International Journal of
Applied Mathematics,
Electronics and Computers
ISSN:2147-8228 www.atscience.org/IJAMEC
Advanced Technology and Science
Original Research Paper

User tracking mechanisms and counter-measures

Asra Ishtiaq[1], Salma Hussain Abbasi[2], Muhammad Aleem*[3], Muhammad Arshad Islam[4]

Accepted : 24/05/2017 Published: 30/07/2017 **DOI: 10.18100/ijamec.2017528829**

Abstract: Online customers have lots of alternatives while making a purchase or discovering information on websites. Now a day, tracking services and passive traffic check are widely used to collect knowledge about user's internet activities and interests. For end users, such tracking has significant privacy implications. So, Privacy becomes a sensitive issue which attracts a lot of user's attention. Discrimination is a way to differentiate, isolate, or make a difference. User discrimination is a tactic used to present personalized content to the user, based on user profile. It helps the web owner to improve the content of website. It also ensures that the content caters to a largest population of the user visiting their website. Our major contribution in this paper is to empirically show the tracking mechanism used for user discrimination on the Web, and also provide a defense mechanism against these tracking mechanisms.

Keywords: *Counter-measures, Discrimination, Privacy, Tracking mechanism*

1. Introduction

Discrimination is an uneven behaviour based on shared agreement, logical or illogical reason presented to an individual or community. It is exploiting several aspects. World Wide Web is evolving to assist improvement and delivery of web pages collected numerous sources. People search for information for motivation, comprehension, empowerment, knowledge, and service to do something or basically to gain knowledge. We live in an information society where different organizations follow user activities on the Web to analyze and gain knowledge about user's interest on the basis of collected information in order to improve their website and also ensure that the content accommodates to the large user population visiting their web pages. For Example, the owner of the website may use this information to present personalized content that goes with users' interests. The advertising agencies may use this information to target ads that may go better with user concern [2, 3, 5, 10, 11].

The use of the web is tracked using both the active and passive techniques [15]. Third-party scripts and plug-in sets are employed for the visited websites to dig out and gather information about a user's every click, the time spent on every page, and the information that can help distinctively classify every browser and customer. At the same time, numbers of trackers permit advanced finger-printing from client-side for user information. Also, a number of tracking mechanisms are employed at server-side tracking, as a download of comparatively easy third-party contents such as CSS files, personalized fonts, or JavaScript libraries do not comprise of tracking code themselves Roesner [1, 6—9]. In passive tracking, client actions are classically dug out from traffic logs collected from traffic monitoring within a network or in a cloud-operated data-hub that hosts several types of the contents. At this time, TCP/IP header information such as IP addresses and application specific information e.g., HTTP header information and payload data as well as information delivered to third-party trackers through unencrypted HTTP transmits can be used to recognize client. The major financial model at the back of the majority of Internet services is to offer service free of charge, magnetize users, gather information, and supervise these users to monetize this information. The gathering of personal information is completed by complicated means [4] and become the concentration of privacy supporter, controller, and the typical media. A simple query is that what is done with all the gathered information? The well-liked reply is the collected information is used for targeted advertising.

In this paper, we identify the tracking mechanisms used to track the user activities and based on those make a user profile. Using user profile, customized content is a target to the user. Based on our extensive study of state of the art, identify the defense mechanisms used against tracking mechanisms. Section 2 contains the detailed description of tracking mechanisms. Section 3 presents the detailed description of defense mechanisms employed against user tracking. Section 4 concludes the paper.

[1,2,3,4] *Department of Computer Science, Capital University of Science and Technology, Islamabad – 44000, Pakistan*
* *Corresponding Author: Email: aleem@cust.edu.pk*

2. User tracking mechanisms

Web search engines assist users to find useful information on the *World Wide Web* (WWW). Each user seeks different information using unique query text. Therefore, the search results should be adapted to users with a different information need. In order to improve the content and rating of the website, user activities are being tracked by using different mechanisms. These mechanisms are briefly discussed in next sections.

2.1. Stateless tracking

Stateless tracking is fingerprint-based tracking mechanism. In this method, website learns properties of browser [3]. The combination of these properties forms a unique identifier. This identifier is used for identification of the user.

2.2. Stateful tracking

Stateful tracking is cookie based tracking mechanism. Websites exploit cookies data for tracking users [13]. Generally, websites embed a unique identifier into stateful web technology. This unique identifier is employed by the website for identification of

the user.

2.3. HTTP cookies

An HTTP cookie is a small piece of information that a server transmits to the user's web browser. It is stored on the user browser and sent back with next request to the same server. This helps the server in identifying that two requests are made from the same browser so that the user's session could be maintained [10]. The main purpose of these cookies is to remember stateful information for example products added in the shopping cart and for remembering data entered into *form-fields* like name, address, and passwords [2]. With the help of HTTP cookies, the server is able to store and retrieve data on the client side.

