Comments to the FCC on Broadband Consumer Privacy

Peter Swire
Huang Professor of Law and Ethics
Georgia Tech Scheller College of Business

April 28, 2015

This document responds to the request from the Federal Communications Commission to participate in the April 28, 2015 Public Workshop on Broadband Consumer Privacy. I have been asked to speak on Panel 2, the application of Section 222 of the Communications Act to Broadband Internet Access Services.

Based on the questions posed to me by staff, of many possible issues for comment, I have tried to highlight areas where my experience and previous scholarship have distinctive things to add to the discussion. My organizing principle is consumer benefit. My remarks do not address protection of one carrier versus another in connection with proprietary information.

Where possible, I have tried to look to the practices of the Federal Trade Commission and other federal privacy regimes as guidance for the treatment of broadband companies under the Open Internet Order. In particular, I draw in a number of places on my experience drafting and then implementing the HIPAA medical privacy rule and the Gramm-Leach-Bliley Act (GLBA) financial privacy rule. In considering these sectoral approaches to privacy, one theme is that the Commission should be cautious about picking winners and losers in markets where there is ongoing competition between entities covered by Section 222 and those not covered.

This document addresses three sets of issues. First, I examine the effect of the Section 222(a) definition of "proprietary information" as compared with the Section 222(c) definition of "customer proprietary network information." (CPNI) My conclusion, based on some analogous provisions from HIPAA and GLBA, is that the Commission should be cautious about founding any additional regulatory requirements under this proceeding based on the language in 222(a).

Second, I examine the intersection of privacy and competition law, drawing on my previous writings in the area. New entry into online advertising, including by broadband providers, could be a new source of competition on privacy attributes. My recommendation to the Commission is to consider the effects of this potential competition on privacy and other non-price aspects of competition, along with price aspects of competition, as part of the overall assessment of how to govern the use of CPNI for broadband providers.

Third, I address priority uses of information that I believe should be permitted in the CPNI context. Although I do not seek to create a complete list of possible exceptions to the general CPNI rule of consumer opt-out, I do emphasize three areas where an opt-out is not generally appropriate – anti-fraud, cybersecurity, and research on network usage. I also analyze the role of de-identification and aggregate information under Section 222, suggesting strategies to preserve the utility of de-identified and aggregate information while protecting privacy. In this discussion, I do not take a position on whether a rules-based, principles-based, or other approach should be adopted by the Commission. Instead, I emphasize that important interests such as anti-fraud and cybersecurity should be taken into careful consideration in whatever approach the Commission pursues.

Background of the witness. I am the Huang Professor of Law and Ethics at the Georgia Tech Scheller College of Business, with appointments by courtesy with the College of Computing and School of Public Policy. Consistent with university consulting rules, I am Senior Counsel with Alston & Bird, LLP.

I have been immersed in privacy and cybersecurity issues for two decades. In 2015, the International Association of Privacy Professionals, among its over 18,000 members, awarded me its Privacy Leadership Award. In 2013, I served as one of five members of President Obama's Review Group on Intelligence and Communications Technology. Prior to that, I was co-chair of the global Do Not Track process for the World Wide Web Consortium. I am Senior Fellow with the Future of Privacy Forum, and a Policy Fellow with the Center for Democracy and Technology.

Under President Clinton, I served as Chief Counselor for Privacy, in the U.S. Office of Management and Budget, the only person to date to have U.S. government-wide responsibility for privacy policy. In that role, my activities included being White House coordinator for the HIPAA medical privacy rule, serving as White House representative to the privacy rulemaking process under the Gramm-Leach-Bliley Act, and helping negotiate the U.S.-E.U. Safe Harbor agreement for trans-border data flows. Under President Obama, I served as Special Assistant to the President for Economic Policy in 2009-2010.

I have testified on privacy and other issues before almost a dozen committees in the U.S. Congress, and worked closely with the Federal Trade Commission and other federal agencies on privacy and cybersecurity issues. In 2011, the Federal Communications Commission asked me to summarize and comment on the day's proceedings for its Workshop on Location Information. Further information is available at www.peterswire.net.

