An analysis of web-based user tracking and online privacy

Date: April 16, 2017

Contents

1. Introduction

Currently there are over 3 billion users on the internet with an internet penetration level of almost half of the world's population (Kende, 2016). People have many uses for the internet including shopping, banking, researching, and maintaining social connections, but with increasing amounts of consumer targeting and surveillance one of the main challenges to users is maintaining their privacy (Alsabah & Goldberg, 2016). The 2014 CIGI-Ipsos Global Survey on Internet Security and Trust found that users have become more aware of data breaches, pervasive surveillance and third-party trading of personal data, and their level of trust of the internet has dropped[1]. In other words, internet users are increasingly becoming more concerned about privacy issues, and for companies and organizations that provide services through the internet, ignorance of privacy issues can have "undesired consequences" such as law suits and reputation damage (Ziegeldorf, Morchon & Wehrle, 2013).

In 1968, Alan Westin, described information privacy as "the right to select what personal information about me is known to what people (Westin, 1968)." There are conflicting consequences to privacy – privacy for those who use the internet for legal purposes also means privacy for users with illegal purposes. There are aspects of privacy related to free speech when "the reality is that freely voicing one's opinions is punishable by law in one country and not in another (Haughly, Epiphaniou & Al-Khateeb, 2016)." In (Schneier, 2015), privacy is called "an inherent human right, and a requirement for maintaining the human condition with dignity and respect." He states; "privacy is innate: mammals in particular don't respond well to surveillance."

Privacy includes; awareness of privacy risks, individual control over the collection and processing of personal information, and awareness and control of subsequent use and dissemination of personal information (Ziegeldorf, et al., 2013). Users are concerned about unauthorized use of personal information, their online habits, their locations, and their physical conditions and how the collection and re-selling of this information could affect their ability to get credit, insurance and employment (FTC, 2015). They also expressed concerns that these risks could "undermine the consumer confidence necessary for the technologies to meet their full potential, and may result in less widespread adoption."

There is evidence that user's information is bought and sold, including their internet usage, social networking and other consumer data (Roderick, 2014) and the information industry treats personal data as goods. Users can attempt to use the internet anonymously but anonymous users are "not lucrative and personal data have a unit price (Peacock, 2014)." Big data and analytics are growing industries and organizations have many uses for mining data. Retailers use data to direct advertising to consumers, politicians use data to target voters and social media data is mined for use by employers and for developing social networks (Schneier, 2015).

[1] https://www.cigionline.org/internet-survey

In the 1960s, Arpanet was launched to serve as a secure communication channel for scientists and laid the foundation for today's internet[2]. Arpanet opened secure communications and increased the performance of connectivity and information exchanges (Peacock, 2014). In the 1980s, domain names for internet servers evolved allowing the mapping of servers to internet protocol (IP) addresses (Longman, 1998). In 1990, Tim Berners-Lee released a prototype web browser based on hypertext markup language (HTML) which was further enhanced by Dave Raggett from Hewlett-Packard in Bristol, England.

In 1994, Lou Montulli developed code for use by HTML that was eventually termed a cookie. Prior to this invention, data exchanges between a user's computer and an internet server were anonymous (Peacock, 2014). This invention allowed for "storing information on the user's computer about a transaction between a user and a server that can be retrieved at a later date by the server (Peacock, 2014)." The cookie is the most common way to track browser activity (Eckersley, 2010).

Since the invention of the cookie, there have been a multitude of new methods for collecting information about internet user behaviors. These include supercookies, evercookies, flash cookies, browser history extraction, traffic analyses, packet inspection and browser fingerprinting (Boda, Foldes, Gulyas, & Imre (2011); Mowery & Shacham, (2012); Olejnik, Castelluccia, & Jane (2014); Mulazzani et al. (2013); Macia-Fernandez, Wang, Rodriguez-Gomez, & Kuzmanovic (2012); Ghaleb (2016); Soltani, Canty, Mayo, Thomas, & Hoofnagle (2009). The use of online tracking methods is "growing at a startling pace" and projected to keep increasing (Mathews-Hunt, 2016).

To combat online tracking, anonymity networks appeared such as The Onion Routing (Tor), the Invisible Internet Project (i2p) and Freenet (Zhioua, Alsabah & Goldberg, 2016; Haughey et al., 2016). These networks help conceal user identities by "providing unlinkability between a user's IP address, his or her digital fingerprint, and his or her online activities (Alsabah & Goldberg, 2016)." Blocking tools for third-party tracking exist and are currently the primary solution (Acar, Eubank, Englehardt, Juarez, Narayannan, & Diaz, 2014). There are also newer tools that inform the user if their information is being collected or shared, such as browser add-ons (Aonghusa & Leith, 2016).

The upcoming internet of things (IoT) may present additional challenges to user privacy (Ziegeldorf, et al., 2013). The Federal Trade Commission (FTC, 2015) estimates there are already 25 billion connected devices and expects that to double by 2020. Examples of IoT are smart homes, connected cars and fitness devices. There is concern that the collected data could "increase the harm caused by a data breach" as these devices collection additional information on location, health, driving and shopping (Kende, 2016).

[2] https://www.darpa.mil/

2. Privacy

A simple definition of information privacy is "the right to select what personal information about me is known to what people (Westin, 1968)." Other definitions include "the right to be left alone" or "the right to prevent the disclosure of personal information (Malandrino & Scarano, 2013)." In Martin (2016), the author defines a privacy violation as occurring when "information is either not controlled or no longer inaccessible."

When another individual or organization has access to large amounts of our information, they have power over us (Angwin, 2014). With the development of cheaper, faster data storage technologies, data records can be kept longer and perhaps permanently. Schneier (2015) argues that "having everything recorded and permanently available will change us both individually and as a society." He notes that exposure is always a risk when a computer stores data, including the possibility that new policies put into place could be retroactively applied to stored data from past years.

Privacy risks include data protection, surveillance, privacy-intrusive commercial solicitations, security risks and exposure to unfair commercial practices (King & Forder, 2106). The FTC describes privacy risks as enabling unauthorized access to personal information, enabling attacks on other systems and risks to personal safety[3]. There are threats to privacy from both large databases that contain data from multiple sources (Big Data) and threats that may target an individual within a database[4]. Behavioral targeting is called "the greatest threat to privacy, since it is the result of tracking and profiling users' browsing habits throughout the Internet, often without their knowledge and consent (Parra-arnau, 2017).

A legitimate debate develops when criminal intentions are driving the desire to remain anonymous. A debate ensues over the balance between privacy and security and "the two are sometimes considered to be at odds with each other (Angwin, 2014)." But, Schneier (2015) points out that the government or Google, for instance, having all our data overemphasizes "group security at the expense of individual security." He argues that privacy can protect criminal activity but that privacy is "fundamental to liberty." Mowery & Shacham (2012) identify "constructive" benefits of fingerprinting in the case of bank authentication. They call fingerprinting "destructive" when the user does not want to be tracked. Haughey et al. (2016) argue that there are ethical issues and arguments that can be made for and against, "due in part to the fact that privacy and anonymity are desired by criminals and terrorists, not just individuals who care about their privacy."

