Nick Feamster
Professor, Department of Computer Science
Acting Director, Center for Information Technology Policy

310 Sherrerd Hall
Princeton University
Princeton ,NJ 08540-5233
+1 609 258 2203
feamster@cs.princeton.edu

RE: Docket No. 16-106, Protecting the Privacy of Customers of Broadband and Other Telecommunications Services.

Chairman Tom Wheeler
Commissioner Mignon Clyburn
Commissioner Jessica Rosenworcel
Commissioner Ajit Pai
Commissioner Michael O'Rielly
Federal Communications Commission
445 12th Street SW
Washington, DC 20554

May 27, 2016

Dear Chairman and Commissioners:

We write in reply to the FCC's proposed rulemaking in WC Docket No. 16-106. We aim to provide some important technical context concerning the scope of the FCC's rulemaking authority with regards to ISPs and privacy and the extent to which it applies to edge providers.

The main premise of this comment is that BIAS providers and edge providers operate extremely differently in terms ot the nature of the data that they collect about customers, and the extent to which customers can control the collection of customer proprietary network data.

The proposed rulemaking aims to constrain the ways service providers who fall under Section 222 of the Communications Act can collect and share Customer Proprietary Network Information (CPNI); it also prescribes how service providers must ask consumers to opt-in to the collection of CPNI and outlines baseline requirements for data security and breach notification. ISPs are also prevented from charging subscribers a premium for providing baseline privacy protections to consumers. The rulemaking has potential far-reaching implications for how ISPs operate and secure their networks, which I will discuss in a separate comment.

A question concerning the rulemaking is whether it might pertain to "edge providers" (i.e., content providers, application service providers), in addition to Broadband Internet Access Service (BIAS) providers.

The similarities and differences between BIAS providers and "edge providers", including content and application service providers warrants discussion. We can discuss the nature of data collection along two dimensions.

- What data is visible, and how it can be tied to CPNI (we discuss two types of data: network traffic data, and customer-provided data), and the choices consumers make about the collection of that data.
- Whether and how that data is used to secure, manage, and operate the service

With regard to data collection practices of edge providers, it is worth making the distinction between (1) network traffic data; and (2) customer-provided data. In the remainder of this comment, we will explain why the privacy concerns surrounding this data is orthogonal to the concerns of this proposed rulemaking. In summary, the network traffic data that edge providers collect cannot be linked to any CPNI that is collected without the user's consent; and, the data that edge providers do have about their customers is willingly provided by the customers themselves.

Although consumers should certainly be concerned about how edge providers are collecting, using, sharing, and retaining this data, these concerns would appear to fall outside of the scope of Section 222, because (1) edge providers do not have any data that is directly linkable to CPNI that these customers do not already explicitly provide in some way to the edge providers; (2) the customers of these services typically have clear choices about whether to reveal this data in the first place, due to the nature of these services.

What Network Traffic Data is Visible. Edge providers, like ISPs, operate networks to deliver network traffic and, in this regard, face many of the same network management tasks as operators of ISPs, ranging from performance management to provisioning to security. As a result, the network operators for edge providers almost certainly collect information about network traffic (in the form of flow records, byte counters, DNS lookups, and packet traces), as well as information about Internet routing (i.e., routing tables and updates from network routing protocols such as the Border Gateway Protocol).

Yet, while the *types* of network data that edge providers may collect are the data that is collected in BIAS networks, the differences lie in the *ability to tie this network data to customers*. To illustrate these differences, consider the following two types of network traffic data:

- **Network traffic statistics** collected with IPFIX or via deep packet inspection may contain source and destination IP addresses. Yet, due to the edge provider's position along the network path, the provider has cannot link this traffic data back to other CPNI (e.g., the name and address of a user corresponding to that IP address). In contrast, a BIAS provider does have the ability—at least in principle—to link information about IP addresses seen in network traffic traces to CPNI from its subscribers.

- **Domain name system (DNS) lookup** information. Edge providers operate DNS domains (e.g., google.com, facebook.com, netflix.com) and, as such, control the authoritative DNS servers corresponding to those domains. As a result, each respective edge provider will have information about the DNS lookups to its domain. However, the nature of the DNS lookup data that an edge provider sees is very different from the DNS traffic that a BIAS provider sees.

