International Journal of Advance Research in Science and Engineering
Vol. No. 5, Issue No. 07 , July 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A SURVEY ON IMPORTANCE OF USER AWARENESS OF WEB TRACKING

[1]Aswini K Nair, [2]Vidhya S S

[1]Computer Science and Engineering,
Vimal Jyothi Engineering College, Kannur, Kerala, (India)
[2]Asst.Professor, Computer Science and Engineering,
Vimal Jyothi Engineering College, Kannur, Kerala, (India)

ABSTRACT

It is an era of internet. We can say that internet has become part and parcel of our life. It is difficult to say "No" to internet. We can have all kinds of information, purchases, social networking and what not through internet. But have you ever know that you are being tracked or your personal information is being tracked? Most probably "No". There is no question that our personal property and our privacy, are precious to us. Tracking is the process of following something or someone. It goes for gathering data about behaviors of clients, area and exercises. It is for knowing the visitor[0] s interests and behaviors and to provide them with personalized content. This paper surveys about the importance of web tracking. Since awareness has the potential to alert users to the corresponding privacy risks and help them in making informed decisions about feasible counter measures.

I INTRODUCTION

Tracking means it can be an act or the process of following something or someone[1]. It gathers data about client's exercises, areas and practices. By utilizing web following, we can recognize guest's of sites and went by website pages. Thus the visitor's interests and behaviors are known and can provide them with personalized content[2].

Web tracking developed from very modest beginnings[3]. As web has evolved, the use of "third party" sites to serve ads on "first party" sites has continue to grow. Third party tacking refers to the practice by which an entity tracks the user[4]. First party sites is the sites that user's choose to visit. The work of third party sites likewise has developed as they move from contextual to behavioral publicizing. In contextual advertising,ads are served just on the premise of the first party site that a client has gone to. Though in behavioral publicizing, the promoters try to track the conduct of clients crosswise over first gathering destinations to fabricate a "profile of client movement and hobbies". Web advertising so as to follow is typically actualized companies. It is for modifying offers to suit client's inclinations. Tracking is implemented by using various tracking components such as cookies, javascripts(JS), IFrames, and Local shared objects(LSO), and relies on third party trackers.

It is said that web following raises a potential protection concern. It happens when information identified with client's exercises and individual data is uncovered to outsiders without their worry. Numerous clients need data that

they are being followed, or regardless of the fact that they knows it they are not taking efforts to establish safety to shield their own information. So it essential to have mindfulness about web tracking among clients in different levels.

This paper has an aim to improve user's knowledge about web tracking. And thus they can have some more secure type of web browsing. This paper initially discuss about privacy issues, tracking tools and some user choice mechanisms for web tracking. Last not the least discuss about the importance of user awareness.

II BASICS OF WEB TRACKING

We realize that sites can gather data about your PC. For instance: Suppose you are asking for a data. The page demand which is sent to site's server incorporate numerous points of interest. It can have data about careful adaptation of programs, it's setup and screen determination the program running in. Other assembled information incorporate the page you originated from, report asking for, IP address and so on. Despite the fact that IP location is alterable in nature, sites can tell what service provider you are with and what city or locale you are in.

We know whenever there is communication with web server, there is deposit of cookie on computer. There are 2 types of cookies: benign cookies and less benign cookies. Benign cookies simply record information needed to make your web experience better whereas less benign cookies can let websites track your movement around internet and can collect information that you don't want to reveal.

The information collected by tracker does not include your name and address. In any case, by assembling all gathered from following parts, web servers can successfully unique mark singular PCs and consequently track clients crosswise over web.

Presently question is the reason organizations utilizing our information or what are these organizations utilizing these information for? They are realizing what you are occupied with and encouraging you more data that you are keen on. The consequence is you are looking at the same things and will become increasingly isolated in a "filter bubble".

2.1 Motivations for Web Tracking

- Advertisement companies actively collect information about users[5] and accumulate it in user profiles. These profiles are used to chose the content of relevance instead of showing random ads to users.
- Usability tests for web applications so that usability problems can be fixed.
- Web analytics focus mainly on performance of web site. Web analytics is the measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage.
- Law enforcement and intelligence agencies use web tracking to solve crimes. It is also important in the fight against identity theft and for prevention of credit card fraud.

