The ISP Column
A monthly column on things Internet

August 2015
Geoff Huston

The Global Village Idiot

I recall from some years back, when we were debating in Australia some national Internet censorship proposal de jour, that if the Internet represented a new Global Village then Australia was trying very hard to position itself as the Global Village Idiot. And the current situation with Australia's new Data Retention laws may well support a case for reviving that sentiment. Between the various government agencies who pressed for this legislation, the lawyers who drafted the legislation, the politicians who advocated its adoption and the bureaucrats who are overseeing its implementation, then as far as I can tell none of them get it. They just don't understand the Internet and how it works, and they are acting on a somewhat misguided assumption that the Internet is nothing more than the telephone network for computers. And nothing could be further from the truth.

The intended aim of this legislation was to assist various law enforcement agencies to undertake forensic analysis of network transactions. As the government claims: "telecommunications companies are retaining less data and keeping it for a shorter time. This is degrading the investigative capabilities of law enforcement and security agencies and, in some cases, has prevented serious criminals from being brought to justice." (https://www.ag.gov.au/dataretention). So what the agencies wanted was a regulation to compel ISPs to hold a record of their address assignment details so that the question "who was using this IP address at this time" had a definitive answer based on the retention of so-called meta-data records of who had what IP address when.

In the world of traditional telephony this makes some sense. Telephone numbers were synonymous with endpoint identifiers, so that a telephone number was uniquely associated with a subscriber, and this association was stable and long-lived. Asking phone companies to hang on to the association of telephone numbers to subscriber names and addresses, or in other words a phone directory, was hardly an onerous imposition to the industry, and considering that the phone directory was public it was hardly a dramatic incursion into untested areas of personal privacy. In the context of these Data Retention laws noone is asking service providers to record and retain conversations. Noone is even asking to keep the records of what numbers were called by subscribers, although my telephone bill clearly demonstrates that my phone company collects and stores all such individual call records. The data retention measures are explicitly limited to the association of telephone numbers to subscriber details. In this world of black bakelite telephones of the 1950's I'm sure that this was a fine idea, and was, in fact, little more than a formal codification of existing practice in many telephone operators then and now.

But that was then and this is now, and today the telephone system is heading to a role of quaint historical artifact, while the Internet continues to take on the role of the global communications platform. So data retention for telephones is hardly useful. Something has to be done about the Internet. Doubtless someone had the bright idea that if they took this concept of the association of telephone number to subscriber, and used a text editor to globally change "telephone number" to "IP address" in the text of a data retention piece of regulation then they would have a bright shiny piece of regulation that would make them all set for this brave new Internet world. After all, the Internet is just a telephone service for computers isn't it?

But that is not the case in today's Internet. It has been a constantly changing environment that has responded and adapted to various pressures over time. One of the more critical long term pressures on the Internet's architecture has been the prospect of address exhaustion, which has been a pervasive influence for over two decades now. The result of this prospect of address exhaustion has been to change the semantics of an IP address. Because addresses were considered to be a scarce resource the change was to use them sparingly, and the way to achieve that was to share an address across multiple devices. This sharing has increased in intensity over the years. The initial model was to place address sharing units, or Network Address Translators (NATs), at the "edge" of the network, where the carriage network connects to the customer's network. In this model the IP address is now shared by all the units located on the customer's network. As a result an IP address is still, in some sense, an edge point identifier, but the endpoint is now a home network, not a single device. But even so that was then and this is now, and the address sharing picture has changed further.

We are now seeing these address sharing units being pulled further back into the service provider network. This has started with mobile networks, but is now also occurring on wired access networks as well. The inexorable pressures of address exhaustion are driving many service providers into these address sharing approaches for their network. What does an IP address mean when its shared in this manner? It's no longer synonymous with an endpoint identifier, as a number of endpoints may be sharing this single public IP addresses. Equally, a single endpoint may use a number of public addresses, even at the same time in some situations.

So what is an "IP address" if it's not an endpoint identifier? It is now an ephemeral shared token whose contextual lifetime in the public network is that of single network transactions, and it's use is never assuredly unique, even within such limited contexts.

So if IP addresses are losing their role as stable endpoint identifiers what has taken their place? What should we be storing in some data retention framework that relates a network transaction to an endpoint? If storing IP addresses makes no sense as an endpoint identification what should we use instead? The hard answer is that we don't have such a concept any more, and it's ok that we don't. We've managed to convince ourselves that the Internet does not need them. And that's a big statement.

Today's Internet has no strict requirement for universal stable fixed endpoint identities. And things work just fine. What we have found is that in a client/server service model there is no need to assign fixed endpoint identities to the clients. They can get away with pulling out ephemeral tokens from some shared pool and everything still just works. And these days we are also experimenting with Content Distribution Network (CDN) service solutions that allows the servers to also use IP addresses in the same ephemeral manner, relying solely on the DNS as the service point identification space. So addresses in the Internet don't mean all that much any more, and increasingly they don't map to endpoint identifiers any more.

But the Australian Data Retention Laws say something has to be stored, and the bureaucrats running the Attorney General's Office of Data Retention say something has to be stored, and the industry players are trying to understand what exactly should be stored, because in shared address-based networks there is nothing around that meets the intended requirements of this law. If the intended requirement of this law is to retain the association of protocol-specific endpoint identifiers with the customer's details, and the network has now managed to eschew the very concept of stable endpoint identifiers, then where have we got to?

It is unlikely that operators of these address sharing networks will refuse to comply with the provisions of the Data Retention laws. It's likely instead that they will use the address sharing logs and retain them. But this starts to get interesting, because in theory in order to retain something the temptation will be to retain the complete log from the network address sharing unit. What is in this log? In this world of Carrier Grade NATs (CGNs) every transaction generates a new NAT binding, and that NAT binding generates a log entry. So every DNS query, every part of every web page, every individual email collected by your device - in short, each and every individual network transaction - will generate a CGN log entry. This is no less than your entire Web browsing history, your DNS query history, and the history of everything else you are doing on the net.

So when the bureaucrats claim that: "The Australian Government is not requiring industry to retain a person's web-browsing history or any data that may amount to a person's web-browsing history." (https://www.ag.gov.au/dataretention) are they just lying, or do they really mean that operators of CGNs do not need to retain any of this data, making CGN-based networks truly opaque and anonymous? I strongly suspect the former - they are indeed lying and everything you do on the net will be logged and retained. However it's not intentional duplicity. They just don't get it. Because we really are trying hard in Australia with our national branding in the Internet. We are trying as hard as we can to retain the role of Global Village Idiot.

Author

Geoff Huston B.Sc., M.Sc., has been closely involved with the development of the Internet for many years, particularly within Australia, where he was responsible for building the Internet within the Australian academic and research sector in the early 1990's. He is author of a number of Internet-related books, and was a member of the Internet Architecture Board from 1999 until 2005, and served on the Board of Trustees of the Internet Society from 1992 until 2001. He has worked as a an Internet researcher, as a ISP systems architect and a network operator at various times.

www.potaroo.net

Disclaimer

The above views are solely the author's opinions, and certainly do not represent the views or positions of APNIC. APNIC will not be legally responsible in contract, tort or otherwise for any statement made in this article.