into packet headers. Most users of a BIAS perform this look-up by sending a query to a server operated by their BIAS provider, which means a BIAS provider has extensive and detailed information about the domains that the BIAS provider's customers have interacted with, regardless of whether those interactions were encrypted.

Finally, we note that a BIAS provider can learn a great deal by observing when and where devices are connected, regardless of what packets they send and receive. A mobile service does not work unless the BIAS operator knows which cell tower is close to a customer device at any given time, and often the BIAS

---

[5] J. M. Peha, "The Benefits and Risks of Mandating Network Neutrality, and the Quest for a Balanced Policy," *International Journal of Communication*, 2007. www.ece.cmu.edu/~peha/papers.html

[6] For a very small fraction of Internet traffic, source and destination address are obfuscated through tools such as The Onion Router (TOR). However, TOR would fail if a large fraction of Internet users attempted this.

[7] J. M. Peha, "The Benefits and Risks of Mandating Network Neutrality, and the Quest for a Balanced Policy," *International Journal of Communication*, 2007. www.ece.cmu.edu/~peha/papers.html

[8] J. M. Peha and A. M. Mateus, "Policy Implications of Technology for Detecting P2P and Copyright Violations," *Telecommunications Policy*, vol. 38, no. 1, pp. 66-85, Jan. 2014. www.ece.cmu.edu/~peha/papers.html

provider knows location far more accurately than that. Fixed BIAS providers can also know location if they allow access from multiple locations. For example, Comcast boasts over 14 million Wi-Fi hotspots,[9] and if a mobile device is configured to look for these hotspots, Comcast knows whenever the device is close to a hotspot. In either case, a BIAS provider that is retaining location information can track a user over time. BIAS providers know where their customers spend their waking hours, and they know where their customers sleep. Once again, encryption is irrelevant.

Sensitive Information Can Be Revealed

Much of the discussion has been about whether BIAS providers could collect information from Internet traffic that is useful for advertising, which could allow BIAS providers to use their unique position to obtain a competitive advantage in some market. Even if all traffic is encrypted, BIAS providers would indeed have this ability, although as this section will show, that may only be the tip of the iceberg from a privacy perspective.

BIAS providers that adopt the strategies described in the previous section can obtain information about the commercial behavior of Internet users that would be valuable for advertising, among other things. Even when data is encrypted, by observing which domain names a customer inquires about, which well-known IP addresses a user exchanges packets with, or other useful indicators, a BIAS provider can tell that a given customer interacts with a specific set of online merchants, travel sites, banks, newspapers, video streaming services, software vendors and much more. A BIAS provider with the ability to locate devices can also determine the bricks-and-mortar stores where a customer shops. All of this information can be valuable for advertising, especially to competitors of these online and real-world businesses.

Once mechanisms are in place to collect the information described above, Internet users have more to fear than targeted advertising. For example, last year I met with an organization that provides social services to victims of domestic violence. I was informed that some of this organization's clients were afraid to reach out. For example, in some cases, a man might become violent if he learns that his wife has been in contact with an organization of this kind, or even if she has just been reading about available services on the organization's website. I provided information on how the organization might help its clients hide these interactions from a spouse or domestic partner. However, none of the advice I gave would help against a BIAS provider, which could easily determine that these interactions were taking place, even with encryption. Could the BIAS provider sell a list of households where domestic violence is suspected to local divorce attorneys?

This is just one example of sensitive information that can be revealed simply by knowing that a user has browsed the website of a specific organization, as revealed to the BIAS provider by DNS look-up, IP address, or other available data. The knowledge that an Internet user has interacted every night with medical sites such as the American Diabetes Association (diabetes.org) or the Alzheimer's Association (alz.org) may be of value to that user's insurance company. Knowledge that the user has interacted often with websites that provide information on bankruptcy may be of value to the user's creditors. Knowledge that the user has interacted with job search sites may be of value to that user's employer.

