IJCTA, 10(9), 2017, pp. 145-152 ISSN: 0974-5572
© International Science Press

Inference Attack on URL Visiting History of the Social Networking

Aaradhana Deshmukh[1], Reshma Gade[2], Albena Mihovska[3] and Ramjee Prasad[4]

ABSTRACT

Privacy is one major challenge of today's rapid digitalization of the human life. A lot of personal information is unintentionally and not knowingly made available by one's participation to social networking sites, simple URL clicks, or accessing various Internet sites. The users interaction with the internet, create a digital presence that may contradict ones requirements for online privacy. Despite the fact that certain URL clicking techniques have been developed for reducing the amount of private information made publicly available, it is still possible to reconstruct user's preferences and interests and use this inference information for offering push advertisement or even privacy attacks. This paper surveys existing protection approaches to use of URL shortened service from the view point of their advantages and disadvantages.

Index Terms: URL shortening service, privacy leak, Twitter, Novel Attack Techniques.

1. INTRODUCTION

The Twitter blogging service allows for exchanging and sharing messages between the general population and companions and has been sustained by an extremely large community. It declares they have 140 million dynamic clients making more than 340 million messages for each day [21] and more than the one million enrolled applications worked above 750,000 designer [22]. The outsider applications add customer applications for discrete stages, such as Windows, Mac, iOS, and Android, and web based applications such as URL shortening services, picture sharing services, and news sustains [20]. To make the use of Twitter more lightweight, it has been proposed that all posted URL links are shortened [23]. When viewing the tweet on twitter.com, either the shortened or the original URL will be displayed. URL shorteners have become very popular because they help reduce the complexity of the longer links and emphasize on the contents that one would like to display to the Twitter audience. There are many available URL shortening tools. Some include full analytics and archives of everything that has been shortened. For these services, the URL would be replaced by a new domain (e.g., kevanlee.com changes to bit.ly) and the permalink would be replaced by a string of numbers and/or letters (e.g., kevanlee.com/best-writing-articles changes to bit.ly/df8jpI1). In addition, many social networks and social media management dashboards also provide a way to shorten long URLs automatically. The shortening services that include the URL analytics, would make available information related to the number of clicks or taps, and in addition may refer to the users that click on the link. Knowing the number of clicks on one's URL link, can be very beneficial for marketing purposes but this also requires that a level of protection of the privacy of the visitors of this URL would be enabled to protect from unwanted attacks on user's identity. Other possible threats originate from the fact that the short URL obscures the target address and, thus, attackers may redirect it to a harmful, banned or

[1] Department of Electronics System, Aalborg University, Denmark, *Email: aad@es.aau.dk*

[2] PG student, Dept. of Computer Engineering, SKNCOE, Pune, *Email: gadereshma5@gmail.com*

[3] Department of Electronics System, Aalborg University, Denmark, *Email: albena@es.aau.dk*

[4] Aarhus University, Denmark, *Email: ramjee@btech.au.dk*

unwanted site. Shortened URLs also can be force-searched to disclose private content pertaining in some cases to the user's identity (e.g., Google maps is prone to disclosing home addresses, etc.).

The authors in [4] proposed an attack technique to infer whether a specific user clicked on certain shortened URLs on Twitter. The fundamental advantage of the preceding inference attack over the browser history taking attacks is that it just requests open data. The aim of these attacks is to know which URLs are tapped on by target clients. The attack strategies include the following: attack to know (i), who clicked on the URL and (ii), which URLs are clicked. To analyze the attack, there are two strategies: (1) To locate the various Twitter clients who disseminate URLs and to research the click analytics of the dispersed URLs and the followers metadata of the Twitter clients; and (2) To make checking accounts that screen messages from all the followers of the target clients and to gather all shortened URLs that the target clients have tapped on. At that point, the shortened URLs and their click analytics would be compared with the target user's metadata. In the following, we have performed a literature survey to assess the proposed ways to privacy protection when using shortened URLs.

