Testimony

submitted by

James X. Dempsey
Vice President for Public Policy
Center for Democracy & Technology[1]

to the

Committee on Economic Revitalization and Business
Chair: The Honorable Rep. Angus McKelvey

regarding

H.B. 2288

I. Introduction

A data retention mandate would require companies in the Internet ecosystem to retain certain information about all their users so that it would be available when sought by the government in investigations.[2] Data retention bills have been proposed in the U.S. Congress since 2006 but have never made it to a floor vote because of concerns about effectiveness, cost, and privacy.

H.B. 2288 would impose a data retention mandate on any company that provides access to the Internet. The exact scope of the data that would be required to be retained under H.B. 2288 is unclear: The bill states that "[t]he required data for the consumer records shall include each subscriber's information and internet destination history information." When data retention is discussed, "subscriber information" often is assumed to include the Internet Protocol ("IP") address associated with the communications of a subscriber.

---

[1] The Center for Democracy & Technology is non-profit public interest organization. Based in Washington, DC and with an office in San Francisco, CA, CDT works to keep the Internet open, innovative, and free. With expertise in law, technology, and policy, CDT seeks practical solutions to the challenges of the digital age. CDT convenes a series of working groups that bring together Internet, communications and technology companies, trade associations, think tank, and advocacy groups from across the political and ideological spectrum for dialogue and consensus building.

[2] One stated use of this data is in identifying the source of child pornography. CDT has long worked to protect children in the online environment while at the same time also protecting Internet users' privacy and other civil liberties. See generally, "Data Retention as a Tool for Investigating Internet Child Pornography and Other Internet Crimes," CDT testimony before the House Judiciary Committee's Subcommittee on Crime, Terrorism, and Homeland Security (January 2011) http://judiciary.house.gov/hearings/pdf/Morris01252011.pdf (hereinafter "CDT Testimony").

This testimony analyzes the costs that a data retention mandate would impose on Internet Service Providers (ISPs), mobile carriers and other businesses.[3] It specifically focuses on developments in Internet addressing practices that will make the costs of retaining just one kind of data – IP addresses -- much larger than previously understood. It also explains why, as a result of those same trends in address allocation, IP address data may no longer reliably identify individual end-user devices, thus reducing the usefulness of a data retention mandate.

First, we describe a major development in Internet addressing: ISPs are sharing Internet addresses among multiple customers, which means that IP addresses no longer uniquely identify the computers or other devices of Internet users. (This development, as we explain below, is especially pertinent to H.B. 2288, which seems premised on the assumption that IP addresses are still unique.) We then explain why this trend in IP address sharing means that a data retention mandate would require the collection of vastly larger quantities of data at considerably greater cost than may have been projected even several years ago. We next discuss how the costs of compliance with a data retention mandate would especially harm small ISPs, such as those that serve rural or less populated areas. Finally, this testimony examines the implications of H.B. 2288 for coffee shops, hotels, and other businesses, most if not all of which use address sharing when they provide Internet access for visitors or employees. These entities, if covered by a mandate, would be forced to either assume the huge costs of data retention alongside ISPs or forgo providing Internet connectivity altogether.

II. Changes underway in IP address sharing would render compliance with a data retention mandate extraordinarily expensive

The high capital and operating costs associated with data retention mandates have long been identified as barriers to legislation.[4] However, recent changes in technology will render such mandates even costlier than previously anticipated.

First, some technical background: In the simplest configuration of Internet access, each device connected to the Internet is assigned a unique Internet Protocol address. The "IP

---

[3] In other memos and testimony, CDT has written extensively about the privacy implications of a data retention mandate. See, for example, CDT Testimony, note 2 above.

