Using Public Click Analytics and Twitter Metadata Occurring Inference Attack on Browsing Antiquity of Twitter Users

Ms. K. Srilekha
[1]PG Scholar,
Department of Computer Science and Engineering,
Vivekanandha College of Technology For Women,
Elayampalayam, Thiruchengode Tamilnadu,
India.

Mrs. R. V. Sudha
[2]Assistant professor,
Department of Information Technology,
Vivekanandha College of Technology For Women,
Elayampalayam, Thiruchengode.Tamilnadu,
India.

***Abstract*— Nowadays all them are using social network as twitter service for sharing short messages (tweets) among friend and communicate with each other, supported by a huge ecosystem. URL shortening services which provide a short alias of a long URL is an essential service for Twitter users who want to share long URLs via tweets having length restriction. Twitter allows client to post up to 140-character tweets containing only texts. Therefore, when users want to share complicated information (e.g., news and multimedia),they should include a URL of a web page containing the information into a tweet. URL reduction servicesbit.ly and goo.gl also provide the shortened URLs' click analytics consisting of the number of clicks, countries, internet service, and referrers of call on . Detect a simple inference attack that can estimate individual visitors from the aggregated, public click analytics using public metadata provided by Twitter. First, To analyses the metadata of client application and location because they can be correlated with those of public click analytics. Next, To perform the simple inference attack.**

Keyword: URL Shortening Service, Twitter metadata, Privacy leak, Inference attack

I. INTRODUCTION

Twitter is a popular online network services and micro blogging service for exchanging messages also known as tweets among peoples and communication each other's , supported by a huge ecosystem. Twitter announces that it has over 140 million active users creating more than 340 million messages every day and over one million registered applications built by more than 750,000 developers. The third-party applications include client applications for various platforms, such as Windows, Mac, IOS, and Android, and web-based applications such as URL shortening services, image-sharing services, and news feeds. Among the third-party services, URL shortening services. Which provide a short alias of a long URL is an essential service for Twitter users who want to share long URLs via tweets having length restriction. Twitter allows users to post up to 140-character tweets containing only texts. Therefore, when users want to share complicated information e.g., news and multimedia, they should include a URL of a web page containing the information into a tweet. Since the length of the URL and associated texts may overreach 140 characters, Twitter users demand URL shortening services further reducing it. Some URL shortening services (e.g., bit.ly and goo.gl) also provide shortened URLs' public click analytics consisting of the number of clicks, countries, browsers, and referrers of visitors. Although anyone can access the data to analyze visitor statistics, no one can extract specific information about individual visitors from the data because URL shortening services provide them as an aggregated form to protect the privacy of visitors from attackers. The main advantage of the preceding inference attack over the conventional browser antiquity stealing attacks is that it only demands public information. The conventional browser history stealing attacks rely on private information, such as Cascading Style Sheet (CSS) visited styles, browser cache, DNS cache, and latency. To collect such information, attackers should (i) prepare attack pages containing scripts/malware and lure target users for extracting the information from their web browsers or (ii) monitor DNS requests for measuring DNS lookup time of a target host. In other words, attackers should deceive or Compromise target users or their networks to obtain the browsing history, which relies on strong assumption. In contrast, anyone can access the metadata of Twitter and the public click analytics of URL shortening services so that passive monitoring is enough for performing our attack. The goal of the attacks is to know which URLs are clicked on by target users. Introduce two different attack methods: (i) an attack to know who click on the URLs updated by target users and (ii) an attack to know which URLs are clicked on by target users.

II RELATED WORKS

A. D.boyd et al

Twitter - a micro blogging service that enables users to post messages of up to 140 characters - supports a variety of communicative practices; participants use Twitter to converse with free spirit, too are more data its like groups, and the public at huge, so when conversations emerge, they are often experienced by broader audiences than just the interlocutors. In this paper inspect, the practice of re tweeting as a way by which colleague can be "in a conversation." While re tweeting has become a convention on Twitter, participants re tweet using different styles and for diverse reasons. They highlight how authorship, attribution, and communicative fidelity are negotiated in diverse ways. behavioral conventions have arisen over time and come to be inscribed in the Twitter technology, such as public yet directed messages using the @ symbol and hash tags (#'s) to mark tweets with topical keywords. Both of these conventions have clear conversational purposes. Honeycutt and Herring examine the conversational aspects of messages with the @ symbol. However, a third behavioral convention is known as the "re tweet", or the copying and rebroadcasting of another participant's message, enable conversations in a different manner. Re tweeting is inconsistent and messy. While conventions have emerged, they have not yet stabilized. Participants have different beliefs about how re tweets are "supposed" to work and this results in varied, and often conflicting, conventions. This is further complicated by third-party apps that use different syntax to mark re tweets. Before analyzing how re tweeting operates conversationally, they start by mapping out different aspects of re tweeting to highlight the variations in this practice. What follows is a discussion of the different syntax used to mark re tweet, how respondents modify re tweets, what content they choose to re tweet and their motivations for doing so.