In the US in 2009, the government became increasingly concerned about privacy on the internet and the growing practice of behavioral advertising[5]. As a response, the Federal Trade Commission (FTC) issued

---

[3] https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[4] https://33bits.org/

[5] www.knowprivacy.org

guidance regarding the notification to consumers regarding the collection of data for behavioral advertising and options for opting out of that collection. Privacy policies have been written to allow users to "make informed decisions about which sites they used based on the site's data collection practices (Gomez et al., 2009)." However, privacy policies can be lengthy, vague and difficult to read and may lead consumers to believe their privacy is protected. Privacy notices can be time consuming and misleading (Martin, 2016). In Gomez et al. (2009), the authors found that users are concerned about privacy and do not want their data collected without permission and it has been found that 63% of users were concerned about third-party data monitoring (Wills & Zeljkovic, 2011).

Users may unknowingly release their information for a better user experience (Mathews-Hunt, 2016) but "once released to the public, data cannot be taken back[6]." Users information can be collected, stored, mined, shared or sold without permission or consent (Malandrino & Scarano, 2013). Unknowingly users can "lose control over which details about their private lives are known, and they have little control over who gets to learn of these details after the data passes into a profiler's hands (Berger, 2010)." With more smart phones, smart cars and smart homes, privacy on the Internet of Things (IoT) may "aggravate privacy issues and introduce unforeseen threats that pose challenging technical problems (Ziegeldorf et al., 2013)."

There appears to be a level of acceptability regarding privacy where consumers appreciate some personalized content but do not appreciate personalized content when dealing with sensitive topics such as health concerns, disabilities, bankruptcy and divorce, gambling, sexual preferences, location and unemployment (Anonghusa & Leith, 2016). In one study, it was found that users were more concerned about suggestive or embarrassing content than tracking (Agarwal et al., 2013). Users tend to not want to share data related to sexual orientation, financial status, race, diseases and pregnancy (King & Forder, 2016).

Other practices deemed unacceptable in data collection from the internet include differential pricing, targeting the vulnerable with scamming, and cases of database error (Mathews-Hunt, 2016, Aonghusa & Leith, 2016). The ability to link data collected with personally identifiable information (PII) is a major concern, particularly due to the rise of identity theft, social engineering attacks, stalking and other criminal acts (Malandrino & Scarano, 2013) and companies are now able to correlate cookies with credit card transactions which ties the browser to a user, eliminating anonymity (Schneier, 2015). Other violations of privacy are price discrimination, unsolicited advertisements, social engineering or erroneous automatic decisions (Ziegeldorf et al., 2013).

For security reasons, there are positive aspects in the case of protecting consumers from criminal elements but when companies ignore privacy issues and are exposed by the user community, there can be undesirable consequences such as reputation damage and expensive lawsuits (Ziegeldorf et al., 2013). One example is the case against KISSmetrics, where the company used Etags and supercookies and was found to violate wiretap laws resulting in a $500,000 settlement[7]. An example of a large criminal component in advertising was the GlavMed case. When consumers searched for specific drugs, targeted links appeared on their

[6] https://33bits.org/

[7] http://www.mediapost.com/publications/article/191409/kissmetrics-finalizes-supercookies-settlement.html

browsers that routed users to GlavMed pharmacy sites where users could purchase controlled drugs from Canadian Pharmacies. Google was aware of the illegal shipments, settled the criminal charges and was fined $500 million (Krebs, 2014).

3. Laws and Regulations to Protect Privacy

The laws and regulations related to privacy and electronic communications vary by region but there is a growing awareness world-wide of a growing need for regulation of online data. More and more countries are establishing data privacy and protection laws. While laws may vary slightly, the Universal Declaration of Human Rights, developed in 1948 "is anchored in the constitutional law of most countries today (Ziegeldorf et al., 2013)."

Europe

The European Parliament of the European Union passed Directive 95/46/EC in 1995 for protection of individuals with regard to processing of personal data and on the free movement of such data[8]. The directive uses the principle of explicit consent which "forbids any kind of data collection without explicit permission from the subject (Ziegeldorf et al., 2013)." The directive was developed to make laws consistent in the member states of the EU. There are plans to enact the General Data Protection Regulation in 2018 which will strengthen the existing laws and address personal data that is exported outside the EU.

Russia

Russia passed laws in 2016 called the Information Protection Act and the Personal Data Act. The regulating authority is the Federal Service for Supervision of Communications, Information Technologies and Mass Media. Principles in the law involve informing a data subject, lawful and fair processing, minimization of use, retention and consolidation of databases of personal data[9].

United States

The United States Privacy Act of 1974 established fair information practices for notice, consent, individual access and control, data minimization, purposeful use, adequate security and accountability (Ziegeldorf et al., 2013). The act defines a system of records as a group of records under the control of an agency from which information is retrieved by the name of the individual or by some identifier assigned to the individual[10]. A subsequent bill called the Privacy Act of 2005 was a bill that would "require the consent of

[8] http://eur-lex.europa.eu/legal-content/

[9] https://www.iclg.co.uk/practice-areas/data-protection/data-protection-2016/russia

[10] https://www.justice.gov/opcl/e-government-act-2002

an individual prior to the sale and marketing of such individual's personally identifiable information, and for other purposes," but this bill died in congress[11].

The Federal Wiretap Act was originally enacted in 1968 and applied to telephone communications and declared that the government "shall not intercept the contents of communications." The modification of the act was called the Electronic Communications Privacy Act of 1986 and it extended government restrictions on wiretaps to include transmissions of data by computer. The act defines the sharing of stored records of online activities, but does not cover sharing TCP headers, essentially making wiretaps of communications "with anyone except a government body … legal (Macia-Fernandez et al., 2012)."

The Consumer Privacy Bill of Rights (CPBR) was introduced by President Obama in 2012. It defines personal data as "any data, including aggregations of data, which is linkable to a specific individual. For example, an identifier on a smartphone or family computer that is used to build a usage profile is personal data (King & Forder, 2016)." The bill of rights was developed with the intention of giving "Americans the ability to exercise control over what personal details companies collected from them and how the data was used," but the effort has not resulted in an enactment of new legislation[12].

Asia Pacific

In Australia, data protection is regulated by the Privacy Act which contains thirteen Australian Privacy Principles (APPs). Australian companies "can obtain customers' informed consent to pass their information to a foreign business that does not have privacy controls that align to the APPs[13]. In Singapore, the Personal Data Protection Act of 2012 requires consent, reasonable purpose and notification to the individual of the purpose. Malaysia has a similar policy called the Personal Data Protection Act which was enacted in 2013. New Zealand is modeling its approaches after the EU and Japan where there is the Act on the Protection of Personal Information. South Korea has the Personal Information Protection Act which was enacted in 2012 and applies higher standards to sensitive personal data and suggests a minimum amount of data collection. Vietnam has the Law on Information Technology that was passed in 2006 which is constructed like other countries' laws in the region.