2.4. Local shared objects

Local shared objects are also known as Flash cookies, are pieces of data stored on the user computer by websites using Adobe Flash [15]. The storage capacity of Flash cookies is more compared to the HTTP cookies. Flash cookies can save up to 100KBs whereas the HTTP cookies cannot store more than 4KBs of data [10]. A flash cookie can store browsing history, location information, and key-logged data. Flash cookies are hard to spot because they do not appear in the list of cookies accessed with help of cookie manager of a browser [2]. Flash cookies are stored as binary data; therefore the ordinary user cannot interpret the contents of these cookies.

2.5. First-party cookies

A cookie is a small piece of information that is placed on a hard drive of a user's computer. The cookie data is placed by the server of the website that is visited by the user in order to recognize specific browser [1]. Within every cookie, domain is specified which is the website by which specific cookie is set. First party cookies refer to those cookies, which are set by the websites the user interacted directly. If a certain user visits a website, for example, *www.xyz.com*, the domain name stored in the cookie will xyz.com and these are called first party cookies [13]. First party cookies involve a high level of trust.

2.6. Third-party cookies

Third-party cookies refer to cookies implanted by websites other than the one user chooses to interact with [10]. If a certain user visits a website, for example, *xyz.com* and domain of cookies stored on user computer are other than xyz.com that cookie is referred as the third party. Third party cookies are set when a request goes from a website to a domain name that is not the website user chosen to interact [13].

2.7. Ever cookies

Ever cookie is a JavaScript API created by Samy Kamkar that creates zombie cookies in a browser. Ever cookies are used for identification of the user even after deletion of HTTP, Flash, and other types of cookies [5]. Ever cookie achieves this by employing several types of storage mechanisms that are present in browsers [2]. Additionally, if a certain user removes cookies from one storage place, it repopulates deleted cookies from other locations where additional copies of the cookies are available. Ever cookies are used for tracking Tor [36] users.

2.8. Zombie cookie

A zombie cookie is an HTTP cookie which can be recreated even after deletion. These cookies are stored in different backup places outsides the web browser which makes them difficult to remove. Therefore, deletion from one place the cookies is recreated from available backups locations [20]. As zombie cookies are stored outside the browser, therefore a user using a different browser on the same machine can easily be tracked. Zombie cookies operate differently compared to the ordinary cookies [12].

2.9. Active fingerprinting

Active fingerprinting is the process of transmitting packets to a remote host and analyzing corresponding replies [9]. This allows the scanner to obtain more accurate results than a passive scanner. This technique also involves the shorter amount of time for tracking.

2.10. Passive fingerprinting

In passive fingerprinting, packets from a certain user are analyzed [9]. In this case, finger printer acts as a sniffer and doesn't put any traffic on a network. Passive fingerprinting is the process of analyzing packets from a host on a network. It is called passive because it doesn't involve communicating with the host being examined [9].

2.11. Canvas fingerprinting

Canvas fingerprinting is a browser fingerprinting technique that tracks users using HTML5 canvas element instead of using cookies [4]. Canvas fingerprinting tracking mechanism requires HTML5 Canvas API and JavaScript [2].

2.12. Audio fingerprinting

This is relatively a new technique of fingerprinting that works by creating fingerprints of machine's audio stack using Audio Context API. Using this mechanism, the audio signal collects the user machine data and uses that for identification of specific device [19]. This fingerprinting technique uses the aspect of audio signals on different machines or browsers and may differ from each other due to the difference of hardware or software platform [20]. As audio fingerprinting is not very common so most of the tracking defense tools do not shield it against therefore it has a capacity of bypassing most of the privacy tools [19].

2.13. Battery Based Fingerprinting

This fingerprinting technique tracks devices by creating fingerprints on the basis of battery status of the device. For fingerprinting, it uses HTML5 features such as Battery Status API [11]. The official World Wide Web consortium states that this API creates a weak fingerprint, which has very less effect on privacy. However, French and Belgian security researchers found that third party based tracking mechanism can easily exploit user visits across multiple sites in short time intervals by exploiting information like battery level, discharging, and charging time [11].

2.14. Font Fingerprinting

Font fingerprinting is about what fonts you have, and how they are drawn [11]. It is based on measuring dimensions of filled with text HTML elements [4]. It is feasible to build an identifier that can be used to track the same browser over the time. Font fingerprinting is hard to counter [6].

2.15. Panopticlick fingerprinting

When you visit a website, online trackers and the sites themselves may be able to recognize you [6]. A uniquely configured system is easily identifiable even if privacy protective software is used. It is easy to configure your browser to prevent such tracking, however, a lot of people don't know how [5]. Panopticlick [37] project that helps users to generate their fingerprints and understand how

unique their browser is [5]. *Panopticlick* identifies users based on the screen resolution and color depth, time zone, version of plugins and extensions; the list of fonts available on the computer; information whether JavaScript is enabled or not; and information whether certain persistent storage requests are accepted by the browser. One of the limitations of *Panopticlick* technique is that it is unable to identify identically configured machines.