B. Bin Liang et al

The existing Web timing attack methods are heavily dependent on executing client-side scripts to measure the time. However, many techniques have been proposed to block the executions of suspicious scripts recently. In this paper dispense a novel timing attack method to sniff Prepare Your Paper Before Styling Using a series of case studies and empirical data, this paper maps out re tweeting as a conversational practice.

A conversation is most commonly bounded by time, space and social context. Whether sitting around a table or talking on the telephone, conversations typically include a known, fixed set of participant who are assembled in real time in a particular social context for the purpose of talking to one another. The growth of computer-mediated communication, social media, and networked publics has shown that conversations can take place asynchronously and unbounded in space or time, but they are most often nevertheless bounded by a reasonably well-defined group of participants in some sort of shared social context. One kind of conversation that does not have a bounded set of participant is the kind described by marketers, celebrities, and politicians when they seek to be "in conversation" with their customers, fans, or constituents. These conversations do not typically involve direct dialogue, but a public interplay of voices that gives rise to an emotional sense of shared conversational context. Thus, the participants are no longer bounded except by a loosely shared social context. Because of Twitter's structure, which disperses conversation throughout a network of interconnected actors rather than constraining conversation within bounded spaces or groups, many people may talk about a particular story at once, such that others have a sense of being surrounded by the story, despite perhaps not being an active contributor in the conversation. users' antiquity histories without executing any scripts. Our method is based on the fact that when a resource is loaded from the local cache, its rendering process should begin earlier than when it loaded from a remote website. We leverage some Cascading Style Sheets (CSS) features to indirectly monitor the rendering of the target resource.

Three practical attack vectors are developed for different attack scenarios and applied to six popular desktop and mobile browsers. The evaluation shows that our method can effectively sniff users' browsing histories with very high precision. We believe that modern browsers protected by script-blocking techniques are still likely to suffer serious privacy leakage threats. One might, therefore, assume that it is infeasible to draw meaningful inferences about transactions of specific users from the public outputs of recommender systems. There show that this assumption is wrong.Our contributions. There develop a set of practical algorithms that allow accurate inference of (partial) individual behavior from the aggregate outputs of a typical recommender system. There focus on item-to-item collaborative filtering, in which the system recommends items similar to a given item. Our key insight is to exploit the dynamics of public recommendations in order to make the leap from aggregate to individual data The most widespread history sniffing attack relies on inspecting the visual style difference between the visited and unvisited links. In modern browsers, Cascading Style Sheets (CSS) can be employed to make visited and unvisited links take different colours or amounts of space. Based on this, attackers can place a list of URLs that they want to inspect in a web page and set the visited links to take a different style than the unvisited ones by using CSS. When a victim opens the page, a client-side script embedded in the page will check the style of links in the list or the positions of other elements, subsequently be determining whether the victim has recently visited a specific URL. For example, attackers can use CSS a: visited selector to set the font colour of visited links to red and unvisited links to green.

If the font colour of a link is red, a request can be submitted to a remote server controlled by attackers to inform them the link has been visited by the victim. Essentially, this kind of attack exploits browser bugs, e.g. to extract the visited status of given links. Fortunately, these bugs are easy to fix. In 2010, Baron of Mozilla Corporation proposed a solution for mitigating this kind of attack. All mainstream browsers, including Firefox, Chrome, Safari, and IE, have adopted this solution. As a result, attackers cannot distinguish visited links from unvisited ones according to their styles. It can be predicted that this kind of history sniffing technique will completely disappear following the update of users' browsers.

C. Z. Cheng et al

Micro blogging services such as Twitter allow users to interact with each other by forming a social network. The interaction between users in a social network group forms a dialogue or discussion. A typical dialogue between users involves a set of topics. There make the assumption that this set of topics remains constant throughout the conversation. Using this model of social interaction between users in the Twitter social network, along with content-derived location information, There employ a probabilistic framework to estimate the city-level location of a Twitter user, based on the content of the tweets in their dialogues, using reply-tweet messages.