Middle East

Some countries like Qatar, Saudi Arabia and the United Arab Emirates recognize data protection and privacy in specific circumstances but up until recently there were not specific "laws or regulators[14]." Qatar

[11] https://www.govtrack.us/congress/bills/109/s116

[12] https://www.nytimes.com/2016/02/29/technology/obamas-effort-on-consumer-privacy-falls-short-critics-say.html

[13] https://www.law360.com/articles/699125/the-latest-cross-border-privacy-rules-in-asia-pacific

[14] https://www.lw.com/thoughtLeadership/data-protection-privacy-laws-middle-east-2013

has recently passed their Data Protection law that will go into effect in 2017. It addresses consent, access and responsible handling practices and imposes penalties for breaches of certain provisions[15].

Africa

The Protection of Personal Information Act 4 was enacted in 2013 in South Africa. The law applies principles including lawful processing of personal information, minimal and necessary in collection, consent, and "in a manner that does not infringe the privacy of the data subject[16]."

4. Cookies

Hypertext markup language (HTML) was developed in the 1990s and eventually evolved to HTML4 in 1997[17]. In 2012, the W3C prepared a recommendation for HTML5 or the 5th major revision to HTML. HTML documents "consist of a tree of elements and text," where elements containing attributes are identified by a start and end tag. Attributes have a name and a corresponding value, separated by the "=" sign. Browsers receive the HTML upon accessing a website and create a local document object model (DOM) tree which is an "in-memory representation of a document."

The earliest cookies were developed in the 1990s and are referred to as hypertext transfer protocol (HTTP) cookies and are the most common tracking technology on the internet (Peacock, 2014). HTTP cookies contain unique identifiers in a piece of code in HTML sent from a web server to the user's browser[18]. The browser stores the information on the user's computer. The web server instructs the browser to send the cookie back according to a set of rules each time the user's browser submits a request to the webserver allowing the webserver to identify an individual user.

The World Wide Web Consortium describes the earliest cookies:

A "cookie" is a mechanism developed by the Netscape Corporation to make up for the stateless nature of the HTTP protocol. Normally, each time a browser requests the URL of a page from a Web server the request is treated as a completely new interaction. The fact that the request may be just the most recent in a series of requests as the user browses methodically through the site is lost. Although this makes the Web more efficient, this stateless behavior makes it difficult to create things like shopping carts that must remember the user's actions over an extended period of time.

---

[15] http://www.mondaq.com/x/561508/data+protection/Qatar+leads+the+way+with+new+standalone+data+protection+law

[16] https://www.iclg.co.uk/practice-areas/data-protection/data-protection-2016/south-africa

[17] http://www.w3.org/TR/html5/introduction.html#background

[18] https://dev.chromium.org/Home/chromium-security/client-identification-mechanisms

Cookies solve this problem. A cookie is a small piece of information, often no more than a short session identifier, that the HTTP server sends to the browser when the browser connects for the first time. Thereafter, the browser returns a copy of the cookie to the server each time it connects. Typically the server uses the cookie to remember the user and to maintain the illusion of a "session" that spans multiple pages. Because cookies are not part of the standard HTTP specification, only some browsers support them: currently Microsoft Internet Explorer 3.0 and higher, and Netscape Navigator 2.0 and higher. The server and/or its CGI scripts must also know about cookies in order to take advantage of them.[19]

A web server sends a cookie by populating the "set cookie" HTTP response header. A simple cookie might look like[20]:

Set-Cookie: <cookie-name>=<cookie-value>

There are other attributes that can be part of the cookie description including the following:

- Name – the name of the cookie
- Value – text value of the cookie
- Domain – web server the cookie will be sent to, defaults to the host portion of the current page
- Path – uniform resource locator (URL) path on the web server that must exist to send the cookie to, defaults to the path of the URL that sent the cookie
- Expires – the expiration date of the cookie, after this date the cookie is no longer sent back to the server

The Internet Engineering Task Force (IETF) RFC6265 documents the standards and additional attributes available for HTTP Cookies and the Set-Cookie header fields[21]. Cookies that are removed when a browser session is shut down (Expires and Max-Age attributes are not set) are called session cookies. Otherwise the cookie is called a persistent or permanent cookie.

Types of Cookies

A first party cookie belongs to the domain in the current address bar or the domain the browser is currently visiting. When the domain is different than the one shown in the address bar, the cookie is called a third-party cookie. Third party cookies are often used by websites as a service to their advertisers that enables the advertisers to "collect data, aggregate information and build personal profiles of Internet users in order to provide free and personalized services (Malandrino & Scarano, 2013)." In Malandrino & Scarano it is

[19] http://www.w3.org/Security/Faq/wwwsf2.html

[20] https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies

[21] https://tools.ietf.org/html/rfc6265#section-4.1

noted that between 56% and 75% of sites analyzed "directly leak sensitive and identifiable information to third-party aggregators." In Mathews-Hunt it was noted that 85% of the 100 most popular US websites "embedded third party cookies on user's browsers." Third parties can collect user browser history, which site the user is currently browsing and potentially other sensitive information such as email addresses (Englehardt & Narayanan, 2016).

In 2005, an advertising company called United Virtualities developed a technique called a persistent identification element (PIE) that utilized local shared objects (LSOs) and was not able to be deleted by commercially available software (Soltani et al., 2009). LSOs became known as flash cookies and were developed to "store client-side data within Adobe Flash[22]." In 2009, flash cookies were found to be present on over 50% of websites (Soltani et al., 2009). Flash cookies can be installed on a user's computer just by visiting a website and are implemented when the website issues an identifier that is stored in multiple locations on the user's computer. The website embeds Flash content or a Shockwave Flash (SWF) onto the computer for Flash banner advertisements or for metrics collection (Soltani et al., 2009). The user may remove the HTTP cookie but the web server restores the cookie by respawning from the flash cookie (Acar et al., 2014). Flash cookies can store up to 100 KB of data, have no expiration date by default and are stored in different locations. In Soltani et al. (2009), the authors found that most Flash cookies belonged to third-party advertising networks.

In Acar et al., (2014), the authors note that "in 2010, Samy Kamkar demonstrated the 'Evercookie,' a resilient tracking mechanism that utilizes multiple storage vectors including Flash cookies, localStorage, sessionStorage and ETags. Kamkar employed a variety of novel techniques, such as printing ID strings into a canvas image which is then force-cached and read from the cached image on subsequent visits. Instead of just respawning HTTP cookies by Flash cookies, his script would check the cleared vectors in the background and respawn from any storage that persists." The evercookie is a JavaScript application program interface (API) that can identify a client even after removing cookies and flash cookies[23].

