LOVELY PROFESSIONAL UNIVERSITY

A Study on User's Comfort and Preferences For privacy in Mobile Devices

A Dissertation Proposal submitted

By

Kashish Thakur

To

Department of Computer Science and Engineering

In partial fulfillment of the Requirement for the

Award of the Degree of

Master of Technology in Computer Science and Engineering

Under the guidance of

Mr. Varun Singla

(November 2017)

LOVELY PROFESSIONAL UNIVERSITY
Transforming Education, Transforming India

TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering

Program : P172::M.Tech. (Computer Science and Engineering) [Full Time]

COURSE CODE : CSE548 **REGULAR/BACKLOG :** Regular **GROUP NUMBER :** CSERGD0334

Supervisor Name : Varun Singla **UID :** 17705 **Designation :** Assistant Professor

Qualification : M-Tech **Research Experience :** 4 yrs.

SR.NO.	NAME OF STUDENT	REGISTRATION NO	BATCH	SECTION	CONTACT NUMBER
1	Kashish Thakur	11610460	2016	K1637	8427439444

SPECIALIZATION AREA : Database Systems **Supervisor Signature:** Varun Singla 17705. 30-11-17.

PROPOSED TOPIC : web Tracking using telephone metadata

Qualitative Assessment of Proposed Topic by PAC

Sr.No.	Parameter	Rating (out of 10)
1	Project Novelty: Potential of the project to create new knowledge	6.75
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	6.75
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	7.25
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	7.50
5	Social Applicability: Project work intends to solve a practical problem.	7.00
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	7.25

PAC Committee Members

PAC Member 1 Name: Kewal Krishan	UID: 11179	Recommended (Y/N): Yes
PAC Member 2 Name: Raj Karan Singh	UID: 14307	Recommended (Y/N): NA
PAC Member 3 Name: Sawal Tandon	UID: 14770	Recommended (Y/N): NA
PAC Member 4 Name: Dr. Pooja Gupta	UID: 19580	Recommended (Y/N): Yes
PAC Member 5 Name: Kamlesh Lakhwani	UID: 20980	Recommended (Y/N): Yes
PAC Member 6 Name: Dr.Priyanka Chawla	UID: 22046	Recommended (Y/N): NA
DAA Nominee Name: Kuldeep Kumar Kushwaha	UID: 17118	Recommended (Y/N): Yes

Final Topic Approved by PAC: **Web Tracking using Telephone metadata**

Overall Remarks: Approved (with minor changes)

PAC CHAIRPERSON Name: 11024::Amandeep Nagpal **Approval Date:** 04 Nov 2017

ABSTRACT

Web Tracking refers to the method of archiving the existing sites and tracking the changes to those sites over time. It is widely used method of monitoring (using specialized software tools) to keep tabs on the activities of the site visitors. But these activities have raised serious privacy concerns among the users, which in most cases are unaware of the fact that they are being tracked. There is an inherent information irregularity in web privacy for the reasons like lack of adequate revelation of data tracking and collection practices, the data use policies by the web-sites and limitations or unawareness of the users in understanding these tracking procedures and purposes. In this study we will consider the different situation factors that may affect the users' preference for being tracked and depict their comfort in these situations. We will then use this information about comfort of users in different situations and using a classifier to find their comfort levels which would then be depicted is more effective way.

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the M.Tech Dissertation/dissertation proposal entitled "Study on User's Comfort and Preferences For privacy in Mobile Devices", submitted by Kashish Thakur at Lovely Professional University, Phagwara, India is a bonafide record of his / her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Varun Singla

Date:

1) Counter Signed by:

Concerned HOD:

HoD's Signature: ________________

HoD Name: ____________________

Date: ___________________

2) Neutral Examiners:

External Examiner

Signature: _______________

Name: __________________

Affiliation: ______________

Date: ___________________

3) Internal Examiner

Signature: _______________

Name: __________________

Date: ___________________

ACKNOWLEDGEMENT

I would like to thank, **Mr. Varun Singla** my dissertation guide for giving me this opportunity to do research on the topic “A study on User’s Comfort in context to Online Tracking in Mobile Devices “. I am very much thankful to him for his co-operation and his assistance throughout my research work. He has been guiding light and without his guidance this report would not have been possible. I am also gratefully indebted to **LOVELY PROFESSIONAL UNIVERSITY** for giving us such research opportunities time to time so that we can gain more knowledge. By participating in an effort like this I become aware of the degree to which other people supported me in by endeavor support me in cascades from family and friends. This completed research report is as much their achievements as it is mine.

DECLARATION

I hereby declare that the dissertation proposal entitled, **A Study on User's Comfort and Preferences for privacy in Mobile Devices** submitted for the M. Tech Degree is entirely my original work and all ideas and references have been duly acknowledged. It does not contain any work for the award of any other degree or diploma.