---

[9] T. Shields and J. Plungis, "GM's Taking Cadillac on Collision Course with Silicon Valley," June 20, 2016. http://www.chicagotribune.com/news/sns-wp-blm-talking-cars-1c1f6a50-36e1-11e6-af02-1df55f0c77ff-20160620-story.html

The specific content about the disease or job or bankruptcy procedure is of secondary importance. Our browsing history reveals a great deal, and some Internet users may prefer that access to such information be limited.

Our applications and devices can also reveal a great deal. This may become increasingly important with the growth of the Internet of things (IoT). As discussed in the previous section, there are a variety of approaches that may allow a BIAS provider to identify a particular kind of IoT device, or the application running on that device. When this succeeds, it can reveal sensitive information. For example, merely detecting the existence of a personal medical device might reveal information to the BIAS provider about a medical condition of the user. Merely detecting the existence of a car ignition interlock and breathalyzer might reveal to the BIAS provider that the user has been convicted of driving while intoxicated.

Implications for Definition of CPNI

Any information that makes it possible to reach conclusions about content, application, data source, or device used in communications deserves appropriate privacy protection, as do the analytic results obtained from this data. As previous sections show, the packet header is filled with data that could be used for this purpose. This includes source IP address, destination IP address, source MAC address, destination MAC address, port number, packet length, differentiated service code point (IPv4), traffic class (IPv6), and header and trailer fields we cannot yet anticipate that will be used in future versions of Internet protocols. The simplest solution is to consider everything in the packets sent by or to customers as proprietary. The time between packets also reveals information, especially when combined with these other fields, as does the duration of TCP sessions, the number of packets and bytes sent per unit time, and other typical NetFlow data. Information obtained from DNS queries made to BIAS providers should also be included. All of these deserve reasonable privacy protection as proprietary information when disclosed or used for purposes other than the transfer of IP packets across a network.

Multiple Devices, More Observations

Some ISPs have also argued that BIAS providers are somehow losing their ability to collect sensitive user information as users expose their information to more BIAS providers via more devices. However, each such device is just another window through which to observe users. It is true that decades ago, a typical Internet user would have accessed the Internet from a single fixed location using a single BIAS provider, as the ISPs assert. Decades ago, this user also would have spent far less time online, and engaged in fewer activities where sensitive information could be revealed. Now many of us are active online throughout the day, and connected 24 hours per day. I now reveal more to the BIAS provider that connects my home computer than I used to, and the fact that I also have a smart phone does not change that. Through that smart phone, I now reveal more to my cellular provider than I used to, and the fact that I have other computers does not change that. In essence, the ISPs are arguing that if I put four more windows in my living room, then the voyeur who has been staring through the existing window will no longer be able to see me

Conclusions and Policy Implications

This comment has shown that the unique position of BIAS providers gives them extensive access to highly sensitive information about their customers. This would be true even under the implausible assumption that soon 100% of content will be encrypted, because sensitive information is observable in control packets and in the headers of data packets. BIAS providers need access to this information for the Internet service to work, and regulators should do nothing to prevent that. However, if BIAS providers wish to (i) retain, (ii) use, or (iii) disclose information derived from Internet traffic for any function that is not strictly necessary for the provision of Internet service, then users should know about it, and should have a choice. ***This requires a broad definition of CPNI*** when applied to BIAS providers.

Much of the debate concerns whether this choice should be opt-in or opt-out. Any time we are deciding between opt-in and opt-out, a key consideration is whether users understand the privacy risks well enough to know their own preferences. If yes, then those who would prefer to limit what BIAS providers can do with this information would know to contact their BIAS provider to opt out, while those who are less concerned about privacy do nothing. Thus, opt-out would lead to a good outcome. However, if users don't even know that privacy is even an issue, then they may never learn what it means to opt out, or consider whether opting out is better for them. This could lead to significant harm for some users. Thus, the less users know, the more reason there is to consider an opt-in approach, whereby BIAS providers have to ask their customers for permission to access user information when it is not necessary, as asking this question may give those customers reason to learn more.