[4] Capital costs associated with data retention compliance include the costs of designing new collection and storage systems, purchasing collection and storage equipment, integrating new and existing systems, and developing systems to identify and deliver requested data to the government in a timely manner. Key operating costs associated with compliance include the costs of operating and maintaining interfaces for accessing the data in a timely manner, data security, compliance implementation staff, law enforcement liaison staff, staff training, system maintenance, and continuing system integration efforts. *See* Cable Europe, GSMA Europe, EuroISPA, ECTA (European Competitive Telecommunications Association), and ETNO (The European Telecommunications Network Operators' Association), Data Retention: Impact on Economic Operators (2009) at 1-2 (hereinafter "EU Joint Industry Statement"), available at https://www.vorratsdatenspeicherung.de/images/DRconsult/csp_joint_statement.pdf

address" of the device that is the source of a communication is associated with that communication as it is transmitted over the Internet. In some cases, the servers at the destination of the communication – for example, the servers that host the website the user is visiting or the instant messaging service being used – log the source IP addresses associated with each communication that they receive as well as the time of each communication. Government agents may obtain the source IP addresses and timestamps from these destination servers or by other means (such as by seizing and searching the computer of the recipient of the communication). With this information in hand, the government can often identify the ISP or mobile carrier that provided the sender's IP address, as publicly available records show which ISPs and mobile carriers use which blocks of IP addresses. The government can then ask the originating ISP or carrier to determine which customer was assigned the particular source IP address during the relevant time period.

Data retention legislation is intended to require ISPs and mobile carriers, and possibly other entities, to retain logs of the IP addresses they assign in order to be able to connect an IP address obtained by law enforcement at the end point of a communication to a particular customer at the communication's starting point.

H.B. 2288 seems to be premised on the simple configuration of Internet addressing described above. The bill defines Internet protocol address as "a numerical label assigned to each device participating in a computer network … ."

Increasingly, however, ISPs are not using the simple configuration of Internet access described above. Instead, in a growing number of circumstances, IP addresses are being shared among many users, so that the IP address that passes over the Internet is no longer unique to a single end-user device. As we explain below, this change makes it complex and extraordinarily expensive for some ISPs to collect and retain the data necessary to retrospectively connect the source IP address as recorded at the end of a communication to an individual customer.

These changes are being driven by a critical shortage of traditional IP addresses, known as IPv4 addresses. In response to this shortage, key Internet stakeholders have embarked on a potentially decades-long transition to a new addressing protocol, known as IPv6. In the meantime, however, some major Internet access providers are adopting a very complex system of assigning IP addresses.

As a means of conserving IPv4 addresses, some ISPs and mobile carriers have adopted a technology known as Network Address Translation (NAT). NAT allows multiple Internet users to share the same IP address. Until recently, NAT was primarily used at a relatively small scale – for example, to have all of the devices within a single household or coffee shop share one address. However, because the pool of available IPv4 addresses is near exhaustion and the transition to IPv6 has only just begun, many ISPs and mobile carriers have begun or are planning to use NAT on a much larger scale. As a result, in some cases, a single IP address may be shared among thousands of customers. Furthermore, because devices that are only capable of understanding one version of IP or the other

need to communicate with each other during the transition phase, newer flavors of NAT have been developed to translate between IPv4 and IPv6.[5]

NAT usage, whether on a small or large scale, greatly increases the amount of data that must be stored in order to connect particular Internet activity to a specific customer. Below, we explain in more detail why NAT so drastically raises the costs of compliance with data retention mandates.

A. Many IP addresses no longer uniquely identify users or end-user devices

Whenever an Internet-connected device communicates on the public Internet, it is identified by a number called a public IP address, which is typically provided by the ISP or mobile carrier that connects that device to the Internet. Just as a street address sometimes identifies one unique individual, a public IP address sometimes identifies one unique Internet-connected device. However, just as a street address often identifies a multiple members of a family or even a large number of families and individuals, such as all those who live in the same apartment building, NAT allows a single public IP address to identify an entire household, all computers in an organization, or thousands of unrelated customers.