There estimate the city-level user location based purely on the content of the tweets, which may include reply-tweet information, without the use of any external information, such as a gazetteer, IP information etc. The present framework for estimating user location does not consider the underlying social interaction, i.e. the structure of interactions between the users. In this project, There calculate a baseline probability estimate of the distribution of words used by a user. This distribution is formed by using the fact that terms used in the tweets of a certain discussion may be related to the location information of the user initiating the discussion. There also estimate the top K probable cities for a given user and measure the accuracy. There find that our baseline estimation yields an accuracy higher than the accuracy of the current state of the art estimation.

Our intuition is that a user’s tweets may encode some location-specific content – either specific place names or certain words or phrases more likely to be associated with certain locations than others (e.g., “howdy” for people from Texas). In this way, There can fill-the-gap for the 74% of Twitter users lacking city-level granular location information. By augmenting the massive human-powered sensing capabilities of Twitter and related micro blogging services with content-derived location information, this framework can overcome the sparsity of geo-enabled features in these services and bring augmented scope and breadth to emerging location-based personalized information services.

III SYSTEM ANALYSIS

Twitter is a popular online social network and micro blogging service for exchanging messages also known as tweets among people, supported by a huge ecosystem. Twitter announces that it has over 140 million active users creating more than 340 million messages every day and over one million registered applications built by more than 750,000 developers. The third-party applications include client applications for various platforms, such as Windows, Mac, IOS, and Android, and web-based applications such as URL shortening services, image-sharing services, and news feeds. Among the third-party services, URL shortening services. Which provide a short alias of a long URL is an essential service for Twitter users who want to share long URLs via tweets having length restriction. Twitter allows users to post up to 140-character tweets containing only texts. Therefore, when users want to share complicated information e.g., news and multimedia, they should include a URL of a web page containing the information into a tweet. Since the length of the URL and correlated texts may exceed 140 characters, Twitter users demand URL shortening services further reducing it.

A. History Stealing Using Timing Attack

The click analytics of the dispense shortened URLs and the metadata of the followers of the Twitter users. To perform the second attack, Create monitoring accounts that monitor texts from all followings of target users to collect all shortened URLs that the target users may click on. Then monitor the click analytics of those shortened URLs and contrast them with the metadata of the target user. Furthermore, Propose an advanced attack method to reduce attack overhead while increasing inference accuracy using the time model of target users, representing when the target users frequently.

B. Proposed Solution

Annually monitor click analytics of shortened URLs to observe its instant changes made by a new visitor. Whenever It notice that there is a new visitor, Its match his or her information with each of our target users to know whether the new visitor is one of our target users. It can estimate information about visitors by checking the differences between the new and the old click analytics.

The periodic monitoring, determining the optimal query interval is important, which depends on the variety of the characteristics of followers. When there are some characteristics to be observed at the same time and their whole values change rapidly, the query interval should be short enough to catch a small change. As It has many followers, the slope becomes so that should have a short interval.

C. Performance Metrics

- At the same time the third-party services, URL shortening services which provide a short alias of a long URL is an essential service for Twitter users the person who want to share long URLs via tweets having length restriction.

- Twitter allows users to post up to 140-character tweets containing only texts. Therefore, when users want to share complicated information news and multimedia, they should include a URL of a web page containing the information into a tweet.

- Since the length of the URL and associated texts may exceed characters, Twitter users demand URL shortening services further reducing it.

- Some URL shortening services also implement shortened URLs public click analytics consisting of the number of clicks, countries, browsers, and referrers of inspector.

- Although anyone can access the data to analyze visitor statistics, no one can extract specific information about individual visitors from the data because URL shortening services provide them as an aggregated form to protect the privacy of visitors from attackers.