3. Cross-browser fingerprinting

Cross-browser fingerprinting is a browser-independent fingerprinting method. In this technique, browser-independent features like fonts and time-zone etc. are employed for user identification [5]. This technique does not store any information on the user's computer; instead extracts certain parameters that are accessible through the web browser, e.g., time zone and screen resolution [5] and identify the user on the basis of these parameters. Like *Panopticlick* fingerprinting technique also unable to identify users with identically configured machines.

3.1. Visited link tracking

Websites mostly use a different color for visited and unvisited links. This information can be extracted using JavaScript [3]. Websites use this single bit of information for multiple-session tracking.

3.2. History stealing

History stealing is basically a process of learning about user's web browsing history. This is done by exploiting link styling as visited and non-visited links are displayed in different colors [5]. This information can be used for learning user [20]. History stealing can be used by one organization to check whether users have visited their competitor's website sites. Advertising networks also often employ history stealing.

3.3. Detection of HTTP Proxies

Proxy detector (sometimes called *proxy checker*) work by testing for HTTP headers, which are commonly set by the proxy servers. If proxy headers are not found it is assumed that you are either using the direct connection or a high secrecy proxy server. Proxy tester won't detect a transparent proxy, which doesn't over wire IP address of the sender and thus does not provide any secrecy [4]. Important to remember that even with high secrecy proxy servers, user activity could be logged and if the connection is not encrypted your ISPs may log requests to the proxy server. One more option is to use the Tor network [36] as an alternative to a proxy server [16]. Unfortunately, a lot of sites detect and block Tor. The finest solution is to use a VPN service supplier with a huge pool of IP addresses [22].

3.4. Online Advertising

Online advertising is also known as online marketing or web advertising [22] It is a type of marketing and advertising that make use of the Internet to convey promotional marketing messages to clients/consumers [7]. Consumers/clients view online advertising as an unnecessary distraction with a small number of benefits and have increasingly twisted to ad blocking for a variety of reasons. It consists of email marketing, search engine marketing, social media marketing, several types of display advertising as well as web banner advertising, and mobile advertising [14]. Online advertising often involves both a publisher (who combine advertisements into its online content) and an advertiser (who give the advertisements to be shown on the publisher's content) [7]. In online advertising, the further possible participants include advertising agencies who help to produce and place the ad copy, an ad server which technically sends the ad and tracks information and advertising affiliates who do autonomous promotional work for the advertiser [18].

3.5. Deep packet inspection

Deep packet inspection is also called complete packet inspection and information extraction, is a type of computer network packet filtering technique [20]. It inspects the data part and possibly the data header of a packet as it transit an inspection position, searching for protocol non-compliance, viruses, spam, intrusions, or defined criteria to make a decision whether the packet may bypass or if it needs to be routed to a different target. Additionally, statistical information has collected that work at the application layer of the *Open Systems Interconnection* model (OSI) [10]. An IP packet contains several headers. Usually, network equipment only employs the first level of these (the IP headers) for usual operations, however, use of the second level header (such as TCP or UDP) is normally considered as a deep packet inspection (typically called stateful packet inspection).

Deep packet inspection and filtering allow advanced network management, consumer service, and security functions as well as Internet data mining [22]. DPI is used in a broad range of applications called "enterprise" level (business and larger organization), telecommunications service providers, and in e-government applications.

3.6. Third-Party Analytics

Web sites that wish to analyze traffic mostly use the third-party analytics like Google Analytic [49]. On such websites, the webpage's visited by the user include a script from third-party analytics. Third party analytic sets cookies on user browser that includes unique identifiers. Their unique identifiers are used by Third-party analytic for tracking visitors. As the third-party analytic script runs in website's context, the resulting cookie is site owned. The unique identifier set by Third-party analytic on each website is different for the same user so cross-site tracking is not possible.

3.7. Third-Party Advertising

Third-party ad serving refers to a common online situation in which a website publisher presents content for users along the advertising content delivered by some other provider [12]. This is a comparatively common approach employed by the websites to get advertising revenue. This technique is also useful for small businesses to market the targeted consumers [13]. Though, privacy issues are a concern for consumers.

3.8. Third-Party Advertising with Popups

Pop-up ads or pop-ups are the type of online advertising on the World Wide Web employed to magnetize web traffic or capture email addresses. Pop-ups are normally small new web browser windows to exhibit advertisements [1]. The pop-up window holds an advertisement that is usually generated by the JavaScript using cross-site scripting. It can also be produced by other vulnerabilities/security holes in browser security. A variation on the pop-up window is the pop-under advertisement, which opens a new browser window hidden under the active window [1]. Pop-ups do not interrupt the user instantly and not seen until the covering window is closed. Pop-up window constructs are more difficult to determine which web site spawned them [14].