Role of 222(a) definition of "proprietary information."

One issue that staff inquired about is what effect to give the definition of "proprietary information" in Section 222(a) as contrasted with the definition of

"customer proprietary network information" in 222(c). Based on my knowledge at this time, I would be cautious about founding any additional regulatory requirements under this proceeding based on that language in 222(a).

This issue received public attention in connection with the FCC's enforcement action against TerraCom and YourTel in 2014, when the definition in 222(a) was mentioned as a possible independent basis for enforcement. My read of the statutory language would make me cautious about finding any such independent authority. Based on the research I have been able to do, I am not aware of any legislative history that suggests that telecommunications carriers have independent duties toward customers arising from "proprietary information" as opposed to CPNI. Writing as one law professor who has taught legislation and helped develop regulations, the usual approach is that the specific definition in 222(c) would seem to give content to the umbrella or introductory language in 222(a). In addition, as a matter of grammar, 222(a) would appear to state the general approach (the section is called "in general") that there should be the protection of proprietary information of: (i) other telecommunication carriers; (ii) equipment manufacturers; and (iii) customers. Then, section 222(b) gives content to that duty as applied to the confidentiality of carrier information and 222(c) similarly gives content to the duty as applied to the confidentiality of CPNI.

Experience with other major U.S. privacy regimes also gives reason for caution about reading these provisions as imposing independent legal obligations under two different definitions of what information is covered. Under HIPAA, the statute uses a broader term "individually identifiable health information," (IIHI) but the obligations of the HIPAA privacy and security rules apply only to the carefully-defined scope of "protected health information."[1] (PHI) PHI is essentially IIHI that is held by the covered entities that have legal obligations under the HIPAA privacy and security rules. Under Gramm-Leach-Bliley, the statute uses two terms for scope – "consumers" and "customers." Based on my experience in the rulemaking process for GLBA, regulators were concerned that these similar-sounding terms could have led to confusion in implementation. As applied by the regulation (and consistent with the statute), the broader term "consumer" includes any individual who establishes "a customer relationship" with a financial institution. The GLBA right to notice applies only to "customers" and not to "consumers."[2] Under both HIPAA and GLBA, in short, the presence of two terms potentially creating different scope was resolved in practice by having one term that sets forth the legal obligations.

This careful delineation of what is covered in a privacy regime, in my view, is sound policy. Definition of what data is covered is a consequential regulatory decision. Data in a modern organization is handled in numerous, complex, and rapidly-changing ways. I write the official textbooks used for the Certified Information Privacy Professional tests administered by the International Association

1 45 CFR Parts 160 and 164.
2 15 U.S.C. § 6803.

of Privacy Professionals.[3] For professionals learning and then administering privacy rules, a fundamental first step is to define the scope of what is covered. Once information is covered by a regulation, numerous additional steps are typically required, such as rules for customer consent, access, audits, training, and so on. Vagueness in the scope of what is covered thus causes particular compliance challenges. Having two overlapping potential scopes of coverage, such as for "proprietary information" and also CPNI, would thus complicate and make more difficult the role of privacy professionals seeking to do their jobs properly. It would be harder for the privacy professionals to communicate to the rest of the organization what actions must be taken, with a consequent risk to the achievement of effective compliance.

The intersection of privacy and competition law – privacy as a non-price basis for competition.

The FCC staff asked us to consider addressing: "What are the competitive implications—whether for ISPs or for other actors in the online ecosystem—of statutory protections for data held by ISPs?" In response, I describe my previous writing on the intersection of privacy and competition law. I then apply that writing to the Section 222 context.

Discussion of the Google/DoubleClick merger in 2007.