Kashish Thakur
Registration No: 11610460

Table of Contents

List of Figures

Chapter 1
INTRODUCTION

Web Tracking refers to the method of archiving the existing sites and tracking the changes to those sites over time. It is widely used method of monitoring (using specialized software tools) to keep tabs on the activities of the site visitors. Usually, the users aren't aware of the fact that they are actually being tracked while using the web. Mostly, the trackers are the providers or the owners of the web-site being visited by the user. The web-site providers may also track their user's activities. They implement the tracking methods or mechanisms to ensure the fundamental features like a shopping basket or customized search results. While there might be some legitimate or real reasons for web-site providers or the owners to track their user's activities on their web-site, the question here arises why the third-parties' trackers, are so much interested in user's behavior. Their main objective is collecting as much user related data as possible. This data is very important to the companies because it presents the user's likes or interests. These interests in turn are the main targets of the advertising organizations, which greatly improves the possibilities that users' clicks on advertisements will simultaneously generate revenue for these companies. Such business processes go along with problems, in particular the users have never agreed to such an information gathering. They don't have any idea what purposes their data is serving and they never approve the processing of their data.

It is normally used to personalize the user's experience online on a website. It is known that the content providers for example Facebook, Google, the service providers like You Tube and some other third parties like Double Click gather a huge chunks of user's personal information from the users when they are browsing through the web. This personalized information captured via browsing histories of users provides general data about financial and health condition, educational information, shopping information and other user preferences. The data can also be used by the government agencies for the purpose of National Security to identify and prevent frauds. Some associated programs like Pay per-per-scale demands tracking to follow the user and can be done through mechanisms like cookie -syncing between the sites where the ads are placed on the sites where the actual purchase is made.

When visiting a web-site, a user receives everything that is relevant for displaying the web-site effectively, for example the image, texts, Java-Script files or the style sheets. Still, there can also be things in the response which might belong to the trackers and are related to Web-Tracking. Those parts may have pictures, the tracking-pixels or any similar things that can be requested from the tracker's server. This request needs the establishment of a connection to the site's tracker and contains referrer-header that contains the currently visited site. By observing this header field, the tracker is able to identify which web-site the user has been visiting. If the site's tracker has the ability to track a user, he is also able to develop profiles for that very user.

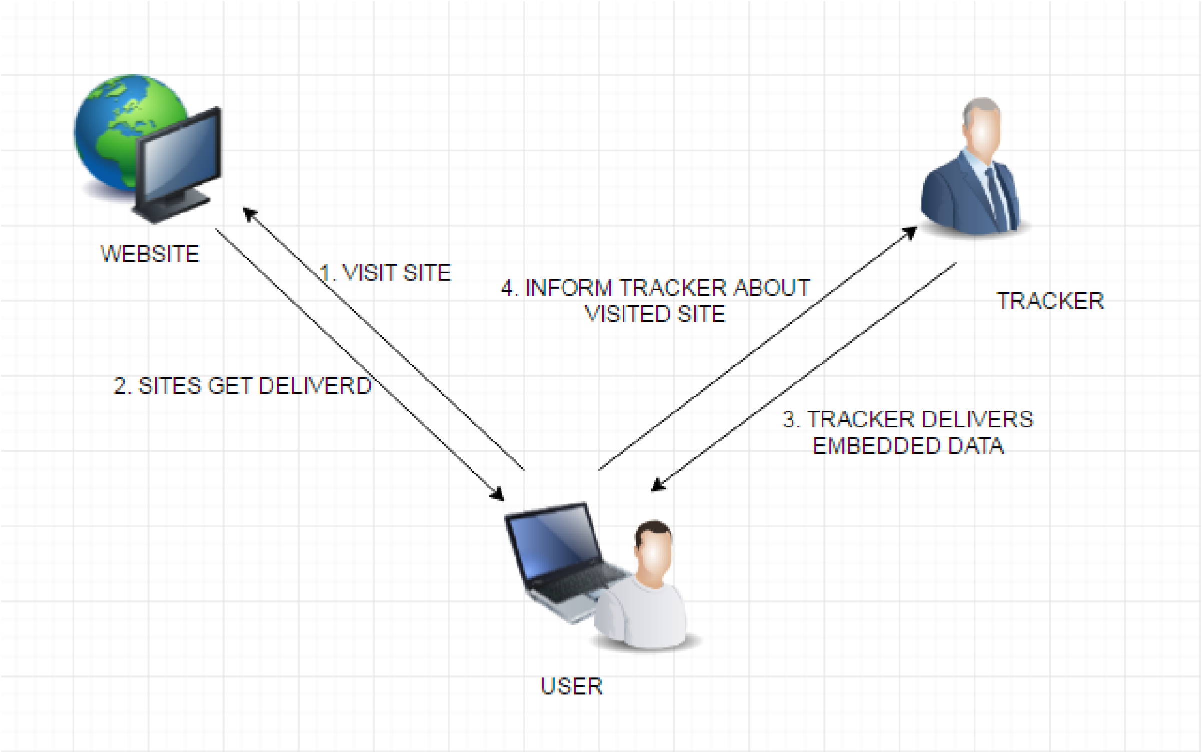
WEBSITE
1. VISIT SITE
4. INFORM TRACKER ABOUT VISITED SITE
TRACKER
2. SITES GET DELIVERD
3. TRACKER DELIVERS EMBEDDED DATA
USER