D. System architecture

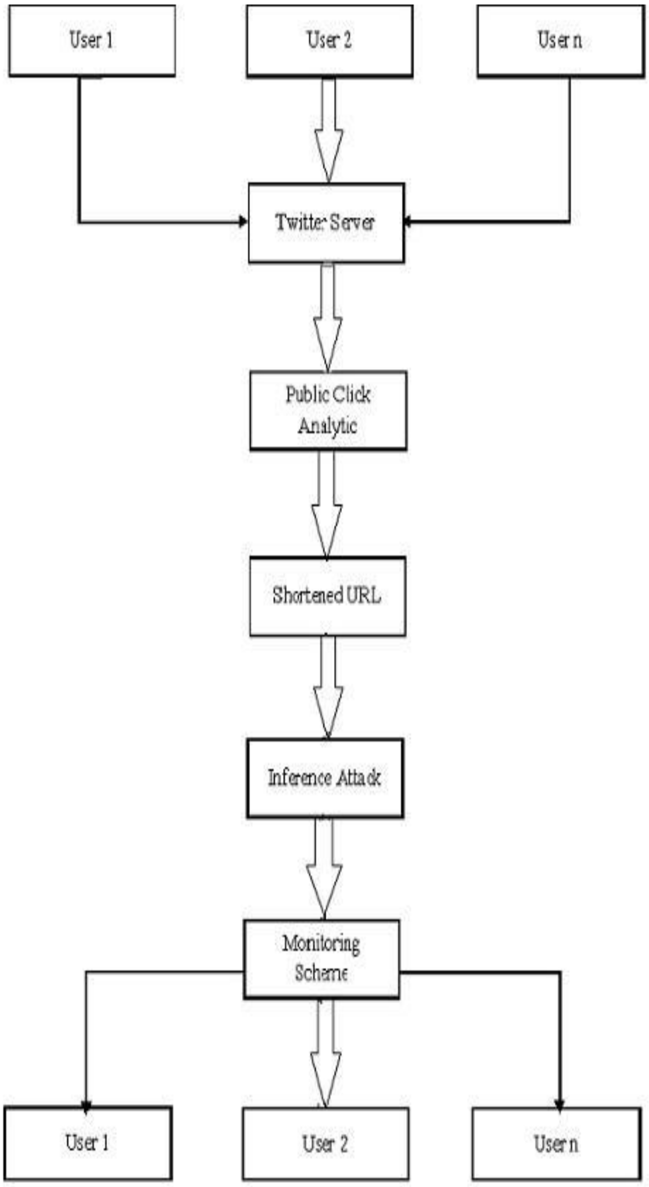

FIGURE D. SYSTEM ARCHITECTURE

IV MODULES

A.Modules Description

1) URL Shortening Service

Twitter is a popular online social network and micro blogging service for exchanging messages also known as tweets among people, supported by a huge ecosystem. Twitter announces that it has over active users creating more than 340 million messages every day and over one million registered applications built by more than developers. The third-party applications include client applications for various platforms, such as windows, Mac, and Android, and web-based applications such as URL shortening services, image-sharing services, and news feeds. Therefore, when users want to share complicated information news and multimedia, they should include a URL of a web page containing the information into a tweet. Since the length of the URL and associated texts may exceed characters, Twitter users demand URL shortening services further reducing it.

2) Twitter Meta Data

However, It detects a simple inference attack that can estimate individual visitors from the aggregated, public click analytics using public metadata provided by Twitter. First, It examines the metadata of client application and location because they can be correlated with those of public click analytics. For instance, if a user, Alice, updates her messages using the Twitter client application for iPhone, Twitter for iPhone will be included in the source field of the corresponding metadata. Moreover, Alice may disclose on her profile page that she lives in the USA or activate the location service of a Twitter client application to automatically fill the location field in the metadata.

3) Privacy Leak

Previous studies have considered attack techniques that cause privacy leaks in social networks, such as inferring private attributes and de-anonym zing users. Most of them combine public information from several data sets to infer hidden information. Some studies introduce de-anonym zing attacks in social networks. Back storm et al. try to identify edges existence in an anonym zed network, and Narayanan and Shmatikov identify Netflix records of known users using only a little bit of data about users. Furthermore, they combine their results with IMDb data and inferred user's political preferences or religious view. Narayanan and Shmatikov also prove that users who have accounts on both Twitter and Flickr can be recognized in the anonymous Twitter graph. Wondracek et al. [30] propose a de-anonym zed attack using group membership information obtained by browser history stealing attack.

