July 2019
Geoff Huston

Not So Private Thoughts at IETF 105

At IETF 105, held in Montreal at the end of July, the Technical Plenary part of the meeting had two speakers on the topic of privacy in today's Internet, Associate Professor Arvind Narayanan of Princeton University [1] and Professor Steven Bellovin of Colombia University [2]. They were both quite disturbing talks in their distinct ways, and I'd like to share my impressions of these two presentations and then consider what privacy means for me in today's Internet.

Firstly, Arvind Narayanan considered some lessons from privacy measurement.

Some 10 years ago the Electronic Frontier Foundation released a service called *Panopticlick* (https://panopticlick.eff.org). The underlying motivation behind this site was that in many cases (evidently over 90% of cases, as originally reported) users' browsers generated a unique browser fingerprint. Irrespective of any browser cookie settings and certainly without the explicit knowledge and consent of the user, the browser could be coerced to leave a digital trace at the web sites it visited that if it were pieced back together it would allow an individual's browsing history to be reassembled. A similar site, amiunique by Inria (https://amiunique.org/), also is based on the assumption that many users are unique, in that they leave behind unique fingerprints in the logs of every site that they visit. It certainly appears that in terms of privacy protection on the Internet, this form of tracking of individual users was "devastatingly effective". So effective was this technique that any efforts to reclaim the pre-fingerprinting state of blissful ignorance, and presumably blissful containment of personal digital privacy, would just be too little too late.

But there is always a risk in such measurement studies of self-selection bias. Subsequent significantly larger volume studies that looked at the effectiveness of browser fingerprinting at large scale by embedding a uniqueness test in several very high-volume websites [3] reduced this 90% uniqueness level to less than a fifth. As the plug-in environment changes and as Flash and Java are phased out in the webspace, this number may well be dropping further. The underlying message is that this larger scale measurement gives us reason to believe that we have not yet lost the privacy battle in this space and tracking and tracing tools based on browser fingerprinting remains ineffectual for any useful purpose. And it's also the case that awareness of this threat has motivated technology developers, particularly in the browser world, to use measures such that each user does not leave behind a unique trace of their activities.

His conclusion from this is that careful measurement can help us to both understand the risks and assist us in working out how to respond to some classes of threats to privacy. Where else could measurement help?

It appears that societal attitudes toward privacy evolve rapidly. Trade-offs between privacy and access to goods and services will change over time, and at times they change rapidly. Technical standards and normative specifications do not have such adaptive capability and can fail to keep track of current expectations and norms.

The widespread use of Facebook "Likes" has been used as a means of predicting personality traits [4]. This was allegedly used by Cambridge Analytica for targeting. Shopping habits are extensively analysed and when combined with behavioural profiling, vectors of suggestibility are exposed. Given the extremely high potential value in such activities when undertaken at scale, it is unsurprising that such technologies that infringe privacy in these ways evolve quickly. To keep pace, we need to bring privacy research into the association between technology standards and technology developers. Research needs to be incentivized to review technology standards in-depth and also study application behaviours in the deployed network.

Underpinning this research is the crucial role of measurement. Use, or abuse, of technology to undermine user's expectations of privacy are held in check through credible public exposure through measurement and analysis. Not only does this inform the public, but it can influence situations of information asymmetry and motivate technology developers to understand how technology can be used and potentially abused. Obviously, such measurement and analysis can also inform, and even help shape, the public policy process.

The issues relating to privacy extend well beyond browsers and profiling. This extends into a world where data collection happens in many ways and the means of monetisation of this data often are completely unknown to the individual. A major US mobile carrier was selling location data of connected devices to data aggregators. Automobile servicing enterprises sell the service data of individual cars. Microphones have been embedded in domestic thermostat controllers. The Internet of Things seems to be nothing less than a looming digital catastrophe in many ways, including the comprehensive destruction of privacy. Again, the feedback loop of observation, measurement, analysis and reporting might offer some productive approaches to combine regulatory responses with technical measures that just might ameliorate the worst of this potential situation.

I would characterise Arvind Narayanan's position as "cautiously more on the optimistic side" on the prospects of the future of privacy. I sensed an opposite stance from Steven Bellovin’s presentation

Steven Bellovin argues that privacy concerns are certainly not a new issue. He cites the Committee on Science and Law of the New York City Bar Association that commenced a formal study on privacy in 1962, leading to Alan Westin's book "Privacy and Freedom" that reported on this committee's work, published in 1967 [5]. Within the United States, a cornerstone of privacy is the concept of Notice and Consent, or informed decision making by individuals, a position that appears to have been based, or at least strongly influenced by this study. Westin noted that "A central aspect of privacy is that individuals and organisations can determine for themselves which matters they want to keep private and which they are willing - or need - to reveal." Subsequent measures in both the United States and in Europe continue this foundation of Notice and consent in regulating an individual's right to privacy.

