Group - User Tracking on the Internet:
Sebastian Schenk, Florian Häntzschel, Karl Adler

Mrs. B.Arch. B.Sc. Darlene Kilian
Hochschule für Technik und Wirtschaft Dresden
Fakultät Wirtschaftswissenschaften
Friedrich-List-Platz 1
01069 Dresden

Dear Mrs. Kilian:

In keeping with our agreement, we are submitting the enclosed report entitled *User Tracking on the Internet.*

As we agreed, the purpose of this report is to provide an overview of the techniques used to track individual internet users and measures the user can take to prevent being tracked. Whilst providing technical background information the report, albeit detailed, is designed for an audience without any computer-science background. For the audience's benefit some information on some subjects may therefore be merely mentioned but not discussed extensively.

We hope this report meets with your expectations.

Kind regards,

Sebastian Schenk
Florian Häntzschel
Karl Adler

Encl.: Project report

Report
on
User Tracking on the Internet

submitted to

Mrs. B.Arch. B.Sc. Darlene Kilian
Hochschule für Technik und Wirtschaft Dresden
Fakultät Wirtschaftswissenschaften
Friedrich-List-Platz 1
01069 Dresden

June 18, 2012

by
Sebastian Schenk

Florian Häntzschel

Karl Adler

This report examines techniques used to track individual internet users. It concludes with the measures the user can take to prevent being tracked.

Table Of Contents

List Of Figures

Abstract

Internet user tracking is a technology that is used extensively for a variety of purposes, ranging from vital information for website designers to commercial personalised advertisement applications. The technology, however, is prone to malicious misuse and may also serve as a means to identify individual internet users, tracking their internet behaviour and collecting sensitive information such as name address and bank details.

The technology used to link a certain identifiable individual to an otherwise anonymous internet IP address is advanced and at times quite complex. The very basics of key logging of IP addresses on a central, website comprehensive, server serve as a means for complex tracking networks that collect and store data from all over the internet. Utilising the cookie technology, where a small piece of software is installed within the web browser, the networks are able to even track visits of websites that are not part of their network. Linking these data together customers using such tracking networks data are able to create unique user profiles with the ultimate intrusion of privacy being the collection of personal sensitive information. Such information may then be used for spam mail and other malicious activities.

Examples of these tools are introduced in form of ‘Conclusion’, a tool to discover active trackers and tracking networks, and ‘DoNotTrack+’ which even automatically blocks those networks after discovery. Serious software providers ensure further security by offering so called md5 sums, which are check sums of binary files. Tools such as ‘MD5 Calculator’ create these md5 sums of downloaded software, allowing for comparison between the generated one and the one stated on the provider’s website.

1. Introduction

As internet and computers become more prevalent in both personal and professional life new problems arise. It is not science fiction but rather a billion dollar business – the transparent user. Out of concerns for privacy, and unwillingness of being spied upon, and made money with, it is important to raise awareness of this issue.

User Tracking, also referred to as Website Visitor Tracking (WVT), in essence is online marketers spying on the users' browser, detecting the user's visited websites, the keywords entered in a search engine, the time spent on specific websites or even what, and how many, products have been purchased. Different software products have been specifically developed by the marketers to acquire this information. As a consequence, internet advertisement is getting customised and optimised whilst browsing and tailored to the users' online profile. The technology is prone to malicious misuse and may also serve as a means to identify individual internet users, tracking their internet behaviour and collecting sensitive information such as name, address, and bank details. Several tools available help the user in keeping this information private.

2. Tracking Techniques:

2.1 General proceeding

To understand the whole proceeding of tracking and identifying users on the World Wide Web it is necessary to know how hyper text transfer protocol (http) and the user's browser works which is beyond the scope of this proposal, so just the basics are outlined.

As a tracker, first of all it is necessary to identify the user. This requires making a user unique in more than one billion internet users. There are different techniques available that use browser version and some other indicators. The problem using this technique is that the IP changes at least every 24 hours and the user might switch between different browsers and devices. To solve this problem a unique key created of mentioned indicators and some random data is stored at the users' device in a small file called a cookie.