Table 1. User tracking mechanisms

Tracking Mechanisms / Research papers	Stateless Tracking	Stateful Tracking	Active Fingerprint	Passive fingerprint	Canvas Fingerprint	Audio Fingerprint	Battery Fingerprint	HTTP cookies	Flash cookies	Ever cookies	Zombie cookies	First party cookies	History stealing	Third party cookies	Font fingerprinting	HTTP Proxies detection	Panopticlick fingerprint	Cross-browser fingerprint	Advertising	Visited link	Deep packet inspection	Third Party Analytics	Third-Party Advertising	Third-Party Popups	Third-Party Networks	Cookie syncing	Cookie re-spawning	history fingerprint
[3], 2012	✓	✓	✓	✓										✓								✓	✓					
[25], 2013														✓								✓						
[4], 2013			✓	✓	✓	✓									✓	✓		✓										
[5], 2012		✓	✓	✓											✓		✓											
[7], 2013	✓							✓	✓			✓		✓					✓			✓	✓					
[8], 2013																												
[9], 2014	✓	✓	✓	✓											✓													
[11], 2012		✓	✓	✓											✓													
[12], 2012	✓							✓	✓			✓		✓														
[6], 2013		✓	✓	✓											✓													
[13], 2016																						✓	✓	✓	✓			
[1], 2012	✓	✓	✓	✓				✓	✓			✓		✓					✓			✓	✓	✓	✓			
[2], 2014					✓					✓																✓		
[14], 2016	✓	✓						✓	✓	✓	✓	✓		✓												✓		
[15], 2015	✓							✓	✓			✓		✓					✓			✓	✓					
[16], 2013								✓	✓			✓	✓	✓														
[17], 2016	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓							✓	✓	✓
[18], 2012								✓	✓			✓		✓						✓								
[23], 2013			✓	✓				✓	✓			✓		✓					✓							✓	✓	
[47], 2011																							✓					
[22], 2014																			✓	✓			✓					
[21], 2016	✓	✓	✓	✓				✓	✓	✓	✓		✓	✓				✓	✓	✓		✓	✓			✓	✓	
[19], 2016	✓	✓	✓	✓	✓	✓	✓											✓										✓
[20], 2012	✓							✓	✓	✓	✓	✓		✓													✓	
[24], 2016	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓			✓	✓	✓
[41], 2011	✓							✓				✓				✓										✓		✓
[42], 2013		✓	✓	✓	✓										✓													
[44], 2014	✓							✓	✓	✓		✓	✓	✓		✓												
[10], 2011								✓	✓	✓	✓	✓		✓		✓	✓	✓			✓							
[46], 2012	✓							✓	✓			✓		✓					✓	✓			✓					
[39], 2012												✓	✓	✓					✓	✓			✓	✓	✓			
[29], 2013														✓					✓			✓	✓	✓	✓			
[30], 2013	✓	✓						✓	✓	✓	✓	✓	✓	✓		✓				✓		✓	✓		✓			
[33], 2013																			✓			✓	✓	✓	✓			
[31], 2012								✓	✓	✓	✓	✓		✓														

integration of data providers with their own platforms so that all of their partners can benefit from the integration [2].

In order to precisely target an audience, advertisers need to include user data from various domains and sources, which occur as part of data-buying agreements and partnerships between different companies. Advertisers are able to attain this by mapping user IDs from one system to another [2]. A case of this would be mapping a user's ID from a *Demand-Side Platform* (DSP) to a *Data Management platform* (DMP). This process is known as cookie syncing [20].

3.9. Third-Party Advertising Networks

An online advertising network or ad network is a corporation that connects advertisers to web sites that want to host advertisements [1]. The key function of an ad network is an aggregation of ad space from publishers and matching it with advertisers' demand. The phrase "ad network" by itself is "Ad Network" or "Print Ad Network", however, it is increasingly used to mean "online ad network" as the effect of aggregation of publisher ad space and sale to advertisers is most commonly seen in the online space [1]. The fundamental difference between traditional media and the Internet is that it can be used as an alternative to analogical media [13].

3.10. Cookie syncing

Cookies are the most important mechanism by which publishers, advertisers, ad networks, ad exchanges, demand side platforms and data exchanges store and track information about users. Cookies are domain specific. It means that a cookie set by domain foo.com cannot be read by a server from domain bar.com. It is referred as the process of mapping user Ids from one system to another [14]. Ad exchanges and DSPs use cookie-synching to allow the

3.11. Cookie re-spawning

Cookie re-spawning is the procedure of recreating browser cookies from information that has been deleted [14]. With cookie re-spawning, a corporation can take information store up in flash cookies and use it to reconstruct a cookie in a browser [2]. There is concern that cookie re-spawning can breach a user's privacy and become difficult for the operation of the computer in the identical way that any kind of cookie storage can eventually challenge an operating system [20].