Specifically, Kamkar notes that the evercookie uses the following available storage mechanisms:

- Standard HTTP Cookies
- HTTP Strict Transport Security (HSTS) Pinning
- Local Shared Objects (Flash Cookies)
- Silverlight Isolated Storage
- Storing cookies in red, green blue (RGB) values of auto-generated, force-cached portable network graphics (PNGs) using HTML5 <canvas> tags to read pixels back out

[22] https://dev.chromium.org/Home/chromium-security/client-identification-mechanisms

[23] https://samy.pl/evercookie/

- Storing cookies in the web history
- Storing cookies in HTTP ETags[24]
- Storing cookies in Web cache
- window.name caching[25]
- Internet Explorer userData storage[26]
- HTML5 Session Storage
- HTML5 Local Storage
- HTML5 Global Storage
- HTML5 Database Storage via SQLite
- HTML5 IndexedDB[27]
- Java JNLP PersistenceService[28]
- Java CVE-2013-0422[29] exploit (applet sandbox escaping)

Evercookies are browser-independent and are stored in folders that are not read by the user's browser. The identifier marking a browser generated by the evercookie has been called "virtually indestructible but vulnerable to "deleting local storages (Boda et al., 2011)." It has been shown that clearing cache and site data in Chrome, Internet Explorer and Firefox did not delete the values in isolated storage and that browsers installed on the same machine can share isolated storage.

Cookie Synching and Matching

Next to cookie respawning is the concept of cookie synching or cookie matching. A cookie synch allows trackers to share and trade user information. One web server can pass an identifier associated with a user (that is stored in a cookie) to another web server, allowing both web servers the ability to share and match the information about the user. Google developer documentation shows how this is done with Google.

---

[24] https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/ETag

[25] http://www.w3schools.com/jsref/prop_win_name.asp

[26] https://msdn.microsoft.com/en-us/library/ms533007

[27] https://developer.mozilla.org/en-US/docs/Web/API/IndexedDB_API

[28] http://docs.oracle.com/javase/8/docs/jre/api/javaws/jnlp/javax/jnlp/PersistenceService.html

[29] https://www.microsoft.com/security/portal/threat/encyclopedia/Entry.aspx?Name=Exploit:Java/CVE-2013-0422

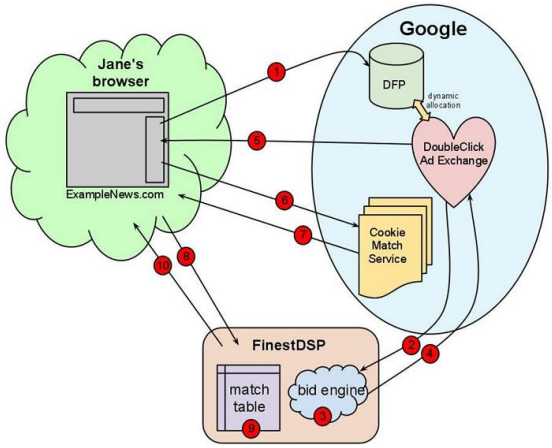

Figure 1 – Google Cookie Matching[30]

This process can only work if the web server has an ad exchange account and a real-time bidder up and running. Bidding is described in detail in the User Profiling section of this research paper.

Web Bugs

Web bugs are used to monitor who is reading a web page and are embedded in a web page HTML code (Gomez et al., 2009). They are also called beacons, clear graphics interchange format (GIFs) or pixel tags. Information collected through web bugs by the web server usually includes the IP address, the web page URL, date and time, browser type and a previously set cookie value[31].

In Gomez, et al. (2009), the top websites with web bugs were identified and the top ten sites were blogspot.com, typepad.com, google.com, blogger.com, msn.com, aol.com, yahoo.com, huffingtonpost.com, photobucket.com, and tripod.com. News sites, reference sites and job search sites were also high ranking for the quantity of bugs per site.

5. Browser Fingerprinting

Browser fingerprinting is called a passive technique, differing from cookies (active techniques) in the sense that there is no information stored on the user's computer that can be destroyed. Browser fingerprinting relies on querying parameters available in the web browser to identify the browser, or the user. The current population of web browsers in use is shown in Table 1. Desktop operating systems are primarily Microsoft Windows (almost 90%), with Mac (around 4%), Linux (around 2%) and other (around 5%).

[30] https://developers.google.com/ad-exchange/rtb/cookie-guide#how-it-works

[31] http://www.pcmag.com/encyclopedia/term/54280/

Browser	Market Share
Chrome	58%
Internet Explorer	20%
Firefox	12%
Microsoft Edge	5.5%
Safari	3.5%
Opera	1.3
Other/proprietary	.2%

Table 1. Current Market Share of Personal Computer Web Browsers[32]

Olejnik et al. (2014) conducted a large study of browser histories. He found that 69% of users have a unique browsing history but when examining users that visited at least four websites they were able to uniquely identify users by their histories 97% of the time. They also found that of the repeat visitors to a website, 38% of the users had identical browsing history.

The ability to uniquely identify users by methods has been evaluated and the results are shown in Table 2.

Method	Identification Rate
User Agent String (UAS)	58%
UAS + IP address	81%
UAS + 24 digit IP Address	79%
Browser Cookies	82%
User Logon ID	93%

[32] www.netmarketshare.com

Table 2. Unique Identification Rates by Method[33]

Identification of a browser by fingerprinting can raise security concerns because it is undetectable from the client-side. In (Mulazzani et al., 2013), an approach to browser fingerprinting was explored. The authors work was motivated by the security scanner *nmap* (Lyon, 2008), which "uses TCP/IP stack fingerprinting to determine the operating system of a remote host." They used the JavaScript engine to develop an approach that was much faster, undetectable, and could be created with a few hundred lines of Javascript. Javascript allows client-side scripting and dynamic websites and is standardized as ECMAScript[34]. The authors note that the browser is "the most prevalent attack vector for malware by far" making reliable identification of the browser a security issue, as the browser, once identified is susceptible to drive by download attacks.

