The disruptive effects of user privacy

Chris Mitchell

24th March 2015

1 Introduction

We are all accustomed to the idea that what we do online is not very private. We may not know exactly who knows what, but we do know from personal experience that organisations connected to the Internet, e.g. websites we visit and Internet advertising agencies, monitor our activity and use it to target advertising. In this context 'activity' includes not only where we browse on the web, but our past purchases, the contents of our emails, and other factors we may not even be aware of.

The means by which we are tracked is not so clear, at least to most internet users. A minority of us understand how cookies can be used to track repeated visits to the same website, and also, through the HTTP referrer field and links embedded in web pages, how advertisers can track us. A smaller minority understand that, even if cookies are disabled, so-called browser fingerprinting techniques enable web servers to uniquely identify platforms. Of course, disabling cookies is rather fraught, since it also disables many of the most useful features of the Web.

Web browser fingerprinting techniques have existed for several years, but their use and effectiveness has only relatively recently become widely publicised, e.g. as described in an article published in December 2014 by Sophos[1]. Indeed, such is the sophistication and maturity of the technique that many companies today offer browser fingerprinting products and services. Essentially browser fingerprinting involves a website using its interactions with a user platform (computer, tablet or smartphone), including through use of JavaScript programmes downloaded to a browser, to learn a host of details about the software and hardware of that platform. This might include what operating system and browser are in use and which versions, what the capabilities of the platform are, e.g. in terms of screen resolution, and what fonts are available. This information is sufficiently detailed to uniquely identify most platforms in use. Of course, IP addresses help with fingerprinting, but the use of anonymising routers doesn't stop fingerprinting.

Whilst our activities can be readily tracked using a variety of means, there is also great pressure to change this, including from legislators, such as the European Commission, who wish to protect citizen's privacy; pressure groups of many types, arguing in favour of greater end user privacy; and standards and other guidelines, which set down codes of behaviour and best practices for websites. Supporting these demands for greater privacy are a range of technologies that help support privacy, e.g. including: encryption; good practice schemes such as the 'do not track' HTTP header field; anonymising routers; anonymous credential systems and other special cryptographic schemes; and homomorphic encryption, which potentially enables processing of encrypted data.

However, despite the plethora of technological aids, in practice we tend to largely rely on regulatory/legal compliance solutions to protect our privacy. Such an approach inherently assumes that those with access to our personal data will behave in accordance with law and regulation. Of course, this is a questionable assumption; moreover the level of legal protection for our privacy varies widely as we travel around the globe. As a result, some in academia and elsewhere advocate a purely technological solution, arguing that use of appropriate technology could prevent any misuse of personal data, however it might arise. However, the consequences of such an approach, if it

[1] https://nakedsecurity.sophos.com/2014/12/01/browser-fingerprints-the-invisible-cookies-you-cant-delete/

could ever be realised (which is, of course, a big if), are profound, and this is the main focus of this article.

2 Possible privacy goals

Perhaps the ultimate goal of privacy advocates is to enable us all to leave no identifiable trace of our activities, if that is what we want. Some would suggest that such an arrangement should even be the default, given that many users have limited technical expertise. However, defining what *no trace* means is problematic.

To see why even defining the type of privacy we might want is difficult, it is necessary to observe that almost everything we do partly identifies us, e.g. we indicate our language, interests, etc. by where we choose to browse. In addition, some activities automatically reveal our unique identity, e.g. when we use a credit card for payment. Perhaps the key property we might wish to achieve is linkability of activities, or rather unlinkability. That is, we might reasonably wish for two distinct interactions with the same or different websites to be incapable of being linked by these websites.

These difficulties in definition highlight the difficulties in effectively anonymising personal data. Such anonymisation has clear benefits, allowing large data sets to be analysed, e.g. to identify new treatments for illness, new solutions to complex problems, etc. However, the risk of de-anonymisation (or re-identification) is always present, so anonymisation needs to done with great care. However, this is a subject for a different article.

3 Disruptive effects

The supposition for the next part of this article is that the privacy advocates are completely successful, and by default all our activities are unlinkable (except where necessary). That is, suppose we can all use the Internet knowing that, unless we choose to reveal who we are, it is technologically impossible to link our various interactions with third parties. 'Hurrah!' we might all say, except that the potential impacts are far-reaching. Most obviously, the service providers would lose their ability to link one user interaction with another, severely limiting their ability to target advertising. Perhaps less obviously it would also affect both security (of users and service providers) and usability in a variety of ways. We next look at these impacts in a little more detail. Perhaps I should observe at this point that the observations made below are not new – many authors have been saying similar things for some time, but it perhaps helps us all to be reminded of the implications of pursuing higher levels of privacy.

3.1 No more free stuff?

Many of the free web services we use on a daily basis are funded through advertising. Examples of such services include search, cloud storage, social networks, messaging (email and instant), Internet gaming, and voice over IP. That is, it includes things that most of us rely on all the time in our daily lives, both at work and outside. It seems evident that loss of targeted advertising could severely impact advertising revenues for these service providers, as well as other possible revenues. With a potentially much reduced revenue stream there will presumably be less free services for us all to enjoy.

How might this affect us? Well, perhaps we will have to start paying for all these services. Alternatively, maybe some service providers and/or services will simply vanish, if it becomes uneconomic to provide them. In any event, we should expect a huge disruption in the economics of the Internet. Some would say this is a small price to pay for greater privacy, but others may disagree.