4. Voiding user tracking on the web

Privacy is a sensitive topic that attracts a lot of users' attention. A user wants typically wants to maintain strict anonymity on the web. In this section, we provide a detailed description of different defense mechanisms used against web tracking mechanisms.

Opt-out cookies defense mentioned in Table II is employed to stop a website from storing future cookies. If the opt-out cookies defense is enabled the visited website will store opt-out cookies in the user browser folder [3]. This cookie will stop that website from installing third party advertiser or other cookies in future in user browser [8]. This will prevent third party advertisers from tracking user preferences from the website. The drawback to using opt-out cookies is that each website creates its own opt-out cookie which will block cookies for that specific website only. Generalized cookie blocking can also be done through browser's cookies settings [1]. Another disadvantage of opt-out cookie defense mechanism is that clearing cookies may also delete opt-out cookies and tracking will be possible again [2]. Moreover, there is a disagreement about what op-out request actually blocks. Some websites only block advertisements but still maintain user profiles. A web tracking industry-wide consensus is required for opt-out cookie success [32].

4.1. Do Not Track Header

Do Not Track (DNT) header approach mentioned in Table II is similar to that of op-out cookie [3]. DNT is a proposed HTTP header field that signalizes the web application that user wants to disable either its tracking or cross-site tracking [9]. Three values can be set for DNT header according to user's choice. If the user does not want to be tracked it is signalized by setting the value of DNT header to 1, DNT header is set to 0 if user allows tracking, and null represents that preference is not expressed by the user [6]. The main advantage of DNT header is that it is generalized and will work for every website and does not require to be set by each advertising network as in opt-out cookies [14]. Since this technology is not standardized, it totally depends on web sites to honor it or not. Therefore, it is possible that even after setting DNT header to 1; tracking is still being done [14].

4.2. Proxy server

The proxy server mentioned in Table II is an intermediary between client and destination server. All the requests and data from the client to destination server or vice versa are passed through the proxy server [32]. So proxy server knows IP address of the client whereas the destination server only knows the address of the proxy server. So by using proxy servers, the client can hide IP address from the destination server [1]. As proxy server will know the address of client so the client has to trust the proxy server. At the proxy server, data can be analyzed by someone who takes control of the proxy server. Therefore, a solution might be to send encrypted data over proxies that cannot be trusted so critical data like passwords remain safe (in the case of intrusion). Even with the encrypted data, critical information (such as IP address) can be saved at a proxy server [21]. To mitigate this problem, a chain of proxy servers can be employed that make it difficult to identify client system IP addresses.

4.2.1. Anonymous proxy server

An anonymous proxy server or web proxy is detectable by the destination server but hides IP of the client. HTTP_X_FORWARDED_FOR header does not send IP address of user instead send the address of proxy or blank. HTTP_VIA header is also sent that shows the user is using a proxy server.

4.2.2. High anonymity proxy

High anonymity proxy hides IP of client and is not detectable by the destination server. The destination server does not know that requests are being forwarded through a proxy server. High anonymity proxy just sends REMOTE_ADDR header showing the client as an ordinary user (surfing without proxy).

4.3. End-to-End Encryption

End-to-End encryption is a mechanism for transferring messages in which only communicating parties can read the message being transmitted [10]. This mechanism defends against eavesdropping [7]. Third parties cannot access the message being transmitted [15].

4.4. Clearing client-side state

Clearing client-side state constantly will help the user get new identity from tracker every time [3]. The tracker will not be able to create a user profile for tracking purpose.

4.5. Blocking popup

Many browsers provide the feature of blocking popups. Blocking popups will enable the user to avoid opening the popups [1]. This will prevent the third party from redirecting the user to other pages.

4.6. Disabling script execution

Disabling execution of Javascript defends against fingerprinting technique based on JavaScript [17]. Firefox extension No Script allows the user to blacklist domains for trying disabling script execution. The disadvantage of this technique is that some websites will not work properly.

4.7. TOR

Tor mentioned in Table II is a network of virtual tunnels designed for anonymous web surfing. Tor was developed for United States Navy for the protection of communication. Today, Tor is open to the public and can be used by ordinary individuals too. Mostly, Tor is associated with criminal activities; however, it is used by law enforcement officials, whistle-blowers, and reporters because of security reasons [11]. Tor uses onion routing technique for anonymous communication. In onion routing, a chain of proxy servers are employed for transmission of data from the user to destination and link between proxies is encrypted. The first proxy in the chain is called Tor proxy, which picks nodes from the Tor network. The message is encapsulated in layers of encryptions applied by Tor proxy (based on symmetric keys). In Tor, the message is transmitted through at least 03 different nodes called onion routers. The message is transferred through nodes selected by Tor proxy hop-by-hop. This process is repeated till massage is transferred to the destination node. Each node removes one layer of encryption and only knows the address of preceding and successive node in the path. So it is very difficult to find the IP address of the user.