In order to recognize a unique user after an IP switch occurred, these data keys are also stored on the web server.

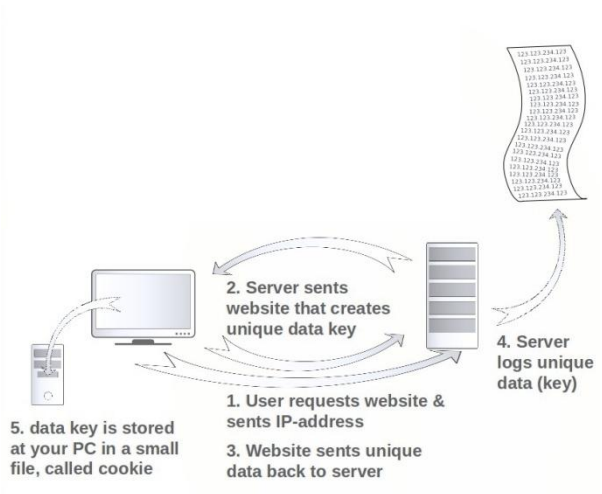

Figure 1: User identifying

Connecting these data to websites, like Facebook, where users are forced to be registered, providers of websites are able to compare the unique key data and track user’s behavior on

different devices. Extending this technique a bit, instead of storing the data only on one server, many websites send their data to a central server, where the keys and the browsing data are stored in a huge database.

Matching these data to known accounts it is easy to collate different devices and track the entire browsing history, worst case down to real name. Such architecture is called a tracking network. Facebook built such a network unobtrusively using its controversial like button, which is basically a tracking page embedded on a million web sites1.

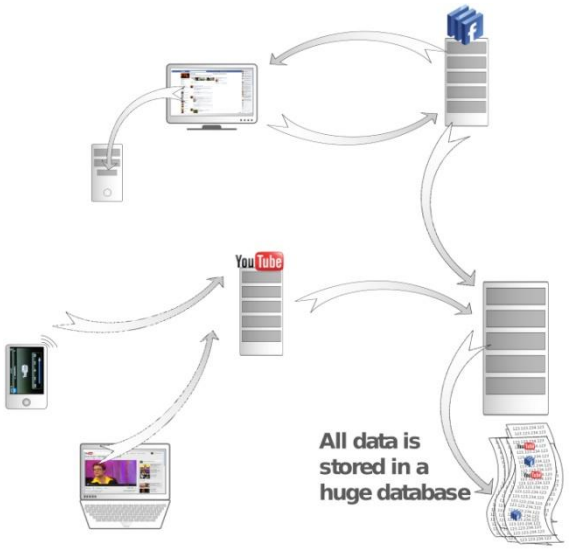

Figure 2: User tracking

2.2 Tracking Networks

Of course not every website belongs to a tracking network, but a huge number of them do, especially the top websites. According to the Alexa ranking most of them use or even provide tracking networks[2]; namely Google, Facebook, Amazon and many more.

[1] "Facebook 'like' button declared illegal," last modified August 19, 2011, http://www.thelocal.de/sci-tech/20110819-37073.html
[2] "The top 500 sites on the web," last accessed June 16, 2012, http://www.alexa.com/topsites.

There are different kinds of tracking networks; they distinguish themselves in purpose, provider and political circumstances like privacy laws. Whilst Amazon's purpose is just to sell their own products, Google's target is to place tailored ads and Facebook provides more than just demographic data. There are also some less known tracking networks for scientific or benchmarking purposes like International Federation of Audit of Circulations[3]. An entire list of known tracking networks can be found at PrivacyChoice.org which is provided by an organization that is trying to make online tracking more transparent[4]. They also offer a browser extension to opt-out for over 150 tracking companies. In spite of that it is necessary to know that there are some really unknown tracking methods and networks.

2.3 More than just a cookie

In addition to the mentioned techniques there are a lot of other methods to easily gain access to data of careless users. Besides hidden malware that comes with infected software downloads or files there is a lot of software, known to be data collecting. E.g. Google Chrome, known as 'the most curious browser' is used by more than 26% of internet users[5]. Whilst more than 70% of Americans reject tailored ads based on tracking activities most of them do not know that a lot of browser extensions like tool-bars are able to log their entire internet history[6].