The way this works is that the ISP or carrier sets up a NAT router serving multiple users. Every device behind the router is assigned a private IP address, one that is not seen on the public Internet.[6] When one of these devices initiates a communication, the communication contains the source's private IP address and a number between 0 and 65,535 that is known as a port number.[7]

When the router behind which the device sits receives the source's private IP address and port number, it records them and then associates them with two new numbers: a public IP address that is possibly being used by many other devices sitting behind the same router and a port number that is not being used by any other device sitting behind the router. The ISP or mobile carrier uses what is known as a translation table (hence the name "Network Address Translation") to convert between the private IP address/port number

---

[5] This is a crucial detail, as machines that are IPv4 compatible and machines that are IPv6 compatible cannot easily communicate with each other. Consequently, ISPs must deploy transition technologies, such as NAT, to enable IPv4-capable devices and IPv6-capable devices to communicate with each other, and the use of such transition technologies will be necessary for the foreseeable future.

[6] This system allows ISPs and mobile carriers to use just one of their assigned public IP addresses to serve multiple customers, thus stretching the limited supply of IPv4 addresses assigned to the access providers.

[7] The port number is typically associated with the specific application or process initiating a communication, but the Internet protocol provides for so many port numbers (65,536 of them) that most of them are never used to identify an application. To facilitate IP address sharing, they have been re-purposed as device identifiers.

combination and the public one and thereby to ensure that the devices that share the same public IP address receive only the data intended for their devices.

Moreover, especially in the context of mobile Internet access, the IP address/port number combination for a particular device can change very frequently. Mobile devices can obtain a new IP address/port number combination as frequently as once every minute and possibly even more frequently.[8]

B. NAT complicates compliance with data retention mandates

Even for ISPs or mobile carriers whose networks use an IP address allocation scheme that does not involve NAT, compliance with a data retention mandate can be quite burdensome. IP addresses within these networks may change on a daily or weekly basis and – as we have discussed in past testimony, memos, and papers[9] – the high costs of retaining logs of these changes for six, twelve, or eighteen months can be quite burdensome.

For carriers and ISPs that deploy NAT, the cost and complexity of compliance with a data retention mandate would be especially burdensome. For some networks, new port assignments can occur as often as once every minute.[10] Depending on the type of NAT used, new data may need to be added to the ISP or carrier's logs each time a new port assignment occurs. This data includes a timestamp, outgoing port number, public and private IP addresses, and a link to the customer's identifying information. For a small or medium size ISP, this may amount to a data storage requirement on the order of terabytes of data per day. Under a data retention mandate, ISPs would be required not only to retain this data but also to have the capability to sift through it to satisfy a government demand. (Imagine re-issuing a copy of the White Pages as often as once a minute but still having to maintain all of the old copies.)

As the IPv4 address shortage becomes increasingly severe and the transition to IPv6 progresses, NAT may see even larger-scale deployment. Ensuring end-user identity with the complexities posed by NAT would require a mandate imposing extensive and expensive recordkeeping requirements on a wide range of entities.

---

[8] M. Balakrishnan, I. Mohomed, and V. Ramasubramanian, "Where's that Phone? Geolocating IP Addresses on 3G Networks," The Proceedings of the 2009 Internet Measurement Conference (Chicago, Illinois: Nov. 2009), *available at* http://research.microsoft.com/en-us/um/people/maheshba/papers/ephemera-imc09.pdf (hereinafter "Geolocating IP Addresses").

[9] Erica Newland and Cynthia Wong, "Data Retention Mandates: A Threat to Privacy, Free Expression, and Business Development," Center for Democracy & Technology, Oct. 2011, http://cdt.org/files/pdfs/CDT_Data_Retention_Paper.pdf; John Morris, Greg Nojeim, and Erica Newland, Memorandum on the Data Retention Mandate in H.R. 1981, Center for Democracy & Technology (July 19, 2011), http://www.cdt.org/files/pdfs/CDT_Letter_HR1981.pdf; CDT Testimony, note 2 above.

[10] Geolocating IP Addresses, note 8 above.

C. NAT adds to the already high costs of data retention

H.B. 2288 would require the retention not only of IP addresses but also "Internet destination history information." We are not aware of any cost estimates of such a mandate, since recent federal proposals have focused only on requiring retention of IP addresses. In this testimony, we focus only on IP address retention.