End to end encryption is not used by Tor so the last node in path removes last encryption layer and send unencrypted data to the destination. Therefore, anyone having access to network between end-node and destination can capture the message being transmitted. Although Tor is better than traditional proxy servers, however, it does not prevent tracking completely. Tor prevents IP-based tracking however it does not provide prevention to non-IP based attacks like fingerprinting and end-to-end timing method [6]. Tor also slows down browsing. Access to Dark web is possible through Tor. For improved privacy, Tor is usually employed in combination with other software's like Privoxy [34]. *Privoxy* is placed at the front end of Tor for filtering purposes.

Table 2. Defense strategies against tracking mechanisms.

Defense Mechanisms / Tracking Mechanisms	Opt-Out Cookies	Browser-based blocking / plugins	HTTP DNT Header	Proxy Servers	Tor	Privoxy	End-to-end encryption	Private Browsing modes	FP Block	Clearing client-side state.	Blocking Pop ups	Share Me Not Defense	Open WPM	Tracking Observer	Disabling Script Execution
IP tracking				✓	✓										
Active Finger Printing					✓				✓				✓		✓
Passive finger Printing		✓			✓				✓				✓		✓
Canvas Finger Printing									✓				✓		✓
Audio Context Finger Printing															✓
Font Finger Printing		✓							✓				✓		✓
Battery API Finger Printing		✓													✓
Cross Browser Finger printing		✓							✓				✓		✓
Cookie Re-spawning					✓					✓					
Ever Cookies	✓				✓					✓					✓
Super Cookies/ Zombie Cookies		✓								✓					✓
HTTP Cookies	✓		✓			✓		✓		✓			✓		✓
Flash Cookies	✓		✓			✓		✓		✓			✓		✓
First-party Cookies	✓	✓	✓			✓		✓		✓			✓		✓
Third-party Cookies	✓		✓			✓		✓		✓			✓		✓
History Stealing										✓					
System-finger Printing Plug-ins		✓			✓			✓					✓		
Online Advertising		✓				✓									
Visited link tracking		✓						✓		✓					
Third-Party Analytics		✓										✓		✓	
Third-Party Advertising												✓		✓	
Third-Party Advertising with Pop ups											✓	✓		✓	
Third-Party Advertising Networks												✓		✓	
Cookie Syncing						✓				✓					✓
Deep Packet Inspection							✓								
Panopticlick Finger Printing															✓

4.8. Privoxy

Privoxy is a non-caching proxy server between the browser and the Internet. It is used for enhancing privacy, eliminating junk from websites, and manipulating cookies [24]. It filter outs web content, ads, script redirection, and outbound personal information contained in HTTP headers and cookies. Privoxy is customizable by the users according to their needs for both stand-alone and

multi-user network. Privoxy can be used along with other proxy mechanisms for enhanced privacy [32]. Privoxy is mostly used in combination with Tor and Sqiud [8].

4.9. Private browsing mode

Privacy mode (within the browser) is a privacy feature that secures browsing experience compared to the normal browsing mode [32]. In private mode, web cache, browsing history, cookies, and local storage are disabled [11]. This mode provides privacy protection on the local machine and hides traces from a person having physical access to a computer [11]. However, tracking can be done for private mode from a web server and fingerprint is also possible.

4.10. Browser-based blocking and extensions

The browser extension is a plug-in that expands the functionality of a web browser in a number of means. A few extensions are authored by means of web technologies such as HTML, JavaScript, and CSS. Browser extensions can be employed to modify the consumer interface of the web browser without the directly disturbing viewable content of a web page; for example, by adding a toolbar etc [6], [9].

4.11. FP-Block

FP-block is a defense against fingerprinting. It detects fingerprinting by detecting scripts through analyzing the JavaScript code embedded in a website. It also prevents leakage of fingerprints to the third party server [34]. FP block focuses on detection of font and plug-in enumeration functionality. When fingerprinting is detected, it blocks the data transmission to third-party servers [6]. FP block works on the principle of "separation of web identities".

For each website, new fingerprints are generated and this identity is not used for other websites. Every new identity is different from the previous identities. Therefore, third party trackers are not able to link users to two websites as both websites have different fingerprints. FP block defends against active and passive fingerprinting [9]. FP block can only deal with HTTP and Javascript based fingerprinting.

4.12. Share Me Not

Share Me Not [1] is a browser extension for protecting user tracking by third-party tracker. The main focus of this extension is to protect the user from being tracked by social widgets [1]. The social widgets supported by ShareMe Not are Facebook, Google, Twitter, Add This, YouTube, LinkedIn, Digg, and Stumble upon. A user can choose to be tracked or not depend on the preferences [1]. The user will not be tracked until widget button is not clicked however if the user explicitly clicks the button tracking for that site will begin.