4) Inference

Other studies consider how to infer the private attributes of users in social networks. He et al. And Linda mood et al. build a Bayesian network to predict undisclosed personal attributes. Shelve and Gentoo show how an attacker can exploit a mixture of private and public data to predict private attributes of a target user. Similarly, Midsole et al. infer the attributes of a target user by using a combination of attributes of the user's friends and other users who are loosely not directly connected to the target user. Celandine et al. Propose algorithms inferring customer's transactions in the recommender systems, such as Amazon and Hunch. They combine public data from the recommender systems and some of the transactions of a target user in order to infer the target user's unknown transactions. Chaabane et al. propose an inference attack predict undisclosed attributes by using only music interests. They derive semantics using Wikipedia ontology and measured the similarity between users.

B. Algorithm Implementation

The outcome of an inference attack is a set of candidate users whose information corresponds to the overlapping information. Therefore, the inference system tries to reduce the number of candidate users as much as possible for better performance.

The first define user and data models. Let U be user information released by the main service. Let D be a dataset released by the third party services. To protect the user's privacy, third-party services provide the online dataset D in aggregate form which consists of attributes a, values v and count of them c.

Algorithm 1 shows the procedure of the inference attack in the case of a single target user and a single third party service this way, It can obtain the service usage history of the user, which is defined as follows:

History = f(t : u) j u exist at time to;

Where (t: u) means that a user u used the service at time t. Algorithm 2 shows the procedure of the inference attack on multiple users. When our inference system obtains _d(t), the system compares it with each user. If UI is a subset of _d(t), the system adds UI into inferred history with time t. Finally, It can have the usage history that shows which users used the service at time t.

Algorithm 3 shows the procedure of the inference attack in case of multiple third parties when each dataset of third-party services has different attributes and has few attributes that the user has.

V CONCLUSION

Every information needed in our attacks is public information: the click analytics of URL shortening services and Twitter metadata. To evaluate our attacks, we crawled and monitored the click analytics of URL shortening services and Twitter data. Throughout the experiments, we have shown that our attacks can infer the candidates in most cases. Previous studies have considered attack techniques that cause privacy leaks in social networks, such as inferring private attributes and de-anonym zing users. Most of them combine public information from several different data sets to infer hidden information. Some studies introduce de-anonym zing attacks in social networks. They combine public data from the recommender systems and some of the transactions of a target user in order to infer the target user's unknown transactions. It proposes an inference attack to predict undisclosed attributes by using only music interests. They derive semantics using Wikipedia ontology and measured the similarity between users.

VI FUTURE ENHANCEMENT

In our inference attack, an adversary does not need to have private information or to exploit complicate techniques. Anyone who can access the released information can be an adversary. An adversary constantly monitors the information released by the services. Then, he tries to extract the private usage information of a target user by exploiting the information released by the connected services. Another essential requirement for the inference attack is that there should be overlapping information between the information released by the connected services. If the overlapping information corresponds with the information of a target user, an adversary can know that the target user has used the connected services.

ACKNOWLEDGMENT

A preliminary version of this paper was presented at WWW 2016 [23]. This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2016-H0301-13-3002)

REFERENCES

[1] Jonghyuk Song, Sangho Lee, and Jong Kim, “inference attack on browsing history of twitter users using public click analytics and twitter metadata,".in proc. IEEE symp. security and privacy (s&p) 2016.

[2] bin Liang. wei you. Liang Kun liu. Wen chang shi, “script less timing attacks on web browser privacy,”. in Proc. IEEE Symp. security and privacy (s & p) 2014.

[3] boyd.d. golden s. and lotan.g, “Tweet Tweet, Re tweet: Conversational aspects of re tweeting on twitter,” 2015.

[4] bugzilla.s, “understanding approval rating of agile project management tools using twitter,”2012.

[5] cheng.z. caverle.j.and leek, are where you tweet: “a content-based approach to geo locating twitter users,”. in Proc. 19th ACM international conference on information and knowledge management (ci km) 2014.

[6] chaabane.a. acs.g. and kaafar m. a, “inference models for twitter user's home location prediction,”. (ndss) 2011.

[7] eric Chen. y. and Zachary Weinberg, “i still know what you visited last summer leaking browsing history via user interaction and side channel attacks,” pros .33rd on automata, language, and programming) 2013.

[8] felten.e.w. and Schneider m. a, “timing attacks on web privacy,”. in Proc. 7th ACM conf. computer and comm. security (ccs) 2013.

[9] kilzer. a. and zhang. s, “securing official account twitter using social media management system: accuracy of the data and information publish with twitter,” 2010.

[10] marie Vasek. john Wad leigh, “hacking is not random: a case-control study of web server -compromise risk,” 2012