III PRIVACY ISSUES IN WEB TRACKING

This section explains about the privacy problems in web tracking. Whenever we have an access on our PCs, smartphones, and tablets,it turns into a data point that trackers can easily collect and share. Sometimes our private data may be tracked by these trackers leads to many privacy problems. Some of the privacy issues are:
these basic techniques to achieve accurate segmentation results.

3.1 Private Data Leakage

Web browsing history is thought to be more identified with individual data. The pages a client visits can give data about her location ,interests, purchases, sexual orientation, financial challenges ,medical conditions, and more.[6] Examining individual page burdens is frequently satisfactory to reach numerous determinations around a user; analyzing examples of movement permits yet more inferences.

When a first-party page embeds third-party content, the third-party website is ordinarily made aware of the URL of the first-party page through an HTTP referrer or equivalent[6].A while later it is found that the free internet dating site Ok Cupid was sending to the information supplier Lotame how frequently a client drinks, smokes,and does drugs. At the point when Krishnamurthy tried [7]search questions on ten famous health sites, they discovered an outsider educated of the clients inquiry on nine of them.

3.2 Policy Views

It is critical that purchasers ought to have some level of control over web following. But the control must be given in which all areas is seemed to be a problem for both stakeholders and consumers Policy views on third-party web tracking vary substantially. EU policymakers believe no tracking should be the default; advertising trade groups have argued tracking should be the default.

3.3 Identifiability

As indicated by an investigation of US Census information, the attribute set Birth date, Gender, Zip represents a danger of individual recognizable proof in light of the fact that these traits can interestingly distinguish 87 percent of the US population[8]. A web searching history is frequently by and by recognized or identifiable. Narayanan as of late proposed in an article around five cement ways[9] in which your character can be joined to information that was at first gathered without distinguishing data. They are:

1) The third party is also a first party: The third party may be a first party in another context, where the user voluntarily provided her identity.

2) A first party sells the users identity: Some first-party websites voluntarily provide (leak) a users identity to third parties for pay.

3) A first party unintentionally provides identity: If a web-site puts identifying information in a URL or page title, it may unintentionally leak the information to third parties. In a 2011 paper, Krishnamurthy [7] examined sign

up and interaction with 120 popular sites for information leakage to third parties. They reported that an aggregate of 48% leaked a user identifier in a Request-URI or referrer.

4) The third party uses a security exploit: A third party may exploit a cross-site security vulnerability on a first-party website to learn the users identity. Cross site Scripting (also known as XSS or CSS) is generally believed to be one of the most common application layer hacking techniques[10].

5) Re-identification: The third party could match pseudonymous browsing histories against identified datasets to re-identify them.

3.4 Harmful Web Tracking Scenarios

It is imperative to concentrate on four variables, when there is a look on harmful First, an actor that causes harm to a consumer. The actor might, for instance, be an malicious employee, hacker, or government agency. Second, a means of access that enables the actor to use tracking data. The information may be willfully exchanged, sold, stolen, lost, or incidentally dispersed. Third, an action that harms the consumer. The action could be, for example, publication, a less favorable offer, or termination of employment. Last, a particular harm that is inflicted. The harm might be physical, psychological, or economic. The endless mixes of these variables result in incalculable conceivable terrible results for consumers.

IV Tracking Components

This section explains about various tracking technologies. Tracking is done by advertising companies to build detailed profiles for pinpoint targeting. On the off chance that you have ever gone to a shopping site and seen promotions for that shopping on different sites later, you have seen it in real life. This area clarifies about the following advancements:

4.1 HTTP Referrer

When you click a link,[11] your browser stacks the page you clicked and tells the site where you originated from. For example, if you clicked a link to an outside website on How-To Geek, the outside website would see the address of the How-To-Geek article you came from. This information is contained in the HTTP referrer Header. The HTTP referrer is also sent when loading content on a web page. For example, if a web page includes an ad or tracking script, your browsers tells the advertiser or tracking network what page you are viewing.

4.2 JAVA Scripts

Numerous sites contain executable Javascript documents that are downloaded by going to users[12]. These files some-times update first-party cookies and send information. Javascript projects have constrained access to client information. However, they can access information stored in the browser including cached objects and the history of visited links.