So being vigilant during browsing is most important, but unfortunately not sufficient. To delete tracks in the net it is not enough to delete the ordinary cookies. There are some other techniques to store data on your device; like flash cookies. Flash is the software that is necessary for most multimedia contents. While Adobe, manufacturer of the flash software,

[3] "International Federation of Audit Bureaux of Circulations," last accessed June 11, 2012, http://www.ifabc.org.
[4] "Tracker List," last accessed June 16, 2012, http://www.privacychoice.org/companies/all.
[5] "W3Counter – Global Web Stats," last modified May 31, 2012, http://www.w3counter.com/globalstats.php.
[6] "Americans Reject Tailored Advertising and Three Activities that Enable It," last modified September 29, 2009, http://ssrn.com/abstract=1478214.

offers ‘tool7’ to remove flash cookies, during flash install there is no information about the storage and usage of flash cookies.

Another possibility to get invisible on the web is offered by newer browsers: When loading websites, the browser sends a message that tracking is not desired. This will become somewhat of a standard, designed by the W3C7 (World Wide Web consortium), but it also needs to be accepted by the online advertisement business.

In conclusion there is no absolute safety to avoid trackers until there will be a worldwide privacy law, but there are a lot of configurations and tools to make it difficult for the tracking companies.

3. Prevent tracking

The Internet was built a platform for sharing data and exchanging information, but not for security. Nowadays chances are malicious people are tracking your internet behavior in the form of scripts, cookies, spyware and even through your camera. In order to prevent people from tracking you it is important to know the methods and processes of removing and eliminating such malicious software. There are three regulations to comply with when you want to reduce the risks of getting tracked.

3.1 Visualizing Trackers:

There are many tools available on the internet that are able to visualize trackers for you. One tool mentioned here is called ‘Collusion’. It is a browser add-on that displays all third parties that are tracking your

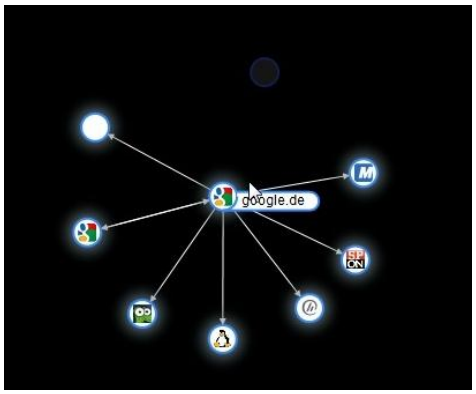

Figure 3: Tracking Visualization Software ‚Collusion‘

[7] “Tracking Compliance and Scope,” last modified November 14, 2011, http://www.w3.org/TR/2011/WD-tracking-compliance-20111114l.

movements across the web in real time. A spider web displays all connections that trackers have with each other and the interaction between the companies and websites (see fig. 1). There are also other browser plug-ins such as Do-Not-Track Plus or TrackerViz that have similar features and functionalities.

3.2 Avoiding Malware:

Malware, short for malicious software, is software to help hackers disrupt users' computer operation, gather sensitive information, or gain unauthorized access to a computer system. Therefore avoiding malware is another important aspect when talking about internet security. In order to do that some basic rules have to be observed:

- ***a.*** *Updating the operating system:* Keeping your operating system up-to-date means you allow your operating system provider to release security updates for your computer. That minimizes security breach exploitation and closes security gaps.
- ***b.*** *Updating software:* Besides enhancing user experience, software updates are made to close critical security gaps as well as correct bugs. By closing such security gaps you decrease chances for trackers to attack your computer.
- ***c.*** *Keeping the antivirus software up-to-date:* Having an updated signature database for the antivirus software is an important aspect to ensure finding vicious viruses, spyware and other tracking tools. Furthermore, it is necessary to have an antivirus check on a regular basis and to have it run in the background.
- ***d.*** *Using one unique antivirus software:* Using one unique antivirus software is essential as more software of this kind prevents each other from working correctly.
- ***e.*** *Using official download webpages:* By using official websites for downloading software you avoid getting spyware. Never use links from any random non-official site.

f. *Checking binary signatures for downloaded software:* One well-known binary signature also referred to as checksum or hashcode is the so called 'md5-checksum'. This is a specific and unique signature of a file that can be produced by the user itself. This signature is given on the official website or in a trusted database. After comparing both codes users can make sure they downloaded the correct software and not a fake installer containing a virus. Figure 2 shows an example of an md5-hashcode calculating tool.