At the federal level, the Congressional Budget Office found that a data retention mandate would impose large up front costs on ISPs.[11] However, it does not appear that the CBO accounted for the added cost introduced by the wider adoption of NAT by ISPs and mobile carriers. Industry representatives, pointing to the new paradigm created by the addressing shortage and transition, have offered far higher estimates of the cost of complying with a data retention mandate.[12] Directly relevant to Hawaii, one small ISP with under 5 million subscribers has told CDT that it could face operating costs of $50 million per year, not including initial capital expenses incurred for the purchase of new equipment and the development of new systems for storing and accessing data. Moreover, in the words of the US ISP Association, cost estimates do not typically account for the "opportunity costs of having [ISPs' technical] experts diverted away from focus on innovating the next generation of Internet-based services."[13]

Finally, the difficulty of retrieving the information sought by the government in a timely manner cannot be overstated. Large-scale data storage increases the likelihood of system crashes and failures; the greater the volume of stored data, the less reliable the integrity of the data and the longer the delays when ISPs respond to demands from government. As the US ISP Association explained in testimony in January 2011, data retention may delay responses in true emergencies because of the slow speed of searching through massive volumes of data.[14] As NAT dramatically increases the volume of data that would be retained, it would also increase the likelihood of delays, errors and crashes.

D. Address sharing reduces the usefulness of data retention mandates

The idea of a data retention mandate was premised on the assumption that an IP address is a reliable Internet identifier. However, with address sharing, to make a match, it is

---

[11] CONG. BUDGET OFFICE, COST ESTIMATE FOR H.R. 1981 at 1 (Oct. 12, 2011).

[12] U.S. House, Committee on the Judiciary. *Protecting Children from Internet Pornographers Act of 2011*. (H. Rpt. 112-281). http://www.gpo.gov/fdsys/pkg/CRPT-112hrpt281/pdf/CRPT-112hrpt281-pt1.pdf.

[13] Written Testimony of Kate Dean (United States Internet Service Provider Association) before the House Committee on the Judiciary, Subcommittee on Crime, Terrorism and Homeland Security on "Data Retention as a Tool for Investigating Internet Child Pornography and Other Internet Crimes," Jan. 25, 2011 (hereinafter "US ISPA Testimony"). *See also* EU Joint Industry Statement, note 4 above ("Furthermore, operational costs are increased by dedicated staff. Often the most qualified engineers, who are being asked to deal with the requests for information from LEAs or to give evidence in Court, are the most expensive and demanded resources.")

[14] US ISPA Testimony, note 13 above.

necessary to know not only the IP address associated with a communication, but also the port number and timestamp. However, the port number information necessary to make a match in a NAT context may not be logged at the destination point. Not all destination servers currently record incoming port numbers and for some it may be difficult or impossible to configure them to do so.

To make a match using NAT tables also requires that the clock used at the destination point to set the timestamp associated with the communication of concern be synchronized with the clock of the originating ISP. However, clocks on the Internet are not perfectly synchronized.[15] If the clocks of the destination server and the Internet access provider are off, even by a few seconds, it may not be possible to make a reliable match, leading to disclosure of data on innocent persons. This can be a problem especially in the mobile context, where the IP address and port number combination for a particular device may change rapidly.

III. Data retention mandates especially burden small ISPs

Many parts of rural America receive broadband services from small ISPs, without which they would remain stuck with slow dial-up services, unable to take advantage of large amounts of the content and services offered through the Internet today. Rural ISPs often serve communities in which larger ISPs have not been willing to invest.