User Agent String

For targeting advertisements and user identification other fingerprinting methods have developed. Panopticlick[35] was considered the first major fingerprinting experiment (Boda et al., 2011) where it was demonstrated that users could be identified based on the user agent string (UAS). The UAS is defined in RFC2616 as a sequence of product tokens and identifies the browser software as well as other system information. An example UAS from Microsoft is shown in Figure 1. The Chromium project (Google) describes the UAS as "identifying the browser version, the operating system version and some installed add-ons".

```
GET / HTTP/1.1
Accept: text/html, application/xhtml+xml, */*
Accept-Language: en-US
User-Agent: Mozilla/5.0 (Windows NT 6.3; Trident/7.0; rv:11.0) like Gecko
Accept-Encoding: gzip, deflate
Proxy-Connection: Keep-Alive
Host: microsoft.com
```

Figure 1. Example of a User Agent String[36]

[33] www.internetsociety.org

[34] http://www.ecma-international.org/publications/files/ECMA-ST/Ecma-262.pdf

[35] https://panopticlick.eff.org

[36] https://msdn.microsoft.com/en-us/library/ms537503(v=vs.85).aspx

The Panopticlick experiment sampled over 470,000 browsers who were informed and visited their website finding that almost 84% of the browsers had "an instantaneously unique fingerprint." When users had Adobe Flash or a Java Virtual Machine the uniqueness rose to over 94% (Eckersley, 2010).

The authors in that study (Eckersley, 2010) use a term called entropy to measure how distributed the uniqueness of the fingerprints were and determined an entropy of around 18, or "if we pick a browser at random, at best only one in 286,777 other browsers will share its fingerprint. Entropy is considered an abstraction of the number of different values a variable can have and a higher entropy is representative of more variance or disorder[37]. The larger the entropy, the more uniquely identifiable are the users. Panopticlick used the UAS, installed browser plugins, screen resolution, time zone and system fonts for their identification mechanism.

With more and more website development, browsers can interpret the code in different ways requiring testing by web developers so that the functionality and appearance is appropriate for the different browser types (Mulazzani et al., 2013). In order to implement features such as rendering a programmatic drawing surface, WebGL, audio and video playback in an efficient manner, browsers need to recognize the operating system and hardware (such as the CPU and GPU) of the computer or mobile device (Mowery & Shacham, 2013). Earlier browsers relied on the central processing unit (CPU) of a computer to present (render) a web page on the user's device. With improvements in the capabilities of graphics processing units (GPUs), browsers can now utilize the GPU to save power and render quicker[38].

Canvas Element

In 2012, Mowery & Shacham presented a methodology for browser fingerprinting based on font and Web Graphics Library (WebGL) rendering[39], introduced with the move to HTML5. The authors noted that "tying the browser more closely to operating system functionality and system hardware means that websites have more access to these resources, and that browser behavior varies depending on the behavior of these resources." The authors examined that ability to use the <canvas> element, pixel extraction, WebFonts[40], and WebGL to observe that a fingerprint could be extracted that is consistent, high entropy, orthogonal to other fingerprints, transparent to the user and readily obtainable.

The canvas element "provides scripts with a resolution-dependent bitmap canvas, which can be used for rendering graphs, game graphics, art or other visual images on the fly[41]" and W3C recommends not using the element when other elements are more suitable. They note that information leakage can occur with <canvas> elements.

---

[37] https://fingerprint.sba-research.org/about

[38] https://www.chromium.org/developers/design-documents/gpu-accelerated-compositing-in-chrome

[39] https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API/Tutorial/Getting_started_with_WebGL

[40] http://www.w3schools.com/css/css3_fonts.asp

[41] http://www.w3.org/TR/html5/scripting-1.html#the-canvas-element

Mowery & Shacham (2012) noted that a website could "render tests to a <canvas>, extract the resulting pixmap, then use a cryptographic hash to obtain a short, convenient fingerprint." The pixmap and resulting hash would be identical on an individual machine, but different depending on different hardware and software configurations.

Methods for Fingerprinting

The Chromium Project provides a list of known methods to perform fingerprinting. These include:

- User agent string, identifying the browser version, OS version, and some of the installed browser add-ons.
- Clock skew and drift: unless synchronized with an external time source, most systems exhibit clock drift that, over time, produces a fairly unique time offset for every machine. Such offsets can be measured with microsecond precision using JavaScript. In fact, even in the case of NTP-synchronized clocks, ppm-level skews may be possible to measure remotely.
- Fairly fine-grained information about the underlying CPU and GPU, either as exposed directly (GL_RENDERER) or as measured by executing Javascript benchmarks and testing for driver- or GPU-specific differences in WebGL rendering or the application of ICC color profiles[42] to <canvas> data.
- Screen and browser window resolutions, including parameters of secondary displays for multi-monitor users.
- The window-manager- and addon-specific "thickness" of the browser UI in various settings (e.g., window.outerHeight - window.innerHeight).
- The list and ordering of installed system fonts - enumerated directly or inferred with the help of an API such as getComputedStyle.
- The list of all installed plugins, ActiveX controls, and Browser Helper Objects, including their versions - queried or brute-forced through navigator.plugins[]. (Some add-ons also announce their existence in HTTP headers.)
- Information about installed browser extensions and other software. While the set cannot be directly enumerated, many extensions include web-accessible resources that aid in fingerprinting. In addition to this, add-ons such as popular ad blockers make detectable modifications to viewed pages, revealing information about the extension or its configuration. Using browser "sync" features may result in these characteristics being identical for a given user across multiple devices.

[42] http://www.color.org/iccprofile.xalter

- A similar but less portable approach specific to Internet Explorer allows websites to enumerate locally installed software by attempting to load DLL resources via the res:// pseudo-protocol.

- Random seeds reconstructed from the output of non-cryptosafe PRNGs (e.g. Math.random(), multipart form boundaries, etc). In some browsers, the PRNG is initialized only at startup, or reinitialized using values that are system-specific (e.g., based on system time or PID).

Browser extensions are popular and allow for customization of the browser. In Saini et al. (2016), browser extensions were studied to see if separate extensions can collude resulting in a security issue. They show that the browser extension's handling of Javascript allows different extensions to share objects. Then note that a particular browser extension can not only send messages to be used by itself, but by other extensions.

Mobile devices generally make use of lightweight browsers. Loading HTML, cascading style sheets (CSS), JavaScript and media files can be slower on mobile devices so lightweight browsers apply efficiency solutions to maintain reasonable web page rendering speed (Pokharel, Choo & Liu, 2016). To improve speed, mobile devices can maintain larger cache and avoid plugins. The browser checks the cache for existence of a previously loaded page and reloads the page from cache if it is available. W3C guidelines recommend 5 MB of storage in WebStorage per website, and lightweight browsers such as GoogleChrome, Mozilla Firefox and Opera use WebStorage which is considered internal and secure. Apps can store data in external storage making the data accessible by apps that have permission to READ_EXTERNAL_STORAGE (Pokharel, Choo & Liu, 2016).

6. User Profiling

User profiling has been defined as "the process of identifying the data about a user interest domain (Kanoje, Girase & Mukhopadhyay, 2014)." Data is usually collected by websites in order to make a website easier to use, provide customization and deliver targeted advertisements (Gomez, Pinnick & Soltani, 2009). Profiling may be used for identity verification, fraud detection and for social networking such as friend finding (King & Forder, 2016). Websites almost universally embed cookies to track internet activity (Mathews-Hunt, 2016).