Figure 1 Basic functionality of web tracking

User's Information on the web can either be made available voluntarily for example while filling the online forms provided on the web or it can be retrieved indirectly without the knowledge of user with the help of different third-party trackers present on the web. This is done through the examining of the HTTP request, IP Headers, or queries placed in the search engines like Google, or can be done through the Java Scripts or Flash programs which are embedded in the web pages. From the collected and analyzed data, we can find information about the browsers; operating system is there, IP address or underlying hardware. We can even retrieve personal information such as the geo-locations of the users, personal preferences and history of the visits on the web page.

1.1 Motivation behind Tracking

There are different reasons for online tracking:

i. First-Part Tracking
ii. Third-Party Tracking

The First- Party Tracking is carried out by the website's owners to provide a customized or unique user experience on the web such as maintenance of the user's shopping activities, personalized ads. The First-Party trackers usually employ Third party trackers for tracking of the user activities.

Third-party services provide an enormous worth to the web helping the first-party sites to implement advertising and their analytics. They also have raised privacy concerns in the recent years among researchers, the civil society organizations or the policymakers to the greater use of third-party sites analyzing their user's browsing task across unconnected first-party trackers.

One of the motivations behind tracking is developing profiles of users. The user's profiles are very useful for many organizations for customizing the user's services which suit their customer's demands. This in turn increases their revenue. The fine motive of this targeting which is also called behavioral profiling is to track users constantly on the web. They construct profiles of users capturing their interest and the characteristics like gender, or their age or which ethnicity they belong to. This also includes other activities like the online shopping tasks. Like the publishing or ad Companies uses behavioral targets to display the ads that jointly reflect the user's interest. Any Online Ads Organization is basically composed of the three main components:

i. The promoter
ii. The maker
iii. The ad connections

The promoter is the component like a product's owner who wants to promote his products and services on the web. The maker is the component like a website's owner or maker which has the possession of one or the several sites and is able to display ads and be paid for it.

Finally, the ad connection is the component that captures ads and promotes and provides them on the maker's sites. If user's clicks on any ad, then the ad connections retrieve the payment from the related promoter. Therefore, there is a strong motivation for the ad 's connection to develop a very precise and proper outlines which can increase revenue.

E-Commerce websites like "Amazon" also use behavioral tracking to recommend users based on their interest, the individual's past behavior (customized searches), on the basis of past behavior of similar users (the social advice) and on the searched products (i.e. items suggestions).

1.2 Privacy Issues

It can be argued that customization of the users while profiling them is usually favorable. With the help of profiles, they receive facts that are relevant to their preference and are related to their search. This is however creating grave privacy concerns as it permits the organizations to concentrate large chunks of information about their users and their web users.

The concern know is that when we are using the web all our activities are monitored and observed and are often correlated. Some of the organizations offer many favors that collect different varieties of information from the users. The collection of this whole information provides a very strong mechanism to precisely develop the profile of the users. Like we have "Google" which being the main third-party tracker and analyzes the users on most of the web-sites.

In addition to that it also manages the most used search engine which stores online histories of the users through their searches, their map search i.e. on the Google map service. Web searches have often been shown to be sensitive. It has actually been shown that it is easy to get the identity of a user from analyzing his web history. The map requests often reveal a lot of the user's information, such as their house address or the favorite locations they like to visit. Finally, Google has the most widely used email tool which is the Gmail, and therefore, has the access to emails of large number of users. By combining these different types of information coming from different sources, Google is able to develop very precise profiles of their customers. Profiling often moves the balance of influence between those that own the sites (i.e. mostly the large organizations) and those that are being tracked (i.e. online users), because the trackers have knowledge to which the users have no idea.

1.2.1 Privacy Issues related to Mobile Devices

The Mobile devices privacy has attracted attraction of many researchers. This is because of the fact that the mobile devices are particularly captivating markets for most of the companies and advertisers as they are belonging to a single user. They provide the signs like location which can lead to better ads.

With the arrival of “Ubiquitous advertising”, which is the application of computational advertising to mobile phones and provides even more sources of user’s information This is the reason why the ads will not only be personalized to users’ online, but also to their profiles. The Advertisements are customized to users’ locations, intellectual activities and possibly their moods. In contrast to a normal PC, a mobile device is usually possessed by an individual person, a more detailed and precise profiles. It is also considered that there will be use of sensors on the mobile phones that will be able to know even the users’ food preferences.