My own writing on this topic was triggered by the proposed merger of Google and DoubleClick in 2007. As a professor of both antitrust and privacy law, I was concerned that the public debate had not accurately understood the intersection of these two topics. For instance, a New York Times article in 2007 stated "Strictly speaking, privacy is not an antitrust issue."[4] I submitted testimony on the intersection of antitrust and privacy while the merger was under consideration, explaining analytically how the issues fit together but specifically not taking a position on the facts of the actual merger.[5]

---

3 Peter Swire & Kenesa Ahmad, "Foundations of Information Privacy and Data Protection: A Survey of Global Concepts, Laws, and Practices" (2012); Peter Swire & Kenesa Ahmad, "U.S. Private-Sector Privacy: Law and Practice for Information Privacy Professionals" (2012).

4 Steve Lohr, *Google Deal Said to Bring U.S. Scrutiny*, N.Y. Times, May 29, 2007, at C1 (quoting Andrew I. Gavil, law professor at Howard University).

5 "Privacy and Antitrust" (2008), available at http://peterswire.net/speeches/; "Google and Privacy: Merger with DoubleClick Prompts New Privacy Guidelines", Dec. 20, 2007, available at https://www.americanprogress.org/issues/technology/news/2007/12/20/3790/google-and-privacy-merger-with-doubleclick-prompts-new-privacy-guidelines/; "Protecting Consumers: Privacy Matters in Antitrust Analysis," Testimony submitted to the Federal Trade Commission on Privacy and Antitrust, October 19, 2007, available at

The basic idea of my writing is that privacy can be an important non-price aspect of competition. Where it is, then a merger or other practice can reduce competition, providing a basis for scrutiny under the antitrust laws. For example, imagine an agreement not to compete on warranties, or a merger where competition on warranties would be greatly reduced. On those facts, there would be an antitrust injury to consumers. The same analysis, I have argued, can apply to an agreement not to compete on privacy, or a merger where competition on privacy would be greatly reduced.

I have argued that the following two questions are relevant to antitrust analysis. First, is privacy a non-price factor (a quality of a product or service) that is important to consumers? Second, will the merger or other action reduce competition in privacy, creating antitrust injury to consumers?

In deciding to approve the Google/DoubleClick merger, the four FTC Commissioners in the majority accepted my analytical framework, and only approved the merger after finding that consumers would not be harmed by reduced privacy competition. The majority decision specifically referenced the basic theory: "We investigated the possibility that this transaction could adversely affect non-price attributes of competition, such as consumer privacy."[6] The dissenting Commissioner Harbour also accepted my proposed analytical framework, citing my testimony as a basis for finding that the merger should not proceed, saying that antitrust law should ensure competition "based on privacy protections or related non-price dimensions."[7]

Application of this antitrust analysis to Section 222 and broadband.

Although the current proceeding does not concern a proposed merger such as Google/DoubleClick, the antitrust and privacy analysis is essentially the same. As in the antitrust context, consumer privacy interests may be helped or hindered in light of market structure and the likely effects on competition.

---

https://www.americanprogress.org/issues/regulation/news/2007/10/19/3564/protecting-consumers-privacy-matters-in-antitrust-analysis/.

6 https://www.ftc.gov/public-statements/2007/12/statement-federal-trade-commission-concerning-googledoubleclick

7 https://www.ftc.gov/public-statements/2007/12/dissenting-statement-commissioner-harbour-matter-googledoubleclick. More recently, FTC Commissioner Maureen Ohlhausen and a co-author have written expressing skepticism about including privacy in antitrust analysis. Maureen K. Ohlhausen and Alexander P. Okuliar, "Competition, Consumer Protection and the Right [Approach] to Privacy," (2015), available at http://ssrn.com/abstract=2561563. In correspondence with the Commissioner, she agrees that she has not addressed my particular approach in her writing to date.