Just how well do BIAS customers understand the privacy issues associated with Internet traffic? At least as of today, probably not well enough to make informed decisions under an opt-out policy. Consumers know that when they provide information to an edge provider, they are revealing private information, but so far I have seen little indication that consumers understand what BIAS providers are able to learn from Internet traffic. Moreover, the comments I have read in this proceeding provide some evidence that the opposite is true. With representatives of ISPs publicly telling the FCC that encryption and use of multiple devices have "eroded whatever expansive access that BIAS providers once might have had to their users' traffic,"[10] how can we expect typical consumers to know that BIAS providers still have access to sensitive information? Will a typical consumer know that even with encryption her BIAS provider can still tell if she reads the website of the Domestic Violence Center or the American Diabetes Association, and could potentially sell this information to others? Not if consumers listen to what ISP representatives are saying. Thus, the FCC can cite ISP comments as evidence that an opt-out requirement may be inadequate for sensitive information derived from observing Internet traffic.

Some BIAS providers are calling for parity, by which they typically mean that the FCC should adopt regulations for BIAS providers that are the same as the regulations that the FTC has applied to some of the edge providers. I have previously explained why this parity objective is ludicrous and unachievable.[11] There are already multiple and conflicting privacy regulations from the FTC and other government agencies. Rather than trying to achieve parity with all of them, which is obviously impossible, the FCC should adopt the regulations for BIAS providers that best serve the public interest, regardless of what other agencies have done with other industries.

---

[10] Comments of T-Mobile USA, Comments in the Matter of Protecting the Privacy of Customers of Broadband and other Telecommunications Services, Federal Communications Commission WC Docket No. 16-106, May 27, 2016.

[11] J. M. Peha, "The Fallacies of Regulatory Parity in Privacy Regulation," Comments in the Matter of Protecting the Privacy of Customers of Broadband and other Telecommunications Services, Federal Communications Commission WC Docket No. 16-106, May 27, 2016. https://ecfsapi.fcc.gov/file/60002079501.pdf

However, if the FCC does decide to seek parity, then the FCC should impose regulations based on the most sensitive information that a BIAS provider has access to, not the least sensitive. Consider the case where a BIAS provider uses its unique position and techniques described in this comment to determine that an individual is spending an hour of every day in a high-data-rate connection with a Cancer Treatment Center, and the characteristics of the data stream are consistent with that of a telemedicine application. The Cancer Treatment center is an edge provider, and its privacy practices are governed by rules developed by the Department of Health and Human Services under the Health Insurance Portability and Accountability Act. This means that the Cancer Treatment Center would require explicit authorization from the patient before disclosing information about patient care to most third parties. Should the BIAS provider be free to disclose information about these sessions when the Cancer Treatment Center cannot? If the FCC decides to use parity rather than public interest as the standard, as some ISPs have requested, then restrictions on a BIAS provider's ability to disclose such information must be comparable to restrictions on the Cancer Treatment Center, and these go well beyond what the FTC would require.

.

About the Author

Jon Peha is a Professor at Carnegie Mellon University, with experience in industry, government, and academia. In government, he served at the FCC as Chief Technologist, in the White House as Assistant Director of OSTP, in the House Energy & Commerce Committee, and at USAID for the Telecommunications Leadership Program. In industry, he has been Chief Technical Officer for three high-tech companies, and member of technical staff at SRI International, AT&T Bell Labs, and Microsoft. At Carnegie Mellon, he is a Professor in the Dept. of Electrical & Computer Engineering and the Dept. of Engineering & Public Policy, and former Associate Director of the Center for Wireless & Broadband Networking. Dr. Peha holds a PhD in electrical engineering from Stanford. He is an *IEEE Fellow* and an *AAAS Fellow*, and was selected by AAAS as one of 40 Featured Science and Technology Policy Fellows of the last 40 years ("40@40"). Dr. Peha has received the FCC's "Excellence in Engineering Award," the IEEE Communications Society TCCN Publication Award for career contributions, and the Brown Engineering Medal. He consults on a wide range of technical and policy issues related to information and communications technology.