New innovations have created serious privacy issues that need be studied more properly. The mobile phones now-a-days have increasing capacities and are fully furnished with multiple facilities like cameras, accelerometers, GPS. The geo-located systems already enable individual user and groups to gather and share various kinds of data. Next the concept of Urban sensing, which is a sensing paradigm that is borrowing users as part of a sensing architecture. It’s the future scope that this concept of urban sensing will be able to provide extra information about users. Most users are unaware that their information that is being collected is extra information, especially in case of participatory sensing. For example, a picture captured by a user may uncover additional context based info examined from the background of any related text. Most of the people are unaware of the fact that the pictures and video clippings that are taken with their smart phones or cameras contain their geo-location info. This info can also be used to localize them during traveling or even uncover their house address.

1.3 User’s Comfort with Tracking

With online tracking of users being so widespread, it is also observed that most of the users are unaware of the fact that they are being tracked. The users comfort and preference with tracking is usually very complex but is based on different situational factors. In the previous study, it is observed that majority of the users are comfortable with being tracked if done for the motive of personalized search outcomes but that also has a condition which says that information used is not sensitive. Most of the users are comfortable with tracking if they are given preliminary privilege to select which type of facts will be shared and also the capability to restrict all the first and the third party trackers from collecting their info.

It also depends on various other situational factors like:

i. Trust on the site
ii. Frequency of visit to the site
iii. General attitudes and opinions
iv. Personalized Search Results
v. Ability to restrict the first and the Third party trackers
vi. Ability to select the type of shared info
vii. Browsing scenarios

There are present methods like Do Not Track, Opt-out Cookies which allows users to prevent the trackers from tracking their activities. But it is found that these methods allow or they deactivate the web tracking or the targeted ads on per-session basis. The Opt-out options often only stop the display of targeted ads rather than the prevention from being tracked.

When we consider the mobile devices we have some ads on factors which may affect the user’s comfort level in tracking. Starting with the cloud, most of the mobile data these days in saved in cloud which has major privacy issues is it shared pool of memory. Moreover, there is no way of questioning this privacy concern as when we consider the mobile devices unlike PC they don’t have the Fourth Amendment Protection. So the users will not know where their cloud data is being used or exploited. Next coming to the Location Privacy issue, it is very common with the mobile devices. Most of our mobile devices are having the geo-location systems which constantly track our location and hence can reveal our current location or even our house address. So the users in the same way do not have any idea of such situations.

Next comes the factor which is “Data never forgets a face”. When the users upload their photos on the web like on Facebook, here usually pictures are posted and tagged, but the users are unaware of the fact that Facebook uses tags associated with theses photos to develop an ever more detailed “face-prints” which shows what the user look like from different angle. These face-prints are in turn being extracted from the web through tracking and used for different purposes like in criminal cases, citizenship applications or criminal cases.

So if we consider these factors their can be different positive and negative aspects in tracking and the users comfort in tracking may vary based on these factors and capturing these comfort on different factors is itself very tedious task to do.

Chapter 2

LITERATURE SURVEY

Jonathan R. Mayers et.al (2012) found that with online web tracking being very beneficial to most of the first party owners and the third parties. There also exists a privacy cost which is gaining attention among the researchers. There is a Fourth Party Web measurement tool implemented as an extension of Firefox. The data analysis using fourth party measurement tool is fast and the results generated in Python. This is the improved version of previous measurement approach. All the Fourth-Party resources are available in http://fourthparty.info. *The* debates on current policy and technology related to third party has been explained. In the policy part, a discussion on the reasons for the privacy concerns in third party tracking and the ways of making better policy for minimizing those concerns. In the technology part a survey is conducted on two different tracking technologies that is stateless and stateful technology is done. An analysis is done of other technologies which provide third party services with lesser privacy concerns. Lastly, a review is carried out on the user's preferences and self-help technologies like cookies, the blocking and Do Not Track has been done.[1]

Arvind Narayan et.al (2016) performed a full measurement of web tracking was performed based on web capture on one million sites. For each site, 15 types of measurements were done which included cookie based (stateful) and fingerprinting based (stateless) tracking browser privacy tools effects and cookie-syncing. A new web privacy measurement platform is implemented called Open WPM that used automated version of full fledged browser. It's different from FP detective from the fact that it conducts stateful measurements and is built using Python and its libraries. In this it was observed that online tracking has "Long Tail ". For fully covering all the sites they have considered three main points like network proxy, browser extension and disk state monitor. They have quantified the influence of tracker and the third-party tracker on HTTP implementation and have depicted that cookie syncing is prevalent. They have Selenium tool a cross platform web driver for Google Chrome, Mozilla Firefox and Internet Explorer is used. For supporting running measurements, the extension like Ghostery, HTTPS Everywhere and Firefox privacy settings (like cookie blocking) are enabled. A ranking metrics for the Third-Party called Prominence is taken which is given by the following formula. Prominence (t) = Summation (edge (s, t) =1/ rank(s), where edge (s, t) => whether third party is there on site s. This metric measures the prevalence with which an average user which is browsing according to the power law will meet a third party. The ranking of third metric is based on how frequently a user encounter a third party tracker embedded in site.[2]