Concerning the last-mile market, I have not written about or taken any public position on the degree of competition or on the overall advisability of reclassifying broadband companies as common carriers. (I may be one of the few Internet law professors in that category, but I have focused instead on privacy and cybersecurity issues.) The first panel of this hearing addresses "privacy implications associated with broadband Internet access services." My understanding is that panel will, among other issues, explore ways in which delivery of last-mile broadband services is similar to, or differs from, practices by other players in the online eco-system, including first- and third-party advertisers. Without going into the issues in great detail, my sense is that there has been considerable change over time in the relative position of last-mile providers and other actors in the online eco-system. At the start of the commercial Internet 20 years ago, the last-mile providers at a technical level had a large advantage over other players in terms of observing a consumer's online activity. Online companies at the edge, such as early web sites, had relatively limited visibility into a user's overall Internet activity. By contrast, last-mile providers today do not seem to have the same relative advantage over other online actors. My experience in the Do Not Track process showed that individual users face large challenges today if they wish to cut themselves off from tracking in the online advertising sector. If a user deletes a cookie, it is often regenerated, with new browsing linked to the same device. The practice of digital fingerprinting of devices also makes it difficult for users to shield their browsing activities from the online advertising eco-system. I look forward with interest to discussion in the first panel at today's proceeding, but my general sense is that the information available to last-mile broadband providers today is very roughly similar to that available to other leading players in online advertising markets.

I do offer some observations about possible ways that competition in privacy might take place in markets other than competition for last-mile services. Based on my experience in the W3C Do Not Track process in 2012-2013, some sectors of online advertising have clear market leaders, including search and social networks. The issue I ask the Commission to consider is the effect of broadband providers' actions on online advertising markets, including their competitiveness.

Based on my experience, I believe that effective entry into major sectors of online advertising occurs only at scale. There are well-known network effects that favor first movers and incumbents. To the extent policymakers would like to have greater competition in those online advertising markets, therefore, the question is who could be effective "mavericks" – what does it take to be an effective new entrant with the plausible possibility of reducing concentration in those markets? For online advertising, major broadband providers are the most plausible such new entrants that I can see.

As a matter of competition law, online advertising is a two-sided market. Greater competition could presumably benefit advertisers, in the form of lower prices and a greater diversity of outlets for effective online advertising. Greater competition could also benefit individual consumers. New entrants have the ability

to create packages of advertising practices that are more attractive to consumers – such as through increased quality or non-price differentiators. As discussed above, privacy is one important non-price potential dimension of competition. In short, new entry by broadband providers into online advertising provides the opportunity for new competition on privacy.

The magnitude and nature of any such competition would be a factual matter for which the Commission could develop a record. Based on my discussions over time with the Federal Trade Commission, there has been ongoing interest in fostering competition on privacy, to improve consumer choice about privacy practices. *In conclusion, my recommendation to the Commission is to consider the effects of this potential competition on privacy and other non-price aspects of competition, along with price aspects of competition, as part of the overall assessment of how to govern the use of CPNI for broadband providers.*

The importance of public policy exceptions and defining the scope of the data covered by a privacy regime.

Properly defined exceptions are an essential aspect of any privacy regime. Based on my experience in drafting and applying privacy regulations, *there are many valuable and essential uses of personal information as well as strong reasons to provide privacy protections.* The balance here is tricky and inherently debatable. Those pushing for strong privacy protections are inclined to draft exceptions narrowly, concerned that broad exceptions will undermine the overall effort of privacy protection. Those pushing for exceptions wish to reserve maximum flexibility for other important public policy values. My own goal, when working on HIPAA and GLBA, was to be open to argument that an exception was needed while also remaining aware that the potential list of exceptions can grow quickly, as one possible use after another of personal information is put forward as in the public interest.

Based on my experience, I would like to comment on three valuable areas related to CPNI and broadband where I believe some exception is justified, as well as related discussion on the conjoined issues of "aggregate information" and de-identification. The three provisions that I examine concern: anti-fraud, cyber-security, and research, including the topics of de-identified and aggregate information.

Anti-fraud, CPNI, and broadband. Entities providing last-mile services, including wireless telecomm providers and broadband providers, have a crucial role to play in reducing the effects of fraud on consumers. My comments here emphasize the importance of enabling effective anti-fraud measures in the provision of last-mile service. I believe it is possible and appropriate to do so while limiting anti-fraud as a basis for a potentially over-broad exception to privacy protections.