4.13. Tracking Observer

Tracking observer mentioned in Table II is a Chrome extension that helps in detecting, measuring, and blocking third-party trackers. Each tracker behaves differently so can be detected based on their behavior [17]. Tracking Observer detects tracker automatically based on their in-browser behavior and categorize them according.

5. Conclusions

Different mechanisms are used to extract useful information from a user profile to provide personalized content to the user. The client profile is a practice of discrimination facilitated all the way through technologies, such as cookies, Font Fingerprinting, Panopticlick fingerprinting, cross browser fingerprinting, session tracking, Deep packet inspection etc. This would help the web owner to improve the website content and also ensure that the content accommodates to the largest population of the user visiting their website. The goal of this study is to determine the tracking mechanism used against the user for discrimination on the web and also find defense mechanisms against these tracking mechanisms. Table 1 shows the Number of tracking mechanisms identified in different papers and table 2 shows the defense mechanisms used against a number of tracking mechanism.

References

[1] Roesner, F., Kohno, T., & Wetherall, D. (2012, April). Detecting and defending against third-party tracking on the web. In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (pp. 12-12). USENIX Association.

[2] Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., & Diaz, C. (2014, November). The web never forgets: Persistent tracking mechanisms in the wild. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (pp. 674-689). ACM.

[3] Mayer, J. R., & Mitchell, J. C. (2012, May). Third-party web tracking: Policy and technology. In Security and Privacy (SP), 2012 IEEE Symposium on (pp. 413-427). IEEE.

[4] Nikiforakis, N., Kapravelos, A., Joosen, W., Kruegel, C., Piessens, F., & Vigna, G. (2013, May). Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In Security and privacy (SP), 2013 IEEE symposium on (pp. 541-555). IEEE.

[5] Boda, K., Földes, Á. M., Gulyás, G. G., & Imre, S. (2011, October). User tracking on the web via cross-browser fingerprinting. In Nordic Conference on Secure IT Systems (pp. 31-46). Springer Berlin Heidelberg.

[6] Acar, G., Juarez, M., Nikiforakis, N., Diaz, C., Gürses, S., Piessens, F., & Preneel, B. (2013, November). FPDetective: dusting the web for fingerprinters. In Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security (pp. 1129-1140). ACM.

[7] Gomer, R., Rodrigues, E. M., Milic-Frayling, N., & Schraefel, M. C. (2013, November). Network analysis of third party tracking: User exposure to tracking cookies through search. In Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 01 (pp. 549-556). IEEE Computer Society.

[8] Bau, J., Mayer, J., Paskov, H., & Mitchell, J. C. (2013). A promising direction for web tracking countermeasures. Proceedings of W2SP.

[9] Kim, D. (2014, May). Poster: Detection and prevention of web-based device fingerprinting. In IEEE Symposium on Security and Privacy (SP).

[10] Schmucker, N. (2011). Web tracking. In SNET2 Seminar Paper-Summer Term.

[11] Broenink, R. (2012, January). Using browser properties for fingerprinting purposes. In 16th biennial Twente Student Conference on IT, Enschede, Holanda.

[12] Chaabane, A., Kaafar, M. A., & Boreli, R. (2012, August). Big friend is watching you: Analyzing online social networks tracking capabilities. In Proceedings of the 2012 ACM workshop on Workshop on online social networks (pp. 7-12). ACM.

[13] Purra, J., & Carlsson, N. (2016, November). Third-party tracking on the web: A Swedish perspective. In Local Computer Networks (LCN), 2016 IEEE 41st Conference on (pp. 28-34). IEEE.

[14] Sanchez-Rola, I., Ugarte-Pedrero, X., Santos, I., & Bringas, P. G. (2016). The web is watching you: A comprehensive review of web-tracking techniques and countermeasures. Logic Journal of IGPL,

jzw041.

[15] Sánchez-Rola, I., Ugarte-Pedrero, X., Santos, I., & Bringas, P. G. (2015). Tracking users like there is no tomorrow: Privacy on the current internet. In International Joint Conference (pp. 473-483). Springer International Publishing.

[16] Bielova, N. (2013). Survey on JavaScript security policies and their enforcement mechanisms in a web browser. The Journal of Logic and Algebraic Programming, 82(8), 243-262.

[17] Englehardt, S., & Narayanan, A. (2016, October). Online tracking: A 1-million-site measurement and analysis. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (pp. 1388-1401). ACM.

[18] Tran, M., Dong, X., Liang, Z., & Jiang, X. (2012, June). Tracking the trackers: Fast and scalable dynamic analysis of web content for privacy violations.In InternationalConference on Applied Cryptography and Network Security (pp. 418-435). Springer Berlin Heidelberg.

