June 2019
Geoff Huston

Network Protocols and Their Use

3 – Security and Privacy

In June I participated in a workshop, organized by the Internet Architecture Board, on the topic of protocol design and effect, looking at the differences between initial design expectations and deployment realities. These are my impressions of the discussions that took place at this workshop.

In this third part of my report I'll report on our experience with security and privacy.

These days any form of consideration about the Internet and its technology base needs to either address the topic of security and privacy in all its forms, or explicitly explain this glaring omission. Obviously, this workshop headed directly into this space, asking whether the IETF was looking at topical and current threat models, and also asking about likely evolution in this space.

Exhortations about security practices for service operators made through standards bodies are often ineffectual in isolation. RFC 2827 is almost 20 years old, and it ignored by network operators to about the same extent that it was ignored at the time of its publication. It may be better known as BCP 38, or packet filtering to prevent source address spoofing in IP packets. It's important because there is a class of DDOS attack using UDP amplification where the UDP response is far larger than the query. It's a fine practice and we should all be doing this form of filtering. Twenty years later the attacks persist because the filtering is just not happening. What may make such forms of advice more effective is the association of some form of liability for service operators, or explicit obligations as part of liability insurance. In isolation, advice relating to security measures is often seen as imposing cost without immediate direct benefit, and in circumstances such as this case, where the defensive approach is only effective when most operators undertake the practice, benefits for early adopters are simply not present.

Another example is the standardization of Client Subnet extensions in DNS. Despite the standard specification RFC 7871 containing the advice that this feature should be turned off by default and users be permitted to opt out, this has not happened. This is in spite of the potential for serious privacy leak through attribution of DNS queries to end users.

The environment of attacks escalates, as the growing population of devices allows the formation of larger pools of co-opted devices that in concert can mount massive DDOS attacks. Given our inability to prevent such attacks from recurring, the reaction has been the formation of a market in robust content hosting. As the attacks increase in intensity the content hosting operators require larger defensive measures and economies of scale in content hosting come into play. The content hosting and associated distribution network sector is increasingly concentrated into a handful of providers. In many ways this is a classic case of markets identifying and filling a need. The distortion of that market into a very small handful of providers is a case of economies of scale coming into play. As with the CA market, the market has now seen the entrance of zero cost actors, which has significantly lifted the barrier to further new entrants in this market. What remains now appears to be simply a process of further consolidation in the market for content hosting.

The threat model is also evolving. RFC3552, published in 2003, explicitly assumed that the end systems that are communicating have not themselves been compromised in any way. Is this a reasonable assumption these days? Can an application assume that the platform is entirely passive and trustable, or should the application assume that the underlying platform may divulge or alter the application's information in unanticipated ways. To what extent can or should applications lift common network functionality into the user space and deliberately withhold almost all aspects of a communication transaction from other concurrently running applications, from the common platform and from the network? Do approaches like DOH and QUIC represent reasonable templates for responding to this evolved threat model? Can we build protocols that explicit limit information disclosure when one of the ends of the communication may have been compromised?

Is protocol extensibility a vector for abuse and leakage, such as the Client Subnet DNS extension in the DNS, or the session ticket in TLS?

And where are our points of trust to allow us to validate what we receive? As already pointed out DNSSEC is not faring well, and the major trust point is the WebPKI. Unfortunately, this system suffers from a multiplicity of indistinguishable trust and our efforts to detect compromise have shrunk to logging, in the form of Certificate Transparency. Such a measure is not responsive in real time and rapid attacks are still way too effective.

A single trust anchor breeds a natural monopoly at the apex and across the diversity of the global Internet there is a lot of distrust in that single point, particularly when geopolitics enters the conversation. This single trust broker is a natural choke point and is one that tends to drift towards rent seeking if operated by the private sector and distrust if operated by the public sector. Designs for trust need to take such factors into account.

The issue of security popups in the browser world can be compared to the silent discard of the response in the DNSSEC world, as they offer two different views of security management. Placing the user into the security model leads to lack of relevant information and an observed tendency for the user to accept obviously fraudulent certificates because of no better information. From that perspective removing the user from the picture improves the efficacy of the security measure. On the other hand, there is some disquiet about the concept of removing the user from security controls. Giving the user no information and no ability to recognize potentially misrepresenting situations that may them seems to be a disservice to the user.