Section 222 itself strongly supports protection of consumers against fraud: a carrier and its agents are permitted to use, disclose, or permit access to CPNI without a customer's affirmative consent "to protect users of those services and other carriers from fraudulent, abusive, unlawful use of, or subscription to, such services." 47 U.S.C. § 222(d)(2). Essentially identical language exists in the implementing CPNI regulation. 47 C.F.R. 64.2005(d).

My understanding of historical practice is that carriers have often focused on the statute's authorization "to protect the rights or property of the carrier" for purposes of preventing fraud. I write to emphasize that the statute says "to protect the rights or property of the carrier *or* to protect users of those services …" (emphasis supplied). In addition, the statute adds that the anti-fraud provision applies for "subscription to" such services. The "or" is a clear indication that the statute supports anti-fraud actions aimed at protecting users of the services, including in their subscription to those services, apart from harm to the carrier itself. This protection of consumers is a logical interpretation of a consumer protection statute, and is clearly indicated in the text.

Based on my recent work with a client in the anti-fraud space,[8] I have come to the conclusion that impersonation of consumers' devices and accounts is likely to be an important risk going forward in provision of last-mile services to consumers. One category of risk is device takeover, such as where the fraudster is able to impersonate the device of a legitimate user. A second is a man-in-the middle attack – the users control their own devices, but the fraudster is able to take action between the sender and the recipient. A third category of risk is account takeover, where the fraudster achieves identity theft, and appears to be an authorized subscriber to the telecommunications service, thus gaining fraudulent access and benefitting from access to the authorized subscriber's accounts. These sorts of risks to consumers were more modest in traditional landline access – it was relatively rare for a fraudster to be able to impersonate a landline call, and gain financial advantage by that access to defraud others. Today and in the future, when so many important consumer activities will happen over the last mile – mobile banking services, purchase of online goods and services, and many more – the incentive for fraudsters to impersonate accounts or devices grows sharply. Therefore, to benefit consumers, strong measures are appropriate to detect fraudulent use of the accounts and devices used in the last mile, such as through wireless phones or other access to broadband networks.

---

8 The client is Payfone, Inc., which provides services that notably for purposes of this testimony address two kinds of fraud on consumers: (1) impersonation of the consumer's device, such as a fake SIM card in a mobile phone; and (2) account takeover by a fraudster. Consistent with my strong belief that it helps consumers to reduce this sort of fraud, I have been working with Payfone and others to clarify the application of the CPNI anti-fraud provision to these sorts of consumer protection.

One important issue, which the FTC has also faced, is how to enable this sort of verification and authentication without creating an over-broad exception for anti-fraud. The exception could become over-broad, for instance, if incidental use of a wireless phone or Internet connection in a fraud were enough to trigger the anti-fraud exception under 222(d). With that sort of very broad exception, there are potential risks to consumers if their CPNI were spread to numerous recipients in the name of addressing fraud; each of those recipients, in turn, could become a potential source of breach of CPNI, and it is possible that effective technical and administrative controls would not be in place to adequately protect the CPNI.

A better fit with the text and purpose of Section 222 would be to recognize that the anti-fraud exception applies to *fraud concerning the use of the telecommunications service itself.* This approach would enable use of anti-fraud measures that concern fraud about the device itself, or account takeover. It would not, by contrast, enable use of the anti-fraud exception based on the mere possibility that the recipient of the CPNI might at some future point use it for anti-fraud purposes.

Based on discussions with FTC officials and FTC statements, I believe that this focus on verification and authentication is consistent with their experience and views. In its 2014 Data Broker report, the FTC recommended a consumer opt-out for other services, such as people search, but did not recommend an opt-out for anti-fraud services designed to reduce the incidence of identity theft.[9] The FTC there also noted the problems with allowing fraudsters to access and correct data, due to the risk of account takeover and other forms of identity theft. Similarly, under Section 222, the anti-fraud provision applies without a consumer opt-out. The reason is intuitive – the fraudsters have strong reasons to opt out of the anti-fraud efforts, and have similar reason to exercise the opt out when they hijack a device or accounts. They should not be provided the opportunity to do so.