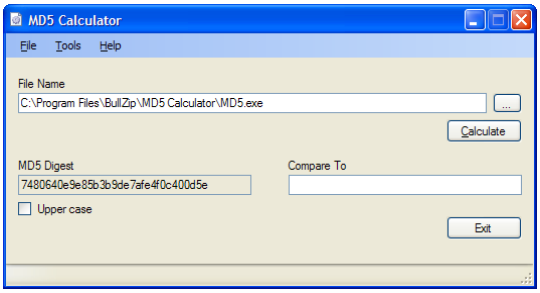

Figure 4: MD5 Calculator

g. *Using a firewall:* A firewall is a software application or a hardware device which controls and restricts the incoming and outgoing data traffic when surfing the internet. Therefore, a correct working firewall is an essential element for safe internet access. Many operating systems have a default firewall, but those firewalls only fulfill the basic safety requirements, thus it is recommended to make use of professional software products like "Kaspersky Internet Security" or hardware solutions such as "Cisco Firewall Router".

h. *Blocking idle ports:* A port is a type of electronic, software- or programming-related docking point through which data flows from the world wide web to your computer and vice versa. Blocking idle ports means they cannot be misused from trackers or hackers as a gate for dangerous data exchange.

i. *Avoiding administrator accounts:* Using the computer as an administrator with full rights entails that trackers and hackers have the same access to the users' private data in case of a computer takeover. Spyware must be much smarter to be used at full potential when surfing the net via a standard user account.

j. *Using uncommercial operating systems:* Generally open source operating systems (e.g. Linux) are safer than commercial systems such as Microsoft Windows as there was only a dozen of known malware programs running under Linux.

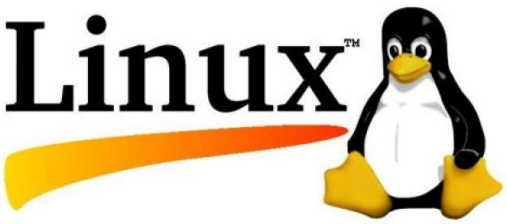

Figure 5: Open source operating system ‚Linux‘

3.3 Avoid getting tracked

Whilst all these points decrease chances of getting malware the following considerations will directly avoid other people from tracking your internet behavior.

a. *Ensuring hardware unavailability for other people:* By making sure your wired network is not accessible to others, no one is able to manipulate the router.

b. *Using a high level wireless network encryption standard:* Informing oneself about the latest encryption standards and using them as a standard transmission service you reduce the possibilities of people tracking you online. E.g. it is highly recommended to use the 2004 WPA2 encryption standard for the router setup.

c. *Avoid using an alien proxy server for the internet connection:* A proxy is an intermediate server that controls ingoing and outgoing data traffic to the internet. It can be accompanied with a higher privacy level but in case you don’t know the proxy server you are forced to trust the unknown random stranger who set it up.

d. *Using encryption for other IT*-aspects: It is important to use encryption for other IT services like mail services. Encryption standard for mail service are for example POPS (Post Office Protocol - Secure) or IMAPS (Internet Message Access Protocol - Secure).

e. *Avoid using IP hidder services:* This implies the usage of proxy servers and therefore entails the same risks as point d.

4. Conclusion

During our extensive research we found that surfing the internet can be extremely dangerous concerning privacy issues. Tools like ‘conclusion’ give evidence of how extensively users are being tracked on the internet. Even in the short demonstration during the presentation the number of trackers became obvious with about 50 tracking us on our five minute visit on a site cnn.com. Tools like this may reduce the risks of being tracked but whilst all these efforts increase internet security it is still important to be vigilant.