The goal of profiling is to amass as much information about the user from as many sources as possible (Boda, 2011) and to find out the user's personal preferences and interests (Olejnik et al. 2014). Profiling is mostly used for "personalization in e-commerce" and for "internal optimization based on customer demographics and interests (Ziegeldorf et al., 2013). Profiling uses analysis tools to derive "information that would otherwise not be available (King & Forder, 2016)." It has been found that the personal identities can be derived from data that does not contain personally identifiable data (King & Forder, 2016).

Profiling can be classified as content based filtering, collaborative filtering and demographic filtering (Kanoje, Girase & Mukhopadhyay, 2014). Content based filtering compares the content of the items with a user profile and presents items that best match the profile. Collaborative filtering organizes users into groups having similar interests and presents items that appeal to that type of user. Demographic filtering groups users based on location, age gender and education.

Methods for profiling using Big Data include data mining, machine learning, social network analysis, predictive analytics, natural language processing and visualization (King & Forder, 2016). Fuzzy clustering methods are also used by assessing web history logs. Personal data may be inferred or derived.

Online Behavioral Advertising

The process of targeting advertisements to web users is called online behavioral advertising (OBA) and means "the collection and use of data on web browsing activity of an internet-enabled device, which allows the device to be added to one or more pre-defined interest categories, to serve advertising based on those … categories (Mathews-Hunt, 2016)." The author notes that websites are increasing their tracking for commercial purposes and that social media derives most of its income from personal disclosures and behavioral attributes. The online advertising revenue in the first half of 2015 was $27.5 billion and is a highly competitive industry (Parra-Arnau, 2017).

Mathews-Hunt (2016) describes implementation methods for enabling OBA, including cookies, web bugs, IP address monitoring, click stream data analysis, deep packet inspection, history sniffing, fingerprinting and location analyses. The author calls flash cookies "one of the most invasive tracking instruments of the digital age."

Behavioral targeting has been called "the process of detecting segments of users with similar behaviors, in order to address effective ads to them (Boratto et al., 2016)." Some of the behavioral targeting tools are Google's AdWords, Facebook's Core Audiences, Amazon's Interest-based ads policy, Yahoo! Behavioral Targeting, DoubleClick, SpecificMedia, Almond Net, Burst, Phorm and Revenue Science.

Macia-Fernandez et al. (2012) give an overview of the process for user targeting by a website through advertising. The website will follow one of two strategies:

- Contextual advertising - use the contents of the web page for targeting an ad.
- Behavioral targeting – consider the profile of the user through navigation history, preferences and interests.

It has been shown that news, arts and sports websites have the most third-party trackers. Adult sites, non-profits and government sites have the least third-party trackers (Englehardt &Narayanan, 2016). The theory is that "these sites provide articles for free, and lack an external funding source, they are pressured to monetize page views with significantly more advertising."

There are several entities involved with ad targeting (Parra-arnau, 2017). These include:

- Publisher – owns the website and places ads for revenue on their web pages.
- Advertiser – wants to display the ads and designate who they want their ads targeted to.
- Advertising platform – group of entities that connect advertisers to publishers, they receive ads and place them with the publishers.

Some examples of ad platforms are DoubleClick (Google), Gemini (Yahoo!), and Bing Ads (Microsoft).

Google and Facebook use the term impression and Facebook describes an impression as a view where "the first time your ad is served to someone in either their News Feed, mobile Feed or as a right column ad, that will count as an impression[43]." With Facebook's cost per link business model, it will optimize "to find people most likely to click on your ad in that way" and on Google "your bids are optimized to favor ad slots that are more likely to become viewable[44]."

Ad Bidding

Bidding on Google AdWords uses "advanced machine learning to automatically optimize bids and offers auction-time bidding capabilities that tailor bids for each and every auction[45]." Targeting algorithms are generally proprietary (Parra-Arnau, 2017). Real-time bidding (RTB) is in Beta on Google[46] and is composed of a user interface, API, server and ad exchange.

Marketing companies are also involved in real-time bidding. Cross-referencing and profiling use the ad bidding system which occurs in less than a half of a second. Real-time bidding is best explained by a digital marketing agency:

> *"So, generally, pretty much always, what happens is a user will visit a page. That page has advertising on it, which is made biddable through ad exchanges. The webpage will call the ad exchange. The ad exchange will package up or announce that impression to people like us that sit there and listen to these impressions coming through exchanges.*
>
> *The ad exchange appends some information to this impression, such as time of day, day of week, the site, the browser, information about the user's computer, nothing personally identifiable. That's illegal, which is good but also sad. It allows us to start to make a decision about the value of that ad impression, whether we want to buy it for your advertising campaign, and as well, how much we want to pay for it if this is a true RTB part of the campaign.*
>
> *So to further enrich or to help us make better decisions, we won't rely just on that little bit of information that the ad exchange passes us, but we'll go and check with other third parties, but most importantly a lot of first-party data from yourselves as well. This is one of the most overlooked things in display advertising - the power of first-party data.*

[43] https://www.facebook.com/business/help/220734457954046

[44] https://support.google.com/adwords/answer/3499086

[45] https://support.google.com/adwords/answer/6268632

[46] https://developers.google.com/ad-exchange/rtb/open-bidder/

So this impression goes into the ad exchange, and the ad exchange announces it. Our DSP is listening to this and making decisions. While it's making that decision, it'll quickly go away and check. Imagine we've set up a campaign, and there's an awareness part of it where you've asked us to purely buy demographics, let's say females between the ages of 34 and 54. There's also a retargeting element to this campaign as well, where we use your site data to be more aggressive with that user because they may have fallen out of your purchase cycle or not[47]."

Using this information, Jellyfish then checks third-party data such as Experian for information like age, sex, income, home ownership based on the cookie matching. They also check for "intent behavior," described as other products a user may have been searching for to use as a basis for cross-checking other third-party data sources.

7. Blocking and Anonymity

Browsers contain mechanisms for managing cookies but research has implied that most users do not regularly delete cookies (Janc, & Zalewski, 2015). Users may block third-party cookies and allow first party cookies, but Janc & Zalewski (2015) point out that first party cookies are identified as belonging to the site the user visited as well as any "content loaded by the browser as a part of full-page interstitials, HTTP redirects, or click-triggered pop-ups."

Browsers also have controls to block flash cookies with extensions such as FlashBlock (Schneier, 2015). Other blocking tools are Adblock Plus which is a plugin for Google Chrome and Safari. Englehardt & Narayanan (2016) determined that Firefox's third-party cookie blocking and extensions like the Ghostery extension were effective at blocking third-party cookies.

There are several sites available that users can access to block advertiser's cookies. The Digital Advertising Alliance "establishes responsible privacy practices across industry for relevant digital advertising, providing consumers with enhanced transparency and control through multifaceted principles that apply to multi-site data and cross-app data gathered in either desktop or mobile environments[48]." Their companion site (youradchoices.com) allows the user to opt-out on participating companies[49]. Another opt-out site is Evidon (from Ghostery)[50] .