3.2 Less effective security

Security and privacy often push in somewhat different directions. We next highlight a few ways in which more effective, technology-driven, privacy provision could affect the provision of security, and might ultimately damage some end users.

Some Network Intrusion Detection Systems (NIDSs) examine DNS messages[2]. As a result, if DNSsec encryption was to be widely deployed, which would, of course, enhance security as well as privacy, then such messages become opaque to the NIDS. That is, by concealing traffic, detecting intrusions becomes more difficult. It has also been widely suggested that DNSsec could make distributed denial of service attacks much more effective, although the degree to which this is true has been disputed[3].

Browser fingerprinting has both positive and negative aspects. Clearly it negatively impacts user privacy. However, it is also widely used as a means of enhancing user authentication[4], by verifying that a user is working via a known platform. That is, if browser fingerprinting was made impossible (actually, very difficult to achieve for anyone other than an expert user) then user authentication would be made less effective.

As is well known, effective user anonymity makes ensuring that users are held accountable for their actions very difficult, if not impossible. That is, efforts to investigate security breaches may be made very much more difficult if all the activity records are unlinkable. More generally, criminal investigations could be made much more difficult. Legal interception may also be made much less valuable to investigators.

3.3 Less effective everything

We finally briefly observe how privacy might impact usability. Browser fingerprinting techniques are used by many websites to understand the capabilities of user platforms, thereby providing content tailored to that platform. For example, content sent to a smart phone can be tailored to display effectively on a small screen, as opposed to content sent to a desktop PC. Indeed, it is hard to see how some details of the end user platform can be withheld from content providers without seriously affecting usability.

Similarly, one of the features of cookies that most of us rely on is the e-commerce 'shopping basket'. We can get part way through our supermarket shopping on line and the contents of our half constructed order will survive even if the platform is rebooted. I know many of us regard this as an essential feature, without which e-commerce would be much less useful. However, this reliance on cookies automatically seriously reduces the degree to which we can stay private.

4 Concluding remarks

I should say at this point that, lest I be marked down on somebody's hit list as an enemy of all that is good and just, this short article is not intended as an argument against enhancing user privacy – it is just intended to point out some of the implications. In fact, implementing complete unlinkability is theoretically possible but very difficult to achieve in the real world. For example our browsers leak vast quantities of information about us; however, the technologies required to fix this are not simple to use. For example, few of us even know what anonymising routers are or what the threat is that

[2] http://en.wikipedia.org/wiki/DNS_analytics

[3] http://conferences2.sigcomm.org/imc/2014/papers/p449.pdf

[4] https://docs.secureauth.com/pages/viewpage.action?pageId=15860199

they address, let alone use them, and it is not very practical to expect users to start with a clean OS installation every time they browse the web.

One possible outcome of any significant reduction in the ability of web servers to track users is that we will be given more overt choices between a reduction in our level of privacy and making a payment for service. Whilst users say they value their privacy, in practice they appear to be reluctant to spend money to do so. For example, AT&T in the US allows gigabit service subscribers to opt out of deep packet inspection for a $29 monthly fee. Apparently most users do not pay the extra fee, although whether this is because they do not value privacy or they don't trust AT&T not to track them even if they do pay the extra is unclear! Such a pricing versus privacy approach is not universally popular, as, for example, an opinion piece in The Guardian on 21st February 2015[5] makes clear.

Another direction much discussed is giving the ability for users to choose, perhaps from one instant to the next, to what degree their privacy is impinged. That is, for some sites users might be happy to take a more relaxed view of tracking in return for free service, whereas for others they may wish to have their interactions remain completely unlinkable. Whilst this is possible in principle today, e.g. by running 'clean' browser instances in separate 'vanilla' virtual machines, for most of this is not a practical option. Of course, this is precisely the sort of thing that HTTP 'do not track' is meant to allow, but it would appear that most websites ignore such a request[6].

A further issue is that if law enforcement and other government agencies cannot access data via interception, then they are likely to try other methods. These other methods may be even more intrusive. For example, there has been much recent discussion of malware distributed by western governments for user monitoring, including the recent damaging revelations about apparently government-originated malware infecting hard drive firmware[7]. Whilst this particular malware appears to have been used only to target high-value individuals, use of such an approach could easily be extended to allow for mass surveillance. If course, it could reasonably be argued that such approaches are likely to be used by governments anyway, even if we don't deploy better privacy technologies.

Getting the balance right is clearly very difficult, so what can we reasonably expect to see happen? Even the most strident advocates of technological privacy solutions do not suggest the legal/regulatory/compliance approach should be abandoned, and this approach will surely continue, as will the development of best practice guidelines/standards. Privacy technology will also continue to advance, and some of it will no doubt be deployed. However, I fully expect security agencies and others to continue to develop ways round deployed technologies.

Of course, highly skilled and highly determined individuals can, as now, make their activities pretty private, but they are essentially irrelevant to the argument. So probably not very much will change, and the possible disruptions discussed above won't happen, unless, of course, legislators demand it. Nevertheless, the potential for huge disruption remains. Perhaps the best way of summing this up is to say that, in the words of the old adage, we should be careful what we wish for.

[5] http://www.theguardian.com/commentisfree/2015/feb/20/att-price-on-privacy

[6] http://www.howtogeek.com/126705/why-enabling-do-not-track-doesnt-stop-you-from-being-tracked/

[7] http://www.bit-tech.net/news/bits/2015/02/17/equation-group/1