Cyber-security, CPNI, and broadband. The Commission's approach to CPNI and broadband should seriously contemplate how to achieve cybersecurity, in addition to the specific risk just discussed of identify fraud concerning a consumer's device or account. In my cybersecurity course at Georgia Tech, we discuss how a cyber strategy that relies primarily on firewalls does not work well in today's computing environment, where so many legitimate (and illegitimate) uses of data cross the boundaries between one organization and another. Instead, cyber-security increasingly relies on a wide variety of information-based strategies.

To give examples relevant to broadband providers, a subscriber may be a bot, operated remotely by a criminal hacker; or, the subscriber may be operating a botnet or doing other cybersecurity attacks. In either case, information about the subscribers available to the broadband provider could prove useful for

9 Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability," at 53 (2014).

cybersecurity purposes. There is a strong public policy case for allowing use and sharing of this cybersecurity information without a customer opt-out. If an opt-out were permitted, then hackers could opt out of use of their CPNI when they were subscribers, and could cause bots under their control similarly to opt out of use of CPNI.

In making these common-sense observations, I am not taking any particular position on the information sharing legislative proposals currently before the Congress. I have written skeptically about some earlier versions of information sharing bills,[10] due to the well-known tensions information sharing can create with existing privacy laws. The Obama administration has emphasized the importance of privacy and civil liberty protections in previous rounds of this legislative debate, and I hope any legislation that does emerge carefully addresses both cybersecurity and privacy concerns. In addition, as shown by my work with President Obama's Review Group on Intelligence and Communications Technology, government access to communications data should be done subject to a careful policy process that creates appropriate limits and strong institutional safeguards.

My main recommendation on this issue is that the Commission should do careful fact-finding, prior to any regulating, about how to promote cybersecurity consistent with privacy goals. As one example, it may be difficult at a technical level to assure that cybersecurity measures operate consistently with a consumer opt-out -- it may be difficult to segregate CPNI attached to an opt-out from other CPNI flowing through a broadband provider's system. Modern cyber-security approaches feature numerous technical measures to examine systems, some of which will detect information connected to CPNI. Other defensive measures, however, may process information but without any mechanism to automatically sort which information is governed by an opt out. In that instance, an opt out may be difficult or impossible to apply in practice, especially for the growing range of defensive measures that respond in real time, often faster than a single second. As another observation, the cybersecurity measures used in one year may evolve into quite different measures within a short time, as new threats emerge. *Any provisions for cybersecurity under the Open Internet proceeding should avoid technology-specific requirements that may quickly become outdated.*

Enabling research about network usage. In drafting the HIPAA medical privacy rule, we knew that it was essential to build privacy protections consistent with enabling ongoing medical research. To do so, the HIPAA privacy rule has four distinct ways to conduct medical research: (1) with individual consent; (2) with approval by an Institutional Review Board (IRB) or similar institution, even without patient consent; (3) by de-identifying the data; or (4) under data use agreements, which allow certain flows of protected health information where there are effective

10 Peter Swire, "Moving Too Fast on Cybersecurity," TheHill, Apr. 20, 2012, available at http://thehill.com/opinion/op-ed/222783-moving-too-fast-on-cybersecurity.

institutional controls against mis-use. As discussed further below, the IRB, de-identification, and data use provisions are done without individual patient consent.

My long-term view is that other privacy regimes should carefully consider ways to enable research consistent with privacy protection. In the short time since I was asked to testify, I was able to talk with academic experts in the area of network usage, and they were positive about recognizing the importance of network research in any privacy regulations or other Commission action under the Open Internet proceeding.

In considering research related to CPNI and broadband networks, a first goal of the Commission should be to do no harm – any privacy-related actions should not inadvertently make it more difficult to do legitimate academic and similar network research. At the same time, as discussed further below, the Commission might consider provisions that would enable research consistent with meeting privacy goals.