Christian Eubank et.al (2013) conducted a study on mobile web tracking which is presented for which five different physical and emulated devices with a desktop device are taken for comparing tracking on each. The crawler is based on Fourth Party, web measurement platform. A mobile web privacy measurement tool is implemented which is based on Fourth Party platform. A crawl on the Alexa top 500 sites is carried out on single personal computer and five mobile devices which includes a Smartphone, two tablets, an emulated tablet and an emulated Smartphone and all of them are running on Android. The data collected is used for comparison with desktop tracking. A survey on growing cookies is conducted. JavaScript is the approach which is used by tracking domains for mobile tracking. This gradually increases with more powerful devices. An analysis of cookie longevity by a device is conducted in which expiry length of a cookie tells about the intention of the third party. It is observed that emulated and physical phones have greater expiry length than the desktops. Lastly, they have examined a intriguing processes of cookies which is used to keep the user browsing history on the client side.[3]

Joshua Tan et.al (2016) found that the Third-party tracking is used to compile user profiles for either for selling information, customer targeting ads or customizing sites. A study is carried out on the possible benefits and risk of tracking and why the tracking is done. They have gathered the browsing histories and conducted interviews of 35 people. A semi structured interview is conducted in which they have observed the participants’ feeling and concerns about tracking in actual browsing scenario. It is based on the user’s experience in the past two weeks. The different situational factors that affect the user’s preferences are identified. For getting the user’s browsing history a web browser plug-in is used. It is founded that the user’s preference depends on following factors like how frequently they visit a site and being uncomfortable with invisible results (like price discrimination, revenue of site etc), user’s trust on a site and more comfort if given advance privilege mechanism to decide type of information being shared. A fuller view was collected regarding the factors which affect the comfort and since it was based on real browsing histories so study is considered more trustworthy. Different classifier was used (like Asymmetric AdaBoost, Support Vector Machine (SVM)) to predict the user's willingness with tracking of page visits based on properties of page visits and user's general conceptions. The data (285 situations) is split into test sets then a classifier on training sets calculates the accuracy of predicting comfort on test sets. With the help of observed results, a design guideline for the future tools is identified.[4]

Mark Juarez et.al (2014) conducted a study on three advance web tracking methods that is Canvas fingerprinting, ever cookies and cookie syncing in conjunction with ever cookie is done. It is found that about 5% of the top 100,000 sites make use of fingerprinting technology. They presented the study ever cookie and respawning in which they have considered like the detection on persistent user ids by leveraging data from two simultaneous crawls on separate machines and applied a rule set to find which elements are identified, Flash cookies respawning HTTP cookie in this we sequentially crawl Alexa top 100,000 sites ,out of them two run with the flash cookies loaded from the sequential crawls and the third runs on a separate machine without any data being loaded from previous crawl. A new ever cookie vector is discovered which is called Indexed. For the cookie-syncing some novel techniques are used like Strace debugging tool for low level monitoring of the browser and the Flash plug-in player, set of criteria for distinguishing and extracting pseudonymous identifiers from the old storage vectors like cookies. This is done for the detection and analysis of ID flows and quantification of privacy-intrusive tricking due to cookie syncing is done. observation is made that even the most experienced users encounter great problem in avoiding the tracking mechanisms in which they found that even single delay in judgment can destroy the privacy defenses considered for the websites.[5]

Monica Chew et.al (2014) found a new privacy technology for Mozilla Firefox called Tracking Protection is presented. It is used to reduce the tracking of user' web activity by blocking the request received by the tracking domains. They demonstrated 67.5 % depletion in the amount of the HTTP cookies decided during the crawl of Alexa peak 200 news sites. The depletion is brought by blocking 11 tracking elements on 50% of the websites. The technology is an API which is based on Google Safe Browsing which is a method for well planned URL-based blocking list which updates and perform lookups. A subset of approx

1500 domains from Disconnects privacy-oriented blocking list is used to find all the unsafe regions and the block list is updated after every 45 minutes to reduce the effects of wrong block list entries. Firefox Nightly is instrumented using the Mozmil tool to visit top 200 news sites. For each website's loading time and data usage is measured with or without Tracking Protection enabled and the reduction is calculated. It has the performance benefits of 44% median depletion in the load time and 39 % reduction in data usage. The overhead elimination brought by the Tracking protection is by decreasing third party tracking rather than the first-party data on each page.[6]

Franziska Roesenerm et.al (2012) conducted a study and found that to better understand the third-party tracking a Client-side mechanism is developed for noticing and clustering five types of third party employees based on how they exploit the browser state. The detection system runs while we are using the web and bout 500 unique trackers are there in the measurement. Most of the trackers have a combination of tracking behavior. For observing the tracker's behavior classification is done based on five different behaviors. It is estimated that on each collects more than 20% of the user's browsing ways based on their web search traces collected from the AOL data. The main aim is to observe the tracking as noticed by the users and this is done by combining web tracking analyzing directly into the web browser. A Firefox extension is developed to calculate the prevalence of the web trackers and their behavior. It is observed that using popup blocking, the third-party method of cookie blocking and the Do Not Track option a major portion of cookie based tracking is stopped without any effect on the functionality of the browser. A new extension called ShareMeNot defends against the third-party tracking while the users can continue to interact with the widgets.[7] .