Do our standards promote and encode the "state of the art" as a means of shedding liability for negligence while still acknowledging that the state of the art is not infallible? Or do they purport to represent a basic tenet of security that is correctly executed is infallible?

The Internet of (insecure) Things is an interesting failure case, and the predatory view of the consumer often distracts from the ethos of care of the customer and the safeguarding customer's enduring interests. Grappling with conformance to demanding operational standards in a low cost highly diverse and high-volume industry is challenging. Perhaps more so is the tendency of the IETF to develop many responses simultaneously and confront the industry with not one but many measures. Already we've seen proposals that use some level of manufacture cooperation, such as nesting public/private key pair, QR code, MUD profile or boot server handshake.

A safe mode of operation would require that the device cannot cold start, nor even continue to operate without some level of handshake. Is this realistic? Will manufacturers cooperate? Will this improve the overall security of the IoT space. Are these expectations of manufacturers realistic? Will a kickstart IoT toothbrush comply with all these requirements? Will these requirements impose factory costs that make the device prone to manufacturer errors and increase the costs to the consumer without any change in the perceived function and benefit of the device? An IoT toothbrush will still brush teeth irrespective of the level of conformance to some generic standard security profile. The failure in the October 2016

botnet DNS attack that used readily compromisable webcams was not a failure of information or protocol. It was a failure of markets, as there was no disincentive to bring to market an invisibly flawed, but cheap, and otherwise perfectly functional product. We tend to see the IoT marketplace as a device market. In contrast, effective device security is an ongoing relationship between the consumer and the device manufacturer and requires a service model rather than a single sale transaction.

The prospect of regulatory impost to provide channels to the retail market that include conformance to national profiles is nigh on certain. Will the inevitable diversity of such regional, national or even state profiles add or detract from the resultant picture of IoT security? Will we end up with a new marketplace for compliance that offers an insubstantial veneer of effective security for such devices? It's very hard to maintain a sunny optimistic outlook in this space.

Human behavior also works against such efforts as well. Our experience points to an observation that users of a technology care a whole lot less about authentication and validation than we had assumed was the case. Most folk don't turn on validation of mail, validation of DNS responses and similar, even when they had access to the tools to do so. When we observed the low authentication rates post-deployment, our subsequent efforts to convince users to adopt more secure practices were ineffectual. Posters in the Paris Metro informing metro users as to what makes a password harder to guess really have not made an impact. In the consumer market users don't understand security and don't value such an intangible attribute as part of a product or service.

Safegarding privacy is a similarly complex space. The last decade has seen the rise in surveillance capitalism, where the assembling of individual profiles consumers has become the cornerstone of many aspects of today's Internet. Many products and services are provided on the Internet free of charge to the user. The motivation to provide such free services comes from the reverse side of this market, where the tool of service is used to assist in the generation of a profile of the individual user, which is then sold to advertisers. Our digital footprint can provide a rich vein of data to fuel this world of surveillance capitalism. Whether its our browser history, logs of DNS queries, our mailboxes, search history, or documents, e-book purchases and reading patterns, all of this data can be converted into information that has monetary value. Our attitude to this activity is not exactly consistent. On the one hand we appear to be enthusiastic consumers of free-of-charge products and services and all too willing to dismiss reports of data mining on the basis that individually none of us have anything to hide. At the same time, we all have experienced those disconcerting incidents where the delivered ad mirrors some recent browsing topic or received email. Why is privacy a common concern when our actions appear to indicate that we are willing to trade it for the provision of goods and services?

One reason for this concern is when a principle of informed consent is violated. Protocols, products and services should not facilitate unintended eavesdropping on a user's actions and activities. When they leak personal information without such informed consent there is a reasonable reaction over what is perceived to be unacceptable surveillance. Another reason lies in the inherently asymmetric nature of the market of personal profile data. Individual users tend to undervalue their profile data, and the relationship with the consumers of such data tends to be exploitative of individual users.

The IETF's position on privacy has strengthened since the publication of RFC 7258 in May 2014, and the IETF's expectation is to go to some lengths with information management in its protocols to contain what is now seen as gratuitous leakage. This includes measures such as query name minimization in DNS queries and encryption of the SNI field in TLS handshakes.

It would be good to think that we have finally stopped using the old security threat model of the malicious actor in the middle. We have more complex models that describe secure and/or trusted enclaves, which an unknown model of the surrounding environment. Is this device security or really a case of "data security"? We need to associate semantics with that data and describe its access policies to safeguard elements of personal privacy.

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