The Commission should also explore whether it can encourage useful academic and similar network research consistent with existing laws. Under current law, the Digital Millennium Copyright Act has a security exception, but it is narrower than many security experts, in my experience, believe is appropriate. The principal federal anti-hacking statute, the Computer Fraud and Abuse Act, is vaguely enough worded that network researchers have similar concerns that what they consider appropriate academic research might instead be found to be an illegal hack. Aaron Burstein, now an attorney at the Federal Trade Commission, has written a useful article about the legal obstacles to online cybersecurity research. Burstein convincingly argues that a major obstacle to conducting research in this area is "the difficulty of determining which of a large set of complex statutes might regulate a given research project."[11] As part of addressing research and privacy in the broadband context, the Commission may be able to offer clearer guidance to researchers, both within and external to broadband providers, about how to conduct network research consistent with applicable law.

In terms of the substance of a research provision, the Commission can consider how to build on the Institutional Review Board approach contained in HIPAA. The HIPAA Privacy Rule supplemented the previous IRB rules about human-subject research by adding new privacy-related questions. Research proposals that answered these privacy-related questions could qualify for the HIPAA research exception, without the need for individual consent. Guidance from the Commission about privacy issues might reduce the uncertainty about what is lawful and ethical research in this area. I believe greater certainty in this area would on average

---

11 Aaron J. Burstein, "Conducting Cybersecurity Research Legally and Ethically," (2008), available at https://www.usenix.org/legacy/events/leet08/tech/full_papers/burstein/burstein.pdf.

increase the willingness of IRBs to authorize research about network usage, even where CPNI is accessed by the research team.

De-identification, research, and aggregate information. As mentioned above, the HIPAA Privacy Rule specifically permitted research where the individuals' information was de-identified. That rule gave two methods for achieving de-identification. The first was to delete a specific list of 17 data fields, such as name, address, and email address. The second, which I believe is a better model for the Commission to follow, was to have an expert certify that there was very low risk of re-identification of the information. The second path is better for a variety of reasons, including that experts can update their assessment of de-identification and re-identification as technology changes. I offer some observations here about reasons for caution around this issue of de-identification, as well as reasons why the concept can be useful in carrying out network research.

In recent years, there has been greater awareness of a range of techniques that can enable re-identification of data that previously was considered de-identified. Re-identification of data becomes generally easier when: (i) there is more information about an individual available to the researcher, and (ii) the researcher can search that information more effectively. Both are far truer today than at the time of the 1996 Telecommunications Act. The Internet contains a growing amount of information about many individuals, drawn from web pages, blogs, social media, and other sources. In addition, since the incorporation of Google in 1998, effective search has become generally available. With more data and better search, some have argued essentially that de-identification is difficult or impossible. The consequent policy recommendation can be to mis-trust all efforts at de-identification and adopt very broad definitions of the types of data that should be covered by privacy regimes.

Based on my work on de-identification for nearly two decades, I urge the Commission not to adopt such a position. Instead, all those involved should recognize that any privacy regime inevitably has a boundary to draw – some information is linked to an individual and thus covered, and other information is so de-linked from the individual or aggregated that it should be considered outside of the regime. For instance, for an individual, knowledge of the gender of a subscriber in my view is de-identified if there are no other ways to link back to the individual. In addition, information is no longer identified or identifiable for aggregate data, such as that a provider has 48% men and 52% women subscribers. Under Section 222, this latter example would presumably be covered under Section 222's definition of "aggregate information." [12]

12 Late in my preparation of this testimony, I became aware of a petition by Public Knowledge about issues of de-identification and the definition of aggregate information. In the short time available to draft these comments, I was not able to come to an informed view about the details of that petition.