Chapter 3

OBJECTIVE OF STUDY

The objective of the study is to address the current privacy concerns that are being raised on the web tracking activities in the mobile devices. This study considers the user’s perspective on this issue. We will first try to identify the different situations or scenarios in mobile devices for which tracking is done or is used. We will then try to evaluate the users comfort level in this situation using classifier.

The research is focused on the following objectives:

i. Identification of key situations factors for which web tracking is done in mobile devices

ii. Collecting user’s information using Surveys

iii. Using a hybrid classifier to display the observed results

iv. Evaluating the user’s behavior on each of the situation and calculating the accuracy of the observed data

In today’s world where web usage and browsing is so regular we should also be aware of how our browsing activities are being used. The user’s privacy should be the main objective of the organizations, for which they should know what the user’s comfort in different situations. They should make necessary changes for avoiding this privacy breach.

Chapter 4

SCOPE OF STUDY

Web Tracking is a method of collecting large chunks of personal info from customers when browsing the web. This huge gathering and observation of personal data becomes the main business of majority of these companies, which use it for business purposes. It is very common now-a days and is being used by both First Party (the site Owners) and third party owners (Double Click) the reasons for such activities could be different like precise targeting, price discrimination, generating revenues or can be just developing user behavioral profiles. But these activities have raised serious privacy concerns among the users, which in most cases are unaware of the fact that they are being tracked. There is an important data in the irregularity in web privacy for the reasons like lack of appropriate revelation of tracking and gathering practices, the data usage policies by the web-sites and limitations of the users in understanding these tracking procedures and purposes.

In this study we will consider the different situation factors that may affect the users' preference for being tracked and depict their comfort in these situations. We will then use this information about comfort of users in different situations and using a classifier to find their comfort levels which would then be depicted is more effective way.

Chapter 5
RESEARCH METHODOLOGY

After the review of all previous approaches for the concept of web tracking we have found that we need to address the privacy concern in this context to a most precise level. So we have taken into consideration the mobile devices ass they belong to individuals rather than a group. So the study will be conducted in this context.

5.1 Identification of different Situations

For the identification of different factors, we have considered first the context of the study that is mobile devices. Next we start with the consideration of different scenarios related to tracking which may raise concern among the users or can be beneficial for them.

Starting with the general attitudes and conceptions of the users regarding their activities being tracked on the web, this might also include the misconceptions that may have about tracking. Next we consider the user perception of outcomes that include two main things:

i. User Noticeable outcomes that includes customization of sites, targeted ads and the legal harms
ii. Invisible outcomes that includes the company's revenue, price discrimination and data linked to user's identity

Then we move on to the next perspective which is factors affecting the preference of tracking. User while considering the tracking usually have different preferences to different factors and these preferences are often very complex as the preference of one user might be very contrasting in relation to other user while considering same factor.

The factors related to information being tracked have two cases:

i. Tracking that contain information properties like user's personal information, social information, search information or has correspondence (disclosure of private info)
ii. Tracking that has non-informational properties like trust, sites frequently visited or infrequently visited, lack of consent and awareness and sharing with 1st parties

Next we take into account the situation which comes into action with mobile devices. This includes the following:

i. Presence or storage of user data on cloud
ii. Location Tracking
iii. Face-printing of uploaded pictures

The data present in the mobile phones is stored on cloud, it might look easy to grab the data from cloud but whatever is written, uploaded or posted on the web belong to online services or the server and they may use it the way they want. Mostly data in the server that is kept for more than 180 days is considered to be abandoned or vague and is of no use to the online service. This data is often requested by different organizations like advertisement companies who analyze and use this data for targeted ads, so the data kept on cloud is often sold to these companies which in turn generate revenues but the user whose information is being used has no idea of it. Moreover, the protection of this data is even more difficult on mobile devices as it does not have the Fourth Amendment Protection as it would have stored on PC.

The issues of location tracking, most of today's mobile devices are builds with geo-location system which constantly keep track of their movement or their current location for example GPS system which helps the users to find their ways. The geo-located systems allow single users and groups to gather and spread different kinds of data. Most users are unaware of the fact that extra information that is collected about them is beyond requested data. For example, a photo which is taken by a user may uncover little contextual information which is observed from the style of any associated text. They are also unaware that the pictures and video clippings taken with their cameras contain geo location data. This data can be used to locate during travels, or uncover their house address. This can be considered as a source of data leakage and may cause a privacy breach.