[47] http://www.jellyfish.net/news-and-views/real-time-bidding-made-real-simple

[48] http://digitaladvertisingalliance.org/

[49] http://youradchoices.com/

[50] https://www.ghostery.com/support/global-opt-out/

Advertising companies have begun paying for their ads to be unblocked, starting what is called "the ad blocking wars (Parra-arnau, 2017). It was found that in 2015 ad blocking cost publishers nearly $22 billion[51].

Online Safety and Protecting Privacy

There are three rules of online safety according to Krebs (2014):

- If you didn't go looking for it, don't install it.
- If you installed it, update it!
- If you no longer need it, remove it!

Ways to protect internet privacy according to Schneier (2015):

- Pay cash
- Refrain from using Google Calendar, webmail or cloud backup
- Use DuckDuckGo for internet (Angwin, 2014 concurs)
- Privacy enhancing technologies (PETs)- Lightbeam, Privacy Badger, Disconnect, Ghostery, FlashBlock
- Encryption – Microsoft's BitLocker, Apple's FileVault, Cryptocat
- Transport Layer Security (TLS) – formerly secure sockets layer (SSL) for encrypting web browsing, use HTTPS Everywhere
- Turn location services off
- Delete cookies everyday
- Distortion – swap affinity cards with neighbors
- Give out false information on forms
- Use Tor

There is a pessimistic outlook for completely retaining anonymity. Peacock (2014) states that "circumventing supercookies is almost impossible, given that much of the web content includes videos requiring widely used applications like Adobe flashplayer." Boda et al. (2011) describes the evercookie as

[51] https://pagefair.com/blog/2015/ad-blocking-report/

creating an “indestructible identifier.” In (Leon, et al., 2011), the authors found blocking and opt-out sites did not “empower study participants to effectively control tracking and behavioral advertising according to their personal preferences.”

Boda et al. (2011) suggest switching browsers and moving to browsers that report a “unified and uncommunicative attribute.” They also suggest making browser extensions capable of spoofing the UAS by setting fake system properties (operating system, time zone, screen resolution) or using universal font lists.

Mowery & Shacham (2012) identify the following ways to avoid browser fingerprinting:

- Turn off the <canvas> capability
- Apply noise
- Make all systems identical
- Mandate user approval for scripts requesting pixel data

They also suggest the use of private browsing or an anonymity network such as Tor. However, the authors state that “although Torbutton disables WebGL, it allows text rendering to a <canvas>, and is thus at present partly vulnerable to our fingerprint.” Acar et al. (2014) suggest possibly implementing crawlers for unwanted tracking but believe they are difficult to deploy.

The Panopticlick project presents methods to reduce fingerprinting including:

- Running Privacy Badger or Disconnect – this sends a Do Not Track
- Tools like NoScript for Firefox
- Using the Tor browser
- Disabling JavaScript
- Using a “non-rare” browser

They note that some trackers can still “slip past the net and that use of the Tor browser will slow down browsing speeds.” They also note that most websites currently ignore the Do Not Track signal[52].

- The Internet Patrol[53] provides the following suggestions:
- Stop browser from running Java, disable JavaScript and Flash scripting

[52] https://panopticlick.eff.org/about

[53] https://www.theinternetpatrol.com/how-to-reduce-your-internet-fingerprint-and-defeat-internet-fingerprinting/

- Use the Firefox NoScript plugin
- Use the phone to browse
- Use private browsing

Anonymity Networks and Privacy Enhancing Technologies

Privacy enhancing technologies (PETs) include other solutions to help a user remain anonymous. These include services on the hidden web (Haughey et al., 2016):

- Dynamic content
- Unlinked pages
- Private websites
- Limited access content
- Scripted pages

These services offer "a high degree of anonymity preservation, as well as much wider access to information" and are useful to both legitimate organizations and criminal organizations. PETs have tools available to find out if user data is being collected, such as Mozilla Lightbeam and PrivacyBadger. These tools have been reported to have "high accuracy in identifying which sources of user data, such as email or web search history, might have triggered particular results from online services such as advertisements (Aonghusa & Leith, 2016). In Lecuyer et al., (2014), the authors were able to track not only which sources triggered an advertisement, but also through correlating data were able to identify with which websites a user's data may have been shared.

Anonymity networks are communication protocols that were primarily designed to provide privacy for Internet users who lived in oppressive regimes, allowing them to evade censorship (Zhioua, 2015) or conceal their identity (Alsabah & Goldberg, 2016). Tor is the most widely used network with over 3000 relays, with estimates of hundreds of thousands of users to millions of users (Zhioua, 2015; Alsabah & Goldberg, 2016). Users include journalists, activists and users aiming to remain anonymous. Other networks include i2p and Freenet (Haughey et al., 2016).

Tor is a public network, operated by volunteers and is easy to use by installing the Tor browser bundle (Alsabah & Goldberg, 2016). Tor conceals the source, destination and contents of a message using encryption layers and uses the volunteer's servers as routers between the users and their destination (Haughey et al., 2016). The client encrypts their message and IP address of the destination. Additional encryption occurs between the entry and exit relays. The protocol uses "onion routing" and modifies traffic by merging small packets together, multiplexing transmission control protocol (TCP) streams and restructuring traffic into fixed-size cells (Zhioua, 2015). There can be delays and long download times that discourage users. Website fingerprinting is not eliminated by Tor with identification estimates as high as

90% (Zhioua, 2015). The Panopticlick project indicates that the Tor browser now includes patches for preventing font and canvas fingerprinting, as well as blocking JavaScript[54].

The i2p network is a private network that allows anonymous public browsing and its users serve as routers (Haughey et al., 2016). The network uses a concept called "tunnels", that are unidirectional for either sending or receiving information. The system relies on a database of participating gateways for inbound tunnels and router information relating to the IP address and TCP port. The i2p uses "garlic routing where packets are "packaged together in bulbs or cloves within one garlic message that is sent over a tunnel, although typically only one clove is sent." Encryption is used throughout the process.

Freenet uses freesites in both Opennet and Darknet modes and allows anonymous social networking and email. Users contribute bandwidth and storage in this distributed system that relies on encryption and non-permanence of data (Haughey et al., 2016). The system uses clusters of nodes and "each node is supposed to be aware only of its own nearest neighbors and requests for files are forwarded on by those nodes if they do not hold a copy, until the request reaches a node that does."

Vulnerabilities have been exposed in the anonymity networks and the networks respond by addressing them. This has been called an "arms race," resulting from the debate between anonymity and privacy (Haughey et al., 2016).

[54] https://panopticlick.eff.org/about

8. References

Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., & Diaz, C. (2014). The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS'14)*, pp. 674-689.

