Incentives for Privacy

Cullen Jennings
fluffy@cisco.com

The reasons for protecting privacy often conflict with the economic incentives of those who make internet applications and even with the habits of individual users. Privacy is far more about people and incentives than it is about technology or cryptography which makes it a difficult topic for an engineering organization to address.

1. Incentives

A wide variety of applications and services on the internet use IETF or W3C defined protocols that impact end users' privacy in one way or another. The bulk of W3C and IETF standards are written by people who work for a set of operators that often derive large revenues from advertising, or else work for vendors that receive a large fraction of their revenue selling to such operators. For the most part, large websites that generate substantial revenue from advertising have very little incentive to protect most forms of end user privacy. On the contrary, knowing more about the users of the site allows them to generate more advertising revenue. Some of the most important issues in what happens in web browsers or HTML 5 are ultimately decided by a very small number of individuals working directly or indirectly for companies that all have large and growing revenue from online advertising.

Designing protocols to protect privacy would typically take more time and energy to design and build. Their end result, frequently, would be to reduce the revenue potential for the organization that provides the service. Internet privacy currently largely rests in the hands of businesses that face serious disincentives to tightening privacy.

2. State of Privacy

Given the incentives of various players, it is interesting to look at what has happened for web privacy.

Selling voice recordings

Long before VoIP was common, service providers were generating revenue selling wiretap to law enforcement organizations. This revenue stream has continued into the VoIP era with large, regulated carriers. It is less clear whether services that provide voice and video communications in an IM-like client also sell this information. The major players in this space refuse to disclose their practices, but it is clear that selling this kind of information is technically feasible.

Keystroke monitoring

More web services, such as search engines, are moving to sending each key stroke or key stroke timing information to the server. This allows for the creation of systems that recognize users by the cadence of their key strokes. The research in this area is proceeding energetically, and it is feasible for many situations. The biggest problem is collecting a large training database.

Voice print monitoring

Recognition of users by their voices is a technology that has been steadily improving over the years. Capturing a large data base of voice samples with known speakers is one of the key things needed to make this work better.

Image and Video monitoring

The state of recognition of faces from pictures or video is steadily improving and works fairly well when the system only has to identify a user from a fairly limited set of possibilities. It does not work anywhere near as well as the number of possible people that might be in the image moves up to the thousands or millions, but that is improving too.

3. Graphs versus Content

I expect to see businesses focus increasingly on tracking who talks to whom. This information, which provides social interaction graphs, is probably more valuable for advertising purposes than the details of what the people talk about. This information can be valuable for other reasons too, and it is much harder to protect than the actual contents of the conversation.

Often casual commentators do not even imagine why it could be an invasion of privacy to know with whom they are communicating. One has only to think of a group on a social networking site dedicated to a particular, banned religion, though, to get a sense of the risks.

4. What Next

There are two minor technical things that might help reduce the issues and tighten privacy protection, but real change is most likely to come from changing the incentives.

Big random numbers are unique

Protocols often need a unique identifier with a temporary life span. Unfortunately much of the technical community is very leery of random numbers so they choose some way to allocate unique identifiers instead. These can be useful for tracking and identifying users. Take MAC addresses for example: if hosts just picked random MAC addresses, the system would probably work just fine.

Its OK to use time

Many designs could be improved by hashing the information transmitted over the internet with a windowed time. This is often rejected as a design on the premise that it forces the endpoint to have a rough idea of time. For the types of endpoints that usefully identify a user, this just seems to no longer be true. It is incredibly cheap and common for devices to have an approximate idea of time and we should stop designing as if they did not.

5. Incentives

Economic incentives to not lose data

Personal data that is only needed for "secondary use" is collected because it can be readily sold for this secondary use. However, it is also often lost or compromised because there is no real incentive to protect it. If there were real economic incentives to safeguard this data, there would be a lot less of it collected.

When is regulation the least bad answer?

The normal knee jerk reaction to regulation in the internet community is that no matter how bad things are, it would be worse if the government was involved.

The current arrangement regarding who pays for the design of privacy on the internet is well beyond appointing the fox to guard the henhouse. At least the fox might stop eating chickens once it was full. The internet privacy situation is more like asking the *National Enquirer* to decide what sort of photos of Britney Spears would be going too far into invading her privacy.

We should consider the analogy to snail mail. In an article called "Conceptualizing Privacy," Daniel J. Solove has observed,

> "We want certain matters to be private, even if we need to create this privacy through the use of law. Privacy is an issue of power; it is not simply the general expectation of society, but the product of a vision of the larger social structure. For example, in America, the privacy of letters was formed in significant part by a legal architecture that protected the confidentiality of letters from other people and government officials. In colonial America, mail was often insecure; it was difficult to seal letters; and the wax often used to keep letters sealed was not very effective. There was widespread suspicion of postal clerks reading letters; and a number of prominent individuals, such as Thomas Jefferson, Alexander Hamilton, and George Washington, decried the lack of privacy in their letters and would sometimes even write in code. As Ralph Waldo Emerson presumed, it was unlikely that 'a bit of paper, containing our most secret thoughts, and protected only by a seal, should travel safely from one end of the world to the other, without anyone whose hands it had passed through having meddled with it. Despite these realities, and people's expectation that letters would not be confidential, the law evolved to provide strong protection of the privacy of letters. " (90 *Cal. L.* Rev. 1087, 1142-43 (2002))

Privacy is not a matter of inertia and drift; it is the product of formal legal and informal social regulation. This regulation imposes sanctions of different kinds for breaches and disruptions of normative, widely shared understandings of what privacy people should have in what circumstances.

The track record of the internet industry with regards to end user privacy is abysmal. We should seriously discuss at what point government regulations and economic disincentives could help make it better.

6. Conclusions

When sites have huge amounts of information about users they can make two enticing offers. First, they can provide somewhat better services to the end user, and, second, they can generate better advertising revenue. The combination of these two does not bode well for privacy on the internet.

The single largest issues around privacy that the IETF or W3C could significantly impact all revolve around what the browser vendors will be incented to implement. Of the four major browser codes bases, three are heavily funded by operators of large search engines, and the other is funded by a company moving strongly into advertising on mobile devices. The motivations for these player are very homogenous, and the lack of diversity in choice for end users will not be good for privacy.

Ultimately, privacy needs to be protected so that the internet can continue to improve lives, not simply by making stuff more readily available for purchase but by ensuring that people can communicate with other people – for all of the many reasons that humans have for communicating.

7. Acknowledgments:

Thanks to Eric Rescorla and Lyndsay Campbell for helpful advice and discussion on this topic.