Most of the technical challenges to privacy have also been understood for many years. The deliberate obfuscation of consent notices to allow all forms of overreaching, and the ease of obtaining such consent when the notice is simply incomprehensible to most consumers have been longstanding issues. The production of metadata and the interpretation of this data to recreate individual profiles is nothing new, as are the perils of massive, yet easily searchable, databases.

Given that these abuses have been occurring for decades it is reasonable to ask if Notice and Consent is adequate today? It appears not to be the case for Steven.

One of the problems of Notice and Consent is that data collection is not an activity that is exclusively centred around the individual. We generate vast amounts of data, from the page reading rate on our ebook readers through to the temperature settings on the home thermostat. Data is shifted across borders with an apparently cavalier attitude to national regulatory measures, and the activity of data aggregation is as much an activity of government agencies as it is a private sector activity. The metadata retention measures in Australia have been used by local councils and even the Royal Society for the Prevention of Cruelty to Animals (RSPCA) Victoria to access the collected metadata of individuals' online activities [6].

It seems that data collection is everybody's business all of the time. Notice and Consent is no longer part of this picture. If you have a credit card, or a car, a digital phone, or go shopping, or use a car, or ride public transport or undertake just about any other human activity, then data about your actions will be generated, stored, traded and analysed in all kinds of ways that extend way beyond your environmental sphere of awareness and way beyond your realm of Notice and Consent. This is not just keyboards, mobile devices and browsers any more. We now have a situation which could be best termed "overcollection" and there are now data wholesalers who aggregate and sell vast quantities of data at a position well removed from the individual subjects.

The notice part of this Notice and Consent structured privacy regime appears to have been comprehensively abused. They are vague, lengthy often obtuse. The consent being sought is often completely open-ended and embedded in fine print that is practically inaccessible. Only in some fantasy world can one assume that users actually read these notices and provide informed and considered consent. It just does not happen.

Steven Bellovin's conclusion is that Notice and Consent is dead. No one knows who collects data, no one knows where it's stored. No one knows what they will do with it.

But if Notice and Consent is dead what should replace it?

In his presentation, he proposes a concept of Use Controls. Rather than providing consent in every individual instance, users should be able to specify how their data can be used by any and all data acquirers, aggregators, analysers and handlers. Whether it's permitted for targeted search advertising, statistical analysis, public census or medical research, it's up to the individual consumer to define a profile of permitted use of data that concerns them.

It may sound simple, but of course, there are many areas of underlying complexity. How are such controls to be defined? What time duration is spanned by such permissions? While it may outlaw misuse, detecting and prosecuting such misuse may be exceptionally challenging. How would it function on indirect data acquisition and how could it apply to profiling data where your digital profile is still of value even without knowledge of your individual identity?

In any case, if we move away from Notice and Consent, which seems only logical today, then we necessarily are moving to a new paradigm of data handling, whether its Use Controls or any other framework. Such a paradigm has to scale across a broad diversity of actors and across times and regulatory regimes. It has to be comprehensible and enforceable and account for both primary data and indirectly generated or inferred data.

Obviously, we don't have such a privacy framework yet. Notice and Consent is failing badly, and its prospects don't look like changing. Use Controls are some way away, even assuming that we'll head in this direction anyway. What can we do now?

What is the message to technology developers? Steven suggests that perhaps the best the IETF can do at present is to force data collection into clear sight, and actively prevent casual indirect data collection through covertly sniffing the digital exhaust fumes that we constantly generate. Traffic encryption at every opportunity helps, in so far as the primary data transaction is limited to the client and the server and intermediaries may be more challenged in their efforts to perform data harvesting by eavesdropping.

This message is certainly bleaker in tone, but he hasn't given up all hope just yet.

Personally, I'm not sure even this will change the current situation and alter its trajectory in any way.

Surveillance Capitalism is now at the core of the wealth of the majority of the world's most valuable public companies. They are valued not by the value of their own production of goods and services in a

traditional sense of the economics but valued as a percentage of the net worth of the now billions of individual subjects of whom they have amassed detailed individual profiles. But even this is not quite the case, as the value multiplier also includes the expected future worth of these same individuals. As my total future spending expectations declines as I age, my net value to this surveillance economy also declines. What does this say about the value and consequent intensity of digital surveillance of our children?