Agarwal, L., Shrivastava, N., Jaiswal, S., Paniwani, S. (2013). Do Not Embarrass: Re-examining User Concerns for Online Tracking and Advertising. *Proceedings of the 9th Symposium on Usable Privacy and Security (SOUPS'13)*.

Alsabah, M. & Goldberg, I. (2016). Performance and Security Improvements for Tor: A Survey. *ACM Computing Surveys*, 49 (2).

Angwin, J. (2014). Dragnet Nation. New York, New York: Henry Holt and Company.

Aonghusa, P., & Leith, D. (2016). Don't Let Google Know I'm Lonely. *ACM Transactions on Privacy and Security*, 19(1).

Berger, D. (2010). Balancing Consumer Privacy with Behavioral Targeting. *Santa Clara Computer and High Technology Law Journal*, 23, pp. 3-61.

Boda, K., Foldes, A., Gulyas, G. & Imre, S. (2011). User Tracking on the Web via Cross-Browser Fingerprinting. *Proceedings of NordSec*, pp. 31-46.

Boratto, L., Carta, S., Fenu, G., Saia, R. (2016). Using Neural Word Embeddings to Model User Behavior and Detect User Segments. *Knowledge-Based Systems*, 108, pp. 5-14.

Eckersley, P. (2010). How Unique is Your Web Browser? *Proceedings of the 10th International Conference on Privacy Enhancing Technologies*, pp. 1-18.

Englehardt, S. & Narayanan, A. (2016). Online Tracking: A 1-Million-Site Measurement and Analysis. *ACM Conference on Computer and Communication Security*.

FTC (2015). Internet of Things: Privacy & Security in a Connected World. Federal Trade Commission. Online: https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.

Ghaleb, T. (2016). Techniques and Countermeasures of Website/Wireless Traffic Analysis and Fingerprinting. *Cluster Computing*, 19, pp. 427-438.

Gomez, J., Pinnick, T, & Soltani, A. (2009). KnowPrivacy. UC Berkeley, School of Information. Online: www.knowprivacy.org.

Haughey, H., Epiphaniou, G. & Al-Khateeb, H. (2016). Anonymity Networks and the Fragile Cyber Ecosystem. *Network Security*, March, pp. 10-17.

Janc, A. & Zalewski, M. (2015). Technical Analysis of Client Identification Mechanisms. Online: https://www.chromium.org/Home/chromium-security/client-identification-mechanisms.

Kanoje, S., Girase, S., Mukhopadhyay, D. (2014). User Profiling Trends, Techniques and Applications. *International Journal of Advance Foundation and Research in Computers*, 1(1).

Kende, M. (2016). Global Internet Report 2016. Internet Society. Online: www.internetsociety.org.

King, N., & Forder, J. (2016). Data Analytics and Consumer Profiling: Finding Appropriate Privacy Principles for Discovered Data. *Computer Law & Security Review*, 32, pp. 696-714.

Krebs, B. (2014). Spam Nation. Naperville, Illinois: Sourcebooks.

Lecuyer, M., Ducoffe, G., Lan, F., Papancea, A., Petsios, T., Spahn, R., Chaintreau, A., Geambasu, R. (2014). XRay: enhancing the Web's Transparency with Differential Correlation. *Proceedings of the 23rd USENIX Security Symposium*. Online: https://www.usenix.org/system/files/conference/usenixsecurity14/sec14-paper-lecuyer.pdf.

Leon, P., Blasé, U., Balebako, R., Cranor, L., Shay, R., & Wang, Y. (2011). Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising. *Proceedings of CHI 2012*, pp. 589-598.

Longman, A. (1998). A History of HTML. World Wide Web Consortium (W3C). Online: http://www.w3.org/People/Raggett/book4/ch02.html.

Lyon, G. (2008). Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning. Insecure.Com, LLC.

Macia-Fernandez, G., Wang, Y., Rodriguez-Gomez, R. & Kuzmanovic, A. (2012). Extracting User Web Browsing Patterns from Non-Content Network Traces: The Online Advertising Case Study. *Computer Networks*, 56, pp. 598-614.

Malandrino, D. & Scarano, V. (2013). Privacy Leakage on the Web: Diffusion and Countermeasures. *Computer Networks*, 57, pp. 2833-2855.

Martin, K. (2016). Understanding Privacy Online: Development of a Social Contract Approach to Privacy. *Journal of Business Ethics*, 137, pp.551-559.

Mathews-Hunt, K. (2016). CookieConsumer: Tracking Online Behavioural Advertising in Australia. *Computer Law & Security Review*, 32, pp. 55-90.

Mulazzani, M., Reschl, P., Huber, M., Leithner, M., Schrittwieser, S., Weippl, E. (2013). Fast and Reliable Browser Identification with JavaScript Engine Fingerprinting. *Proceedings of Web 2.0 Security & Privacy (W2SP 2013)*.

Mowery, K. & Shacham, H. (2012). Pixel Perfect: Fingerprinting Canvas in HTML5. *Proceedings of Proceedings of Web 2.0 Security & Privacy (W2SP 2012)*, pp. 1-12.

Olejnik, L., Castelluccia, C. & Jane, A. (2014). On the Uniqueness of Web Browsing History Patterns. *Annals of Telecommunications*, 69, pp. 63-74.

Parra-Arnau, J. (2017). Pay-per-tracking: A Collaborative Masking Model for Web Browsing. Information Sciences, 385-386, pp. 96-124.

Peacock, S. (2014). How Web Tracking Changes User Agency in the age of Big Data: The Used User. *Big Data & Society*, July-December, pp. 1-11.

Pokharel, S., Choo, K., Liu, J. (2017). Mobile Cloud Security: An Adversary Model for Lightweight Browser Security. *Computer Standards & Interfaces*, 49, pp. 71-78.

Roderick, L. (2014). Discipline and Power in the Digital Age: The Case of the US Consumer Data Broker Industry. *Critical Sociology*, 40(5), pp. 729-746.

Schneier, B. (2015). Data and Goliath. New York, New York: W.W. Norton & Company.

Soltani, A., Canty, S., Mayo, Q., Thomas, L., & Hoofnagle, C. (2009). Flash Cookies and Privacy. Online: https://papers.ssrn.com/sol3/papers.cfm?abstract-id=1446862.

Westin, A. (1968). Privacy and Freedom. *Washington and Lee Law Review*, 25(1).

Wills, C. & Zeljkovic, M. (2011). A Personalized Approach to Web Privacy – Awareness, Attitudes and Actions. *Information Management & Computer Security*, 19(1), pp. 53-73.

Zhioua, S. (2015). The Web Browser Factor in Traffic Analysis Attacks. *Security and Communication Networks*, 8, pp. 4227-4241.

Ziegeldorf, J., Morchon, O. & Wehrle, K. (2013). Privacy in the Internet of Things: threats and challenges. *Security and Communications Networks*, pp. 2728-2742.