4.3 IP Addresses

The most essential method for recognizing you is by your IP address.[13] Your IP location distinguishes you on the Internet. Nowadays, its probable that your PC offers an IP address with the other arranged gadgets in your home or office. From your IP address, a site can focus your harsh land area not down to road level, but rather for the most part your city or region. In the event that you have ever seen a spammy ad that tries to look legitimate by saying your area, this is the manner by which the ad does it. IP addresses can change and are often used by multiple users, so they aren't a good way of tracking a single user over time. Still, an IP address can be combined with other techniques here to track your geographical location.

4.4 Browser Fingerprinting

Browsers are really one of a kind. Websites can focus on your operating system, browser version, installed plug-ins and their versions, your operating systems screen resolution, your installed fonts, your time zone, and other information.[13]If you have disabled cookies entirely, that is another piece of data that makes your browser unique. The Electronic Frontier Foundations Panopticlick website is an example of how this information can be used. Only one in 1.1 million people have the same browser configuration. Browser fingerprinting is a powerful tool for tracking users along with IP addresses, cookies and supercookies. This type of tracking, called stateless or passive tracking, is problematic since it is hard to detect.[12]

4.5 Cookies

Cookies are little bits of data sites can store in your browser. They have many legitimate uses. For instance, when you sign into your online banking website, a cookie remembers your login information. When you change a setting on a website, a cookie stores that setting so it can persist across page loads and sessions. While third-party cookies also have legitimate uses, they are often used by advertising networks to track you across multiple websites. If two different websites use the same advertising or tracking network, your browsing history across both sites could be tracked and linked.

Scripts from social networks can also function as tracking scripts. For example, if you are signed into Facebook and you visit a website that contains a Facebook Like button, Facebook knows you visited that website. Facebook stores a cookie to save your login state, so the Like button knows about it.

4.6 Supercookies

You can clear your browser's cookies. However, clearing your cookies isn't fundamentally an answer since these days super cookies are progressively basic. One such super cookie is evercookie. Super cookie solutions like evercookie store cookie data in multiple places for example, in Flash cookies, Silverlight storage, your browsing history, and HTML5 local storage. At the point when a website sees that you have erased piece of the super cookie, the data is repopulated from the other area. For instance, you may clear your browser cookies and not your Flash

cookies, so the website will duplicate the estimation of the Flash cookie to your browser cookies. Super cookies are exceptionally strong.

It is said that ”Super cookies” track you, even in privacy mode[14]. For example, let’s say you use a regular browser to shop on Amazon and use Facebook. Then you launch privacy mode to visit a website that deserves more discretion, like a controversial blog. Supercookies include flash cookies and evercookies:

1) Flash cookies: : A standout amongst the most unmistakable supercookies is the alleged Flash cookie. Flash cookies are broadly utilized by prominent sites. It is[12] a sort of cookie kept up by the Adobe Flash plugin for the benefit of Flash applications implanted in web pages. Since these cookie files are put away outside of the browsers control, web browsers did not customarily give an interface to view, oversee and erase these cookies. Specifically, clients are not informed when such cookies are set, and these cookies never terminate.

2) Evercookies: Evercookies produces extremely persistent cookies in a browser[13]. It is a javascript API available. Its goal is to identify a client even after they have removed standard cookies, flash cookies and others. When creating new cookie, it uses following storage mechanisms when available: Standard HTTP cookies, Local shared objects(Flash cookies), Silverlight isolated storage etc.

V USER CHOICE MECHANISMS

Some of the user choice mechanisms that helps the users to have a control over web tracking are given in this section:

5.1 OPT-OUT Cookies

Opt out cookies are utilized by online organizations to recall clients inclinations about the gathering and utilization of information for online behavioral advertising[15]. When a client picks not to get online behavioral advertising from specific organizations on the consumer Opt-Out page, those organizations put an opt-out cookie in the clients program to advise the company not to convey such advertising in future.