ISPs serving rural or sparsely populated areas typically operate with very small profit margins. The many capital and operational costs of data retention[16] – from the purchase of new equipment to the development of data security measures[17] and systems for retrieving data in response to government demands – would be especially difficult for these ISPs to absorb, especially because small ISPs may deploy NAT in a more complex or layered fashion than do the larger ISPs. The National Telecommunications Cooperative Association (NTCA), a trade association for small and rural telecommunications cooperatives,[18] estimates that complying with the data retention mandate found in H.R. 1981 would create capital costs for a typical rural broadband

---

[15] See, e.g., Paul Krzyzanowski, "Clock Synchronization" (2009) http://www.cs.rutgers.edu/~pxk/417/notes/content/08-clocks.pdf

[16] See note 4 above.

[17] In Europe, despite data security requirements that are written into the data retention law, small ISPs have found it difficult to appropriately secure data. A recent European Commission report found the high cost of implementing security rendered these providers "unable to implement top IT security solutions protecting [retained data.]". *See* Article 29 Data Protection Working Party, Report 01/2010 on the Second Joint Enforcement Action (July 13, 2010) at 6, http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp172_en.pdf.

[18] These cooperatives are often customer owned and supported by the government's Universal Service Fund.

provider that amount to between 5 and 7.5% of its annual revenue.[19] Such a requirement would likely run some of these ISPs out of business, thereby reducing broadband deployment in the United States and exacerbating the digital divide.[20]

IV. Hotels, coffee shops, airports, airplanes, buses, parks, libraries, convention centers and a host of other access providers also use NAT

HB 2288 has an extremely broad definition of Internet service provider: "a company that provides access to the Internet." This could cover not only ISPs but also coffee shops, hotels, airports, and others that offer Internet access to visitors as well as any business that provides Internet access to its employees.

Coffee shops, hotels, convention centers, airports, buses, trains, airplanes, schools, libraries and other entities providing Internet access to users or visitors very likely use NAT technology to distribute IP addresses within their networks. (Indeed, the use of NAT by small establishments predates its adoption at the carrier level.) All of a coffee shop's customers, for example, may sit behind a NAT router with a single IP address. The same complications for data retention that NAT creates for mobile carriers and ISPs are created for the small coffee shop, the hotel, the bus, and the airport. In almost all these cases, whether covered by the bill or not, the public facing IP address passed through the Internet by these entities and recorded at a destination point will not be the IP address assigned to an individual end-user device. Even if a regular ISP were to keep a record of the Internet address assigned to its customer (the coffee shop, hotel, employer), that customer could run a NAT router providing Internet access simultaneously to dozens or even hundreds of other people.[21]

---

[19] National Telecommunications Cooperative Association (NTCA), "Dynamic IP Address Assignment and Tracking," 2011. The costs will vary for each ISP as each network is different. The quoted cost range is for two different models for compliance that NTCA considered. In developing its cost estimates, NTCA made various assumptions about rural telecommunication companies and their existing infrastructure, the need to fully upgrade new infrastructure, the cost of equipment, and the cost to send a technician to each subscriber location (if required under the compliance approach) . These assumptions should not be assumed to be accurate for every network. According to NTCA, the loans required to finance these capital investments would very often be provided by the USDA Rural Utilities Service. However, due to the stringent loan review processes that are in place to ensure the appropriate use of taxpayer dollars, the loan approval process can take up to two years.

[20] Letter from Shirley Bloomfield, CEO, National Telecommunications Cooperative Association to Rep. Lamar Smith, Chair (July 26, 2011)("Finally, the nation's 1,150 rural providers are small businesses that operate on thin margins and lack the economies of scale to absorb a large, sudden cost. The rural telecom industry bears little resemblance to the largest providers, but it is essential to connecting the entire country. NTCA members serve areas where there is no business case for service and others refuse to serve. If rural providers were to exit their markets there would typically be no provider ready to step in and provide the kind of area-wide service that the local and national economies rely on.").

[21] NAT can be layered on NAT. The bus or train that uses NAT may receive its service from a carrier that uses NAT.