If the major economic entities in today's economy are ruthlessly exploiting the weaknesses in existing privacy structures, then what hope do we have to change the rules to rein in the very behaviour that is generating this vast economic wealth for a select few?

Public Policy: If lobbying politicians cause favourable outcomes, then Alphabet's $20M spent lobbying in 2018, and Facebook's $13M should assist them in ensuring that their data collection and analysis practices are not reined in through changes to public policies in the US at least. [7]

Public Opinion: What sways public opinion? Increasingly it seems that we are victims of these digital natives, where cynical manipulation of public sentiment is being practiced by both enterprises and local and foreign government agencies. That massive data vacuum, social networking, appears to be a remarkably hostile social weapon.

There is a bleaker conclusion that can be drawn from this situation, and its one that I am reluctantly drawn to.

For me, it's not the erosion of personal privacy that is the issue here. I think that is largely a lost cause in this world, and all the regulation, fines, encryption and use control tagging is not going to bring it back. Personal privacy is hopelessly lost in today's world, and it's never coming back.

Our actions are comprehensively observed, archived and analysed both now and in the future, and in a myriad of ways that we only dimly understand as individual subjects of this scrutiny. It's not just a trade of data collection and predictive analysis. The most valuable rewards of Surveillance Capitalism are available to those who can accurately identify those moments of human vulnerability, where just the right prompt in the first of a suggestion takes on an unstoppable momentum. As Hal Varian, the Chief Economist of Google said many years ago (in 1998 at a Global Internet Summit that I attended, and I'll paraphrase his words here) spam [or unwanted advertising in all its forms] is merely a failure of information. If the advertiser had access to better information about each consumer then each and every ad would be either a helpful suggestion or an impossible to resist temptation!

At this stage, we appear to have reached a rather odd place. If the result of this comprehensive digital observation and analysis is gaining a thorough understanding of what we want and what we could obtain, then what's the problem? If my personal digital world is devoted to identifying my personal preferences and informing me of how I could fulfil my needs and desires, then surely all this effort is being devoted to achieving for me what I personally want. Where's the problem?

I worry that the outcome of this digital pandering to my supposed needs is not necessarily a better world for all, or even for me. The digital divide still exists, and the benefits that are realised by the privileged few often come at a price that others who are less fortunate have to pay. The benefits of this digital world may be little more than a digital dividend for the first world, leaving a trail of the disenfranchised and exploited behind them.

What we place at risk as a human society is the very fabric of why we chose to live in societies in the first place. What profoundly worries me about such digital selfishness is that we risk the very cohesion of our human society as we know and understand it. We may lose the essential recognition that by working together and pooling our individual efforts to a common good we can sustain a society that fairly delivers benefits to all.

If the enduring cost of the industrial age was the destruction of our natural environment, will the true cost of this digital age and its inexorable thirst for data turn out to be nothing less than the destruction of our societal structure of common humanity?

References

[1] Narayanan: Lessons from Privacy Measurement, IETF 105 Technical Plenary, July 2019. https://datatracker.ietf.org/meeting/105/materials/slides-105-ietf-sesse-lessons-from-privacy-measurement-arvind-narayanan-00.pdf

[2] Bellovin: Privacy: modern Concerns, IETF 105 Technical Plenary, July 2019. https://datatracker.ietf.org/meeting/105/materials/slides-105-ietf-sesse-privacy-modern-concerns-steven-m-bellovin-00.pdf

[3] Gómez-Boix et al.: Hiding in the Crowd: An Analysis of the Effectiveness of Browser Fingerprinting at Large Scale. WWW 2018. https://www.doc.ic.ac.uk/~maffeis/331/EffectivenessOfFingerprinting.pdf

[4] Kosinski et al: Private traits and attributes are predictable from digital records of human behavior. PNAS 2013. https://www.pnas.org/content/110/15/5802

[5] Westin: Privacy and Freedom, Ig Publishing, First Published 1967.

[6] https://www.zdnet.com/article/61-agencies-after-warrantless-access-to-australian-telecommunications-metadata/

[7] https://www.businessinsider.com/r-google-spends-big-on-us-lobbying-amid-antitrust-bias-battles-2019-1

Disclaimer

The above views do not necessarily represent the views or positions of the Asia Pacific Network Information Centre.

Author

Geoff Huston B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region.

www.potaroo.net