[19] Laperdrix, P., Rudametkin, W., & Baudry, B. (2016, May). Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints. In Security and Privacy (SP), 2016 IEEE Symposium on (pp. 878-894). IEEE.

[20] Hoofnagle, C. J., & Good, N. (2012). Web privacy census.

[21] Lerner, A., Simpson, A. K., Kohno, T., & Roesner, F. (2016). Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In 25th USENIX Security Symposium (USENIX Security 16). USENIX Association.

[22] Zarras, A., Kapravelos, A., Stringhini, G., Holz, T., Kruegel, C., & Vigna, G. (2014, November). The dark alleys of madison avenue: Understanding malicious advertisements. In Proceedings of the 2014 Conference on Internet Measurement Conference (pp. 373-380). ACM.

[23] Krishnamurthy, B., Naryshkin, K., & Wills, C. (2011, May). Privacy leakage vs. protection measures: the growing disconnect. In Proceedings of the Web (Vol. 2, pp. 1-10).

[24] Bujlow, T., Carela-Español, V., Solé-Pareta, J., & Barlet-Ros, P. (2015). Web tracking: Mechanisms, implications, and defenses. arXiv preprint arXiv:1507.07872.

[25] Eubank, C., Melara, M., Perez-Botero, D., & Narayanan, A. (2013, May). Shining the floodlights on mobile web tracking-a privacy survey. In Proceedings of the IEEE Workshop on Web (Vol. 2).

[26] Liu, B., Sheth, A., Weinsberg, U., Chandrashekar, J., & Govindan, R. (2013, November). AdReveal: improving transparency into online targeted advertising. In Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks (p. 12). ACM.

[27] Malandrino, D., Petta, A., Scarano, V., Serra, L., Spinelli, R., & Krishnamurthy, B. (2013, November). Privacy awareness about information leakage: Who knows what about me?. In Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society (pp. 279-284). ACM.

[28] Nikiforakis, N., Invernizzi, L., Kapravelos, A., Van Acker, S., Joosen, W., Kruegel, C., ... & Vigna, G. (2012, October). You are what you include: large-scale evaluation of remote javascript inclusions. In Proceedings of the 2012 ACM conference on Computer and communications security (pp. 736-747). ACM.

[29] Saez-Trumper, D., Liu, Y., Baeza-Yates, R., Krishnamurthy, B., & Mislove, A. (2014, October). Beyond cpm and cpc: Determining the value of users on osns. In Proceedings of the second ACM conference on Online social networks (pp. 161-168). ACM.

[30] Yuan, N. J., Zhang, F., Lian, D., Zheng, K., Yu, S., & Xie, X. (2013, October). We know how you live: exploring the spectrum of urban lifestyles. In Proceedings of the first ACM conference on Online social networks (pp. 3-14). ACM.

[31] Keil, F., Schmidt, D., Burgiss, H., Rian, L., & Rosenfeld, R. (2012). Privoxy-Home Page [online], Boston USA, Privoxy Developers.

[32] Dingledine, R., Mathewson, N., & Syverson, P. (2004). Tor: The second-generation onion router. Naval Research Lab Washington DC.

[33] Eckersley, P. (2010, July). How unique is your web browser? In International Symposium on Privacy Enhancing Technologies Symposium (pp. 1-18). Springer Berlin Heidelberg.

[34] Wills, C. E., & Tatar, C. (2012, October). Understanding what they do with what they know. In Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society (pp. 13-18). ACM.

[35] Weinberg, Z., Chen, E. Y., Jayaraman, P. R., & Jackson, C. (2011, May). I still know what you visited last summer: Leaking browsing history via user interaction and side channel attacks. In Security and Privacy (SP), 2011 IEEE Symposium on (pp. 147-161). IEEE.

[36] Mulazzani, M., Reschl, P., Huber, M., Leithner, M., Schrittwieser, S., Weippl, E., & Wien, F. C. (2013, May). Fast and reliable browser identification with javascript engine fingerprinting. In Web 2.0 Workshop on Security and Privacy (W2SP) (Vol. 5).

[37] Verleg, P., van Eekelen, M. C. J. D., & Vranken, H. P. E. (2014). Cache Cookies: searching for hidden browser storage.

[38] Yen, T. F., Xie, Y., Yu, F., Yu, R. P., & Abadi, M. (2012, February). Host Fingerprinting and Tracking on the Web: Privacy and Security Implications. In NDSS.

[39] Guha, S., Cheng, B., & Francis, P. (2011, March). Privad: Practical privacy in online advertising. In USENIX conference on Networked systems design and implementation (pp. 169-182).

[40] Omidvar, M. A., Mirabi, V. R., & Shokry, N. (2011). Analyzing the impact of visitors on page views with Google analytics. arXiv preprint arXiv:1102.0735.