  An edge provider will see DNS queries from what is called a recursive DNS resolver—typically, a machine that resolves domain names on behalf of many users (e.g., all subscribers from a particular ISP). Generally speaking, the operator of an authoritative DNS server for an edge provider domain will only see lookups to its own domain—and even then, only from IP addresses from recursive DNS resolvers, not IP addresses of individual users. In contrast, because BIAS providers often operate recursive resolvers on behalf of their subscribers, they will see what is sometimes referred to as "below the recursive" DNS resolver traffic—in other words, DNS lookups from IP addresses corresponding to identities that can be subsequently mapped back to subscribers. They can also see DNS lookups to *all* domains for a particular IP address, as opposed to only the DNS domains for just one domain (as is the case for edge providers).[1]

- **Routing information** such as routing table updates for Internal Gateway Protocols (IGPs) for internal topology and the Border Gateway Protocol (BGP) for interdomain routing decisions do not contain any CPNI. Operators from both BIAS and edge providers certainly collect this data to assist with network management, operations, and security, but the collection of this data is orthogonal to discussions of customer privacy.

For each of these classes of data, the customer does not have a choice as to whether a network operator collects the data. In the case of ISPs, the collection, sharing, and retention of this data warrants a reasoned discussion about the

[1] Google, of course, appears to be an exception to this characterization; although they may be deemed an "edge provider", they *also* operate their own open recursive DNS resolvers that many users rely on for DNS resolution. As a result, they also have access to DNS lookup information for individual users.

acceptable uses of and risks associated with this data. In the case of ISPs, this data does contain CPNI and sensitive information about user activity, and yet there are good reasons for ISPs to collect and share this information, even if users do not explicitly opt in. In the case of edge providers, however, this data does not typically contain CPNI in the first place, because of where this data is being collected. As a result, the collection, sharing, and retention of the above data doesn't warrant the same types of discussions about consumer privacy protections as we need to have when considering BIAS providers.

What Customer-Provided Data is Visible. Although the network traffic data that edge providers collect is less pertinent to CPNI, edge providers often have much richer datasets about consumers: a social network may have information about a user's activities, or other personal information (photos, friendship relationships, marital status, age, religion, etc.). A streaming video service may have information about a customer's preferences and viewing habits. A search engine provider may retain a user's search history. By many considerations, this data is much more sensitive than what might be cleaned from network traffic data; it is also much more evident (e.g., the edge provider may know a user's marital status because the user told them).

On the other hand, *users have choice* when deciding whether to reveal this information to an edge provider. For example, in many cases, a user may register with an edge provider using a pseudonym. The user may simply elect not to provide certain personal information or data to a social network, or even to not use the social network at all. So, while there are certainly privacy concerns with the collection, use, and sharing of data that edge providers collect, it is fair to say that the concerns are somewhat different than those that we face when considering privacy protections for BIAS providers.

Securing, managing, and operating a service. With regards to the data discussed above, operators of both BIAS networks and edge providers rely on network traffic data to operate, manage, and secure the network. As discussed, the use and sharing of this data requires more careful consideration in the case of BIAS networks because it can more easily be linked with customer data. On the other hand, the data that customers provide *by choice* to edge providers does have privacy concerns, but it has nothing to do with operating, managing, or securing the network.

In summary, the rulemaking appears most concerned about data about customers that is gathered without their consent and might be shared without their consent. With some exceptions (e.g., Google's public DNS service), the data that edge providers have is either (1) not directly linkable to customer proprietary network information (as in the case of network traffic); or (2) not collected without the user's consent (as is the case with customer-provided data). Based on this, we conclude that the data that edge providers collect is outside the scope of Section 222 and the FCC's proposed rulemaking.

Sincerely,

Nick Feamster
Professor of Computer Science
Director, Center for Information Technology Policy
Princeton University

Supporting Signatories:

/s/

David Farber
Distinguished Career Professor of Computer Science and Public Policy
Carnegie Mellon University

Yan Chen

Professor of Electrical Engineering and Computer Science
Director, Lab for Internet and Security Technology
Northwestern University

Doug Comer
Distinguished Professor of Computer Science
Purdue University

Jim Hendler
Tetherless World Professor of Computer, Web and Cognitive Sciences
Director, The Rensselaer Institute for Data Exploration and Applications
Rensselaer Polytechnic Institute