The pictures uploaded through out smart phone on social sites like Facebook, are usually tagged and posted, these sited use tags associated with these photos to develop a detailed "Face-Prints" of what user looks like from every angle and is used for the purpose of citizenship application, criminal cases and security checks.

5.2 Data Acquisition

After the identification and consideration of the different situations that might exist related to the user views or behavior towards tracking we then will collect the related data for proceeding with the study.

For the data collection we will be using Online Survey Forms which will be consisting about 10-15 questions in context to different situation which re considered in the first step of the study. The survey form would be circulated in such way that users with different occupations and ages could be included like students, business related user, general web users. This will help in capturing the view of different not just different age groups but also different variety of online users.

There might be many users who could be completely unaware of the concept of web tracking or the idea that their web activities are actually being tracked. So considering this fact the questionnaire would be more general in such a way that the situation which are more relatable or frequently observed by the users will be considered and the users will be effectively give their views and opinions.

Finally, the data which will be collected through these online forms will be then used for the further processing.

5.3 Data Preprocessing

The data collected through online forms reviewed by different users will then be preprocessed that is the data which will be providing incomplete or irregular results will either be ignored or alternative will be considered for using that data in more effective manner (data cleaning).

After the data is satisfactorily cleaned and is without any kind of anomaly, the data would then be transformed into more integrated form where only the most persisting situations will then be considered, this will lead to data reduction. The reduced data will be then put to further use.

5.4 Feature Selection

The feature in our study will be different situations which will be identified in our study. After the data will be preprocessed we will them be short listing the key situations under consideration for the study for which the further evaluation will be conducted.

The basis on which the situation (feature) will be selected will be the following:

i. the one which has varying opinions from the user
ii. the most prominent situations
iii. the situation which have higher privacy concerns

During the selection of the situations, if particular situation will have very contrasting result with respect to different users then that particular situation be neglected or ignored. The min motive of the study is to get more precise and comprehensive view of the users with respect to the situations.

5.5 Classification Algorithm

For the classification of data, we will be considering different classifiers:

i. SVM
ii. AdaBoost
iii. EM

Support Vector Machine (SVM) is a supervised classifier which learns hyperplane in order to divide the data into two classes. A hyper plane is the function which is like equation of a line. For a simple classification task with only two features, the hyperplane will be a line. It can perform methods to display the data into greater dimension, after being projected to higher dimension it separated data into two classes. Then it consists of margins, which is distance between the hyperplane and its two closest data points from each of their class. It attempts maximize the margin so that hyperplane is closest to one class it is from the other. This decreases the chance of misclassification.

Expectation Maximization is an unsupervised classifier which is used as clustering algorithm (like the k- means) for discovering knowledge. It loops and improves, the possibility of observing the observed data while it is calculating the parameters of a statistical model with unknown variables. It begins by making assumption at model parameters. It follows an iterative type of mechanism:

i. E-step – this is based on the parameters of model and it calculates likelihood for assigning of each data pt to a cluster
ii. M-step – this step updates the model parameter on the basis of the cluster assignments from the first step
iii. It repeats until parameters of the model and the assignments of clusters are stabilized i.e. convergence

This algorithm is very simple and straightforward to implement. It not just optimizes the model parameter but also iterates to make guesses about the missing data.

AdaBoost is a supervised algorithm and is also called the boosting algorithm which construct classifier, where boosting joins learning algorithm by taking different learning algorithm and then combines them. The main objective is to group learners which are weak and then joining them to build a strong learner. Weak learner is the one with accuracy barely above the chance and a strong learner is the one with much higher accuracy. It is simple and very flexible as well as versatile since can be used in any learning algorithm and can work with different variety of data.

The classifier which will be having the best performance will be chosen for the final evaluation of the study.

5.6 Evaluation

The final stage will be evaluation in which on the basis of classifier which is chosen, the corresponding tool will be selected. Using the selected tool, we will depict and display the results of the study in the most effective way possible. The result will then show the direction in which the study has proceeded and what can be the future prospects of the research.