Client decision in current online advertising self-regulation is executed with opt-out cookies. There are a few issues with this methodology. To begin with, it requires manual updating. To opt out of new third parties, a client needs to put in new cookies. Second, cookies lapse, so a client needs to occasionally reestablish opt-out cookies. Third, clients may clear their cookies, unintentionally uprooting their opt-out preferences. Fourth, opt-out cookies are delicate; it is simple for a third gathering to shamefully set or erase an opt-out treat. Fifth, opt out cookies scale inadequately.

5.2 Private Browsing Mode

All significant browser vendors now incorporate private browsing modes into their browsers, under different names. In Safari and Firefox, this element is called Private Browsing, in Chrome Incognito mode, in Opera Private Tab/Window and in Internet Explorer InPrivate Browsing. In private mode, typically cookies and other browser persistence mechanisms are disabled. These modes have two goals[16]. As a matter of first importance, sites went to

while browsing in private mode ought to leave no follow on the clients computer. Second, clients must have security from a web attacker. If you are in PB mode then any feature that requires storing data to a file on the hard drive is disabled.[17] All cookies created while in PB mode are session cookies that expire if you end that PB mode session. The known issue is that If you are in PB mode then some menu items and other features will be disabled or not available.

5.3 Do Not Track

In September 2011, the W3C contracted the Tracking Protection Working Group. The group is working on a Do Not Track (DNT) standard[3]. Every single browsers have submitted themselves to actualize the standard (and most have as of now so executed the HTTP header), however there remains, amongst those partners who will respect the DNT:1 request, an open exchange on parts of the deliberate standard. A few partners have shown they won't respect the DNT flag for different reasons.

Microsoft has announced that its Internet Explorer[18] web browsers will no longer have the Do Not Track setting enabled by default. This appears like a noteworthy hit to online privacy, however it presumably wont have a lot of an impact on a great many people who browse the Web. Web companies aren't the main gatherings overlooking Do Not Track. Given the willful way of Do Not Track, it most likely doesn't matter[19] all that much whether you turn it on or off at any rate.

5.4 Browser-Based Blocking And Plugins

Most browsers give approaches to keep users from being tracked,[5] either straightforwardly or by means of browser extensions. In any case, end users should be suspicious when putting in new browser extensions from obscure sources, as there have been cases of malicious extensions social affair scanning information without client assent.

1) Tracking Protection Lists: Microsoft Internet Explorer 9 has out of the box support for Tracking Protection Lists[21]. These lists define domains from which content is only fetched if entered into the address bar or directly clicked on by the user, thus preventing cross-browser tracking from third-party domains.

2) Torbutton: Torbutton is a Firefox extension which lets users enable or disable Tor in the browser with one click. It provides additional privacy enhancing functionality.

3) Adblock Plus: Adblock Plus is a browser extension which underpins Mozilla Firefox and Google Chrome. It's fundamental usefulness is blacklist-based covering up of advertisements. The user chooses from a number of subscription lists, which are regular expression based filters that stop unwanted content embedded into web pages from being downloaded.[24] The most visible benefit to the user is the reduced number of ads. But the problem is this can be used to steal your password which is even worse than stealing your token.

4) NoScript: NoScript is a browser extension accessible for Mozilla Firefox that specifically blocks JavaScript, Java, Silverlight, Flash and other executable content. While NoScript is chiefly utilized for security reasons, it additionally incapacitates web tracking administrations that depend on active client-side content. Since many

modern web sites use scripts and plug-ins for legitimate reasons, the whitelist-based approach implemented by NoScript has usability drawbacks and requires frequent user intervention.

5) NoTrace: NoTrace is another add-on for the Firefox web browser that provides you with protection against privacy threats on the Internet. Unlike many other extensions, it also attempts to raise awareness of issues.

6) Ghostery: Ghostery is a proprietary freeware[22] privacy-related browser extension for Mozilla Firefox, Google Chrome, Internet Explorer, Opera, and Apple Safari owned by the advertising and privacy technology company Ghostery, Inc.(formerly Evidon). It enables its users to easily detect and control web bugs, which are objects embedded in a web page, invisible to the user, which allow collection of the user's browsing habits. Ghostery also has a privacy team that creates profiles of page elements and companies for educational purposes.