In my thinking about this issue, one crucial but often under-appreciated aspect concerns the complementary role of technical and organizational controls. Technical controls include techniques such as properly hashing the personal information and methods to suppress or fuzz small cell sizes, such as when a census tract only has one residence and thus the individual in the house can be identified. Organizational controls, meanwhile, include separation of identifiers and data within an organization, as well as contractual obligations about how a third party that receives the data may use it. Many of the studies that have reported the ability to re-identify data concern public databases that were posted on the Internet.[13] These databases had technical controls to de-identify users, but any researcher could try attacks on the data until successful. My point is that organizational controls can be an essential and effective complement to technical controls. This idea is developed in detail in a paper from the Future of Privacy Forum, on which I assisted.[14] One critical idea, too often over-looked, is that defense in depth helps here – many databases are maintained (more or less) securely, and the vast majority of researchers seeking to do re-identification cannot get into the database. Then, if good technical controls also exist, the occasional attacker who gets into the database quite possibly is not technically equipped to do the re-identification. In a risk-based system, the likelihood of re-identification is then quite low.

The dual roles of technical and organizational controls have received strong support from the Federal Trade Commission. On the FTC's view, a company's data would not be reasonably linkable to a particular consumer or device to the extent that the company implements three significant protections: "First the company must take reasonable measures to ensure that the data is deidentified." This means reasonable technical measures should be in place. "Second, a company must publicly commit to maintain and use the data in a de-identified fashion and not to attempt to re-identify the data." Third, "if a company makes such de-identified data available to other companies – whether service providers or other third parties – it should contractually prohibit such entities from attempting to re-identify the data."[15]

---

13 One well-known example is Arvind Narayanan & Vitaly Shmatikov, "Robust De-anonymization of Large Datasets (How To Break Anonymity of the Netflix Prize Dataset)" (2008), available at http://arxiv.org/PS_cache/cs/pdf/0610/0610105v2.pdf. For present purposes, my point is that these researchers had full access to the database, and thus could experiment and find a way to re-identify the data. If the data had been held securely in an organization, by contrast, the organizational controls of keeping the database off of the Internet would have greatly reduced the likelihood of any re-identification of the data.

14 Yianni Lagos & Jules Polonetsky, "Public vs. Nonpublic Data: The Benefits of Administrative Control," 66 Stan. L. Rev. Online 103 (2013), available at http://www.stanfordlawreview.org/online/privacy-and-big-data/public-vs-nonpublic-data.

15 Federal Trade Commission, "Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers," at 21 (2012).

Applied to Section 222, this analysis leads me to two recommendations at this stage. First, the Commission should be cautious about an over-expansive definition of what counts as individually identifiable CPNI, or an overly narrow definition of aggregate information. There is considerable utility from analysis of data for many consumer service, consumer protection, and other purposes.[16] The FTC test was developed under the guidance of a noted expert, its then-Chief Technology Officer Ed Felten, and the test applies to data that is "reasonably linkable" rather than simply "linkable" to an individual. The latter – any information that is "linkable" (including under counter-factual conditions) – runs the risk of being far too broad. Second, the Commission should be aware of the complementary role of administrative and technical controls. For instance, the FTC three-part test offers a promising path for network research. Along with technical controls to provide reasonably effective de-identification, research could be authorized conditional on data use agreements or similar administrative controls so that researchers with access to the data undertake an obligation not to re-identify or publicly release the data set.[17]

Conclusion on exceptions. My remarks on exceptions such as anti-fraud, cybersecurity, and research should not be understood as a call for specific, detailed regulations on CPNI and broadband in general, including for the priority uses that I have discussed here. Rather, I offer these observations on important public policy issues so that these compelling interests are recognized early in the Commission's consideration of Section 222 as applied to broadband. These compelling interests should be taken into careful consideration in whatever principles-based, rules-based, or other approach that the Commission might adopt.

Conclusion

I thank the Commission for the opportunity to participate in this important proceeding. The translation of the privacy protections of Section 222 to the broadband sector is not a simple task. There are considerable technical and market differences from the telephone market governed by the 1996 CPNI rules. I commend the Commission for its thoughtful investigation into the technical and market realities of broadband service, for today and for its future development.

16 See Felix Wu, "Defining Privacy and Utility in Data Sets," 84 U. of Colorado L. Rev. 1117 (2013).

17 In mentioning data use agreements, I am not endorsing the specific requirements that HIPAA sets forth for data use agreements, but instead endorse a risk-based, enforceable regime or organizational controls.