5.7 Flow Chart Process

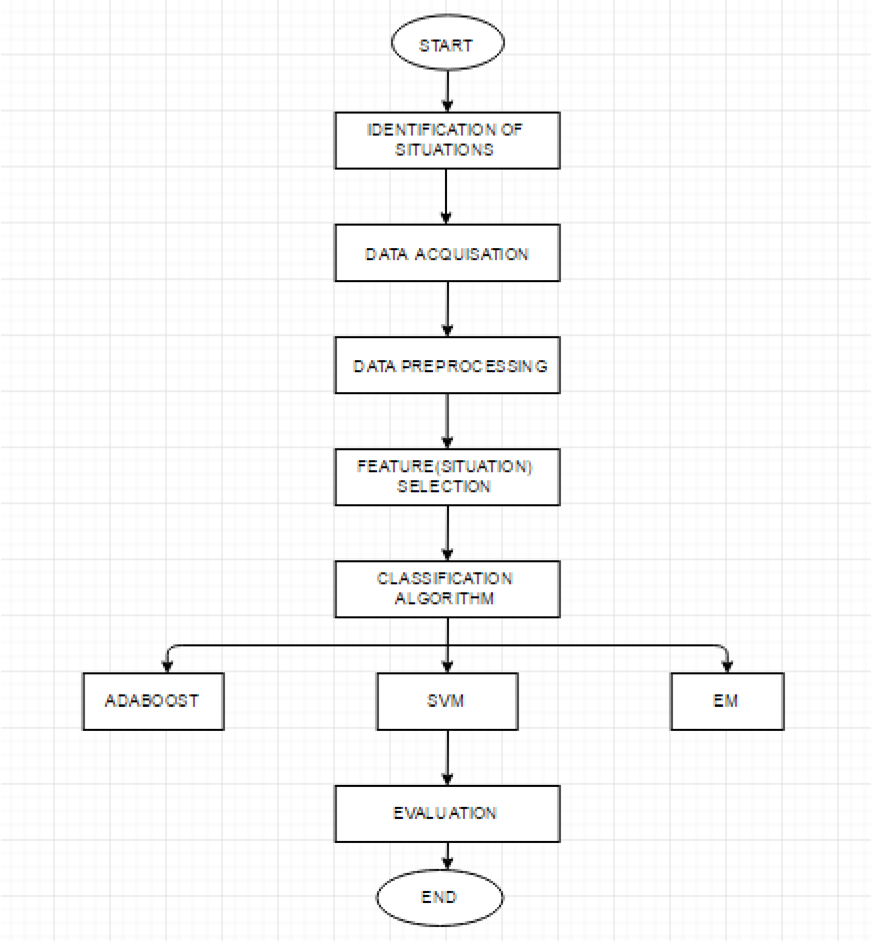

Figure 2 Flow Chart Process

5.8 Tool Be Used

i. Sci-kit Learn (used for the AdaBoost algorithm)

ii. Weka Tool (used for Expected Maximization algorithm)

iii. MATLAB Tool (used for SVM algorithm)

Chapter 6

EXPECTED OUTCOMES

The expected outcomes of this study are that we have a comprehensive view of how the user’s preferences and comfort for tracking vary when they come across different situations. Through the results generated from this results, the organization could implement new tools which could focus on these situations while they are being developed. This could help in incorporating the privacy issues of the users in their tools.

With the help of this the organization or web services could get a glimpse of what are the user’s opinion about being tracked and how these opinion and views could be used by the organization to provide their user with the most important commodity that user’s demand now-a-days i.e. “User’s Privacy “. This can be done with the help of new tools or incorporating specialized features for privacy in their existing features.

Chapter 7

CONCLUSION

So considering the current trends on the web where web tracking by the first-party (site owners) and the third-party (like Double Click) are more prominent. But it also gives rise to the problem of privacy of the users using the web. The companies following these practices should take into account how these practices are actually causing a privacy breach. On one hand we observe that tracking has some positive outcomes like customized search results but one can not neglect the fact that these services should not be provide at the cost of user's right to privacy of his or her information and activities.

Hence there is need of having the idea of how and in which situations the online tracking is actually impacting the users. The thorough examination of these situations and their counter reaction by the user can further help the organization in making necessary changes in their policies and technology. Not just that it can also help in developing new tools which can take into theses constraints into account and can thus help in reducing the privacy concern of the online users.

REFERENCES

[1] J. R. Mayer and J. C. Mitchell, "Third-party web tracking: Policy and technology," pp. 413–427, 2012.

[2] S. Englehardt and A. Narayanan, "Online Tracking: A 1-million-site Measurement and Analysis," *Proc. 2016 ACM SIGSAC Conf. Comput. Commun. Secur. - CCS'16*, no. 1, pp. 1388–1401, 2016.

[3] C. Eubank, M. Melara, D. Perez-Botero, and a Narayanan, "Shining the floodlights on mobile web tracking—a privacy survey," *W2Spconf.Com*, 2013.

[4] W. Melicher, M. Sharif, J. Tan, L. Bauer, M. Christodorescu, and P. G. Leon, "(Do Not) Track Me Sometimes: Users' Contextual Preferences for Web Tracking," *Proc. Priv. Enhancing Technol.*, vol. 2016, no. 2, pp. 1–20, 2016.

[5] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz, "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild," *Proc. 2014 ACM SIGSAC Conf. Comput. Commun. Secur. - CCS '14*, pp. 674–689, 2014.

[6] G. Kontaxis and M. Chew, "Tracking Protection in Firefox For Privacy and Performance," *IEEE Web 2.0 Secur. Priv.*, 2015.

[7] F. Roesner, T. Kohno, and D. Wetherall, "Detecting and defending against third-party tracking on the web," *Proc. USENIX Conf. Networked Syst. Des. Implement.*, no. Nsdi, p. 12, 2012.