VI USER AWARENESS ABOUT WEB TRACKING

There is no doubt that our own property and our protection, are valuable to us. Consistently clients are liable to confront different online security hazard because of open nature of the Internet. With the assistance of web tracking components, organizations gather data about clients. By and large, accumulation, handling and spread of personal[23] data can raise genuine protection issues among users when they go online for a variety of daily activities such as online banking, business transactions, online shopping, social network interactions and so on.

The facebook company has officially[20] pronounced that it will be tracking your online utilization and conduct on Facebook, and off Facebook as well. Like Google, Facebook needs to comprehend client conduct to serve more focused on promotions. Furthermore, for that Facebook has been utilizing your likes and posts on Facebook. Facebook will track you over any website that you visit, if that web page has a "like" button, a Facebook login, or any Facebook code. Also, Facebook will track your reading habits, regardless of the fact that you don't login by means of Facebook or click on "like" button on the site.

Most users of Facebook are unknown about this and they blindly post their personal data in facebook. Facebook has managed this via an automatic opt-in feature albeit one which you can opt out of. However, activists and analysts point out that companies like Facebook, Google etc should only be allowed to use your data by turning it on through your explicit permission (manual opt-in instead of automatic opt-in) and not by having to turn it off through an opt-out. For advertisers, Facebook is aiming to build a bank of data to help target advertising better. The raw data is processed by big data brokers like Acxiom and Datalogix to develop profiles of users, which are then made available to advertisers.

It is based on the logic that, if we are interested in a particular things, then we will have a look more on that things. That logic is being used by Facebook. Suppose you regularly surf sites that discuss fashion and accessories. You will be profiled as a fashion geek, and advertisers interested in fashion geeks will be able to serve ads to you on Facebook. This can be big business for Facebook. Google earned as much as 13 billion dollar off of data-based targeting. Facebook is looking to take a larger piece of this growing market. For consumers, this is the hidden cost of Facebook: your data. Facebook doesn't charge users because it makes money by selling access to demographics of users. This often results in advertising that seems to "follow you around," which can feel intrusive.

The problem is that users are not asked if they consent to being tracked, ie, there is an automatic opt-in process, instead of this being a manual opt-in process where users can choose to allow this to happen or not. Instead, all that Facebook gives users, is the option to 'opt-out' but the problem is that experts know that less than 2% of users actively go through the opt-out process.

Researchers caution that advances in online tracking have made it difficult even for sophisticated computer clients to secure their protection and call for further regulatory intervention. In a research paper, computer security experts from Princeton University and KU Leuven University in Belgium describe three recently developed online tracking mechanisms that can be used to track and potentially identify users across different websites without their knowledge or consent. These technologies canvas fingerprinting, evercookies, and cookie syncing represent what the researchers characterize as an ongoing arms race against privacy. Built using recently developed Web APIs, these tracking techniques are designed to be less susceptible to erasure and blocking than traditional HTTP cookies, which can be cleared and avoided through browser controls.

We cannot say that tracking is completely a bad thing. It helps us in various things that we already discussed in the section of this paper. Online advertising companies want to understand consumer behavior online and they gain this understanding by building interest profiles based on the websites individuals visit. But when people clear the cookie files that websites place on their computers or block them, advertisers may be left in the dark about who is seeing their ads.

To preclude this possibility which makes advertising less effective and less profitable online advertising companies have been experimenting with more reliable ways to get information about website visitors. The researchers say that they found 5% of the top 100,000 websites using canvas fingerprinting. This is a tracking technique that utilizes HTML5's Canvas API to draw an invisible picture in the user's browser window. This picture is then converted into an alphanumeric code so it can serve as a "fingerprint," a unique identifier associated with a specific user. In and of itself, this code does not reveal the user's identity, but identity can often be determined through other means and may end up being associated with other user data.

A single online advertising company, AddThis, is responsible for most of the canvas fingerprinting (95%), according to the paper. Canvas fingerprinting scripts were also found associated with 19 other domains or companies, including Ligatus, a German digital marketing firm, and Pof.com, operated by Canada's PlentyofFish Media.