HB 2288, if enacted, will have one of two results: small businesses like coffee shops will be covered and will be required to collect and maintain complex records and systems for associating the IP addresses they assign to customers with the public-facing data they pass to the Internet, or coffee shops, hotels and many hundreds of other establishments become a gaping hole in the coverage, and hence the effectiveness, of the legislation. For entities that were covered, the infrastructure needed to store months' worth of records about each customer's behavior would require substantial investment in expensive equipment: the NAT routers these establishments typically use are incapable of keeping persistent logs – they simply don't have the storage capacity. Compliance with a data retention mandate would require these businesses to discard their current equipment and purchase all new equipment at considerable cost. Under HB 2288, many small businesses would likely be unable to continue to offer Internet access.[22]

V. Conclusion

It is widely recognized that a data retention mandate would have serious privacy consequences. Retained information would be available to the government for purposes other than those that prompted introduction of the legislation. Stored data could be vulnerable to hackers or to inadvertent disclosure. There is evidence that the data retention mandate in Europe has had a chilling effect on use of the Internet for provision of important services.[23] A data retention mandate is also likely to chill political use of the Internet and other free speech.

In this testimony, however, we focused on the costs of data retention and, to some extent, on its effectiveness in light of ongoing technological changes.

We recognize that ISPs and mobile carriers retain certain authentication data and certain IP address data for business purposes. Service providers are diligent in cooperating with government officials to provide whatever data they store. However, there is a world of difference between collecting and retaining data for business purposes and collecting, retaining and being able to retrieve that data for the purposes the government has in mind

---

[22] Regulatory burdens are, as a general matter, disproportionately borne by small businesses since they tend to be ill-equipped to absorb and comply with unfunded mandates. Nicole V. Crain and W. Mark Crain, *The Impact of Regulatory Costs on Small Firms*, U.S. Small Business Administration, Office of Advocacy (September 2010) at iv, *available at* http://archive.sba.gov/advo/research/rs371tot.pdf ("Small businesses . . . bear the largest burden of federal regulations. . . . [S]mall businesses face an annual regulatory cost . . . which is 36 percent higher than the regulatory cost facing large firms.").

[23] See Axel Arnbak, Plenary Presentation at the Taking on the Data Retention Directive Conference in Brussels: What the European Commission Owes 500 Million Europeans (Dec. 3, 2010) at 3, available at http://www.edri.org/files/Data_Retention_Conference_031210final.pdf (finding that as a result of a German data retention law, "half of Germans will not contact marriage counselors and psychotherapists" via e-mail), citing a German-language study by FORSA, "Opinions of citizens on data retention," June 2, 2008, available athttp://www.eco.de/dokumente/20080602_Forsa_VDS_Umfrage.pdf.

(uniquely identifying end-user devices). In this testimony, we have explained why that gap between business practices and a data retention mandate is growing even wider. Increasingly, the data retained for business purposes (at the beginning point of a communication, at the network level, and at the end points) is very different from the data that would have to be retained under a data retention mandate.

In the changing Internet ecosystem, data retention has become far more complex than even we at CDT understood several years ago. The evolution of IP address assignment practices has vastly increased the amount of data providers would have to retain in order to comply with H.B. 2288. Even with modern storage capabilities, the volume is so huge that the costs would be enormous, hurting especially small carriers serving rural communities, as well as coffee shops, hotels, and others that provide Internet access. This would slow or even reduce broadband deployment and divert financial and technical resources away from innovation.

Meanwhile, under current law, government already has the authority to require carriers to provide addressing data regarding specific accounts. State and local, as well as federal investigators in Hawaii have the authority, under 18 U.S.C. 2703(f), to require providers to preserve IP address and other information retrospectively on specific accounts. In addition, providers have a current obligation to preserve identifying information associated with child pornography that they find on their systems. These methods are highly effective in that they focus on specific users or accounts. These methods provide investigators with information relevant to a specific investigation and do not require the retention of massive amounts of information that will never be part of an investigation.

Mr. Chairman, members of the Committee, we appreciate the opportunity to submit this testimony. We would be happy to answer any further questions that you or your staff would have. Fell free to contact Jim Dempsey (jdempsey@cdt.org) at 415-814-1712.