There are some defenses available, such as Disconnect. But the researchers expect individuals will have problems trying to protect their privacy. "It is doubtful that even privacy-conscious and technologically-savvy users can adopt and maintain the necessary privacy tools without ever experiencing a single misstep,". The researchers conclude by urging standards bodies like the World Wide Web Consortium (W3C) to consider the privacy implications of new Web technology at the design stage. They suggest that a viable approach to online privacy needs to include technical efforts buttressed by regulatory oversight.

Revealing and measuring the many commercial tools that invisibly track web users is a key step toward improving transparency and privacy on the Internet, according to a set of privacy and technology experts. But awareness among

the people must be given. Awareness has the potential to alert users to the corresponding privacy risks and help them in making informed decisions about feasible counter measures. Users must take interest in reading the privacy policies. Not only that there must be awareness programs about the problems of web tracking in earlier stages of school as students are using social networking sites or other sites at their young age itself.

VII CONCLUSION

This paper surveyed about the importance of awareness about web tracking among people. At first it surveyed about various technologies and privacy problems in web tracking. Then it discuss about the various user choice mechanisms in web tracking.

REFERENCES

1) A. Hamed and H.K.-B. Ayed, Privacy scoring and users' awareness for web tracking, Information and Communication Systems (ICICS), 2015 6th International Conference on, April 2015
2) S.Mlot, Parents worried about webfirms tracking their kids, http://www.pcmag.com/article2/0,2817,2412313,00.asp,2012, access date: March 2014
3) Web tracking and privacy:respect for context, transparency and control remains essential, 2012, International Working Group on Data Protection in Telecommunications.
4) T. Kohno F. Roesner and D. Wetherall, Detecting and defending against third party tracking on the web, in proceedings of the 9th usenix conference on net- worked systems design and implementation, 2012.
5) Niklas Schmucker, Web tracking, (2011),J.R. Mayer and J.C. Mitchell, Third-party web tracking: Policy and technology, Security and Privacy (SP), 2012 IEEE Symposium on, May 2012.
6) B. Krishnamurthy and C. Wills, Privacy leakage vs. protection measures: the growing disconnect, in proceedings of the web 2.0 security and privacy workshop, may 2011
7) D. Irani and Webb, Modeling unintended personal-information leakage from multiple online social networks,Internet Computing, IEEE 15 (2011)
8) A. Narayanan, There is no such thing as anonymous online tracking. [online]. available: http: //cyberlaw.stanford.edu/node/6701,
9) Cross-site scripting (xss) attack - http://www.acunetix.com/website security/cross- site-scripting/.,
10) Franziska Roesner, Security and privacy from untrusted applications in modern and emerging client platforms, (2014),
11) Arvind Narayanan, Claude Castelluccia, Privacy considerations of online behavioural tracking, (2012),Enisa.
12) Htg explains: Learn how websites are tracking you online-http://www.howtogeek.com/115483/htg-explains-learn-how-websites-are- tracking-you-online/
13) Jose Pagliery,"super cookies" track you, even in privacy mode 14) http://money.cnn.com/2015/01/09/technology/security/super-cookies/, (2015).

15)Ad choices: The digital advertising alliance of canada (daac) - http://youradchoices.ca/faq.

16) E. Bursztein G. Aggrawal, An analysis of private browsing modes in modern browsers in proceedings of the 19th usenix security symposium, 2010.

17)Issues related to private browsing http://kb.mozillazine.org/

18)Nathaniel Mott,Do not track wont stop tech companies from tracking you, so who cares that microsoft disabled it?, (2015).

19) Microsoft will remove do not track as the default setting in its new browsers http://techcrunch.com/2015/04/03/microsoft-disables-do-not-track-as-the-default-setting-in-internet-explorer/.

20)With 'Like' button, Facebook tracks you across the web: http://timesofindia.indiatimes.com/tech/tech-news/With-Like-button-Facebook-tracks-you-across-the-web/articleshow/36715577.cms

21)Microsoft, internet explorer 9 tracking protection: http://ie.microsoft.com/testdrive/browser/ trackingprotectionlists/.

22) Ghostery : https://en.wikipedia.org/wiki/ghostery.

23) Andrea Petta Del na Malandrino, Privacy awareness about information leakage: Who knows what about me?

24) Is adblock(plus) a security risk? http://security.stackexchange.com/ questions/52361/is- adblock-plus-a-security-risk.