2. LITERATURE SURVEY

In [1] author presents calculations, which take a little measure of the assistant data about a client and derive this current client's exchanges from worldly changes in a little measure of assistant data about a client and derive this current client's exchanges from worldly changes in the public yields of a recommender framework. Our derivation assaults are detached and can be carried out by any Internet client. They go out on a limb of community oriented filtering. In [2] the author introduces a novel planning assault technique to clients' scanning histories without executing any scripts. This technique depends on the way that when an asset is loaded from the neighbourhood reserve, its rendering procedure ought to start sooner features to in a roundabout way screen the rendering of the objective asset then when it is loaded from a remote site. They influence some css. The assessment demonstrates that the technique can viably clients' skimming histories with high accuracy. They trust that present day programs secured by script blocking systems are still prone to endure genuine protection spillage dangers. In [3] the proposed system, the act of retweeting is used to find out, which members can be in a discussion. While retweeting is turned into tradition inside Twitter, members retweet utilizing distinctive styles and for diverse reasons. They highlight how initiation, attribution, and informative constancy are arranged in diverse ways. Utilizing a progression of contextual investigations and exact information, this paper maps out retweeting as a conversational practice. In [4] the induction attack that surmises shortened URLs that are tapped on by the objective client. All the information required in this attack is open data; that is, the snap investigations of URL abbreviation or shortening administrations and Twitter information of clients. Both data are public and can be gotten to by anybody. They consolidated two bits of open data with construed applicants. To assess this framework, they crept and checked the click investigation of URL shortening administrations and Twitter information. In [5] author proposed, Experiments on Twitter occasion discovery exhibited that technique can successfully remove reliable tweets while barring bits of gossip and noise. Furthermore, a similar execution investigation exhibited that technique outperforms existing directed learning plans utilizing tweets physically marked or tweets produced in view of catchphrase coordinating as the preparation set.

In [6] author Spearheading the information obtaining of history-based client inclinations. Examine the impact of css-based history location also show practically of conducting practical assaults within significant assets. In [7] Stricter substance taking care of principles that totally obstruct the assault, the length of the focused on web webpage does not make certain blunders. Introduce a general type of this attack can be made to work in any program that support or backings css, regardless of the possibility that JavaScript is handicapped or unsupported. In [8] to play out an expounded examination to uncover extra exploitable program systems. With more dynamic and smart components exhibited in programs in present times. Introduced another planning assault strategy for sniffing clients' perusing histories. In [9] author take both

Sr. No.	*Explanation*	*Data collection used for attack*	*Attack Techniques*
1	In [1] They develop an algorithm that allows to take small amount of auxiliary information about customer and inferring customer's transactions in recommender system.	Combine public data of recommender systems and some private data from user's unknown transactions.	Privacy leaks from public information. This attack is inference attack.
2	In [2] they shows that the attack allow a malicious web site to determine whether or not the user has recently visited web pages.	Confess other type of information gathering by web sites like cookies.	This attack is history stealing attack.
3	In [4] they shows that attack infers shortened url that are tapped on by target client.	Information is required in this attack is publicly open information.	This attack is inference attack.
4	In [5] They propose different method to evaluate user trustworthiness in twitter.	A bad-natured user attacks without mask whenever it has chance.	This is persistent attack behavior.
5	In [6] They exhibit an algorithm for recognizing visited links and its implantation of JavaScript. They analyse behaviors of every browser related to CSS visited styles and manufacture a system to distinguish browsing history of the clients.	Information of users obtained by ad networks and social networks.	This is history stealing attack.
6	In [7] They general form of attack that uses style sheet import to steal confidential information from victims website.	Information is combined by web sites like yahoo mail, IMDb.	This is history stealing attack.
7	In [8] They begun a timing attack method to clients browsing histories without complete any scripts.	They use CSS animation, scrollbar customization and media queries to gathering users browsers history.	This attack is history stealing attack
8	In [9] They disclose that personal attributes can be inferred with great precision when people associated with solid relationship.	All information is needed in this attack is combine from social networks.	Privacy leaks from public information and This is inference attack.
9	In [10] They presents all the more effective tracking strategies based on caching different kinds of records or files.	Information is gathering by cache, visited links etc.	This is history stealing attack.
10	In [12] they introduce the practical attack that makes utilization of gathering data on person to person social networking sites. They demonstrate that group membership of client in social network may enough information about user to recognize her/him when visiting web pages from outsiders or third parties.	Information is gathering from group membership of user that is the group inside a social network in which client is a part.	This attack is de-anonymization attack.
11	In [16] this paper they proposed the attack technique which leaks hidden or missing attributes using public attribute.	Information is gathering from Facebook application. Information is needed in this attack is public information.	This attack is inference attack.
12	In [17] they analyze inference technique using released social networking data to predict private information about users.	Information is needed in this attack is social networking data.	This attack is inference attack.

Figure 1: Comparative Study of Attack Techniques.

social network structures and influence quality of social relations into thought. Investigated the issue of security inference in social networks using Bayesian network. In [10] author Propose that a general same-origin principle should be apply uniformly across various types of data or information stored on a web client's machine. In this also create ways for users to limit tracking, in the form of browser extensions that are accessible for download. In [11] Figure out if they can recognize a client's area by just taking a gander at what that client tweets. Furthermore, found that a client's nation and state. Look at the data installed in client profile area fields. Through this investigation, they have demonstrated that numerous clients select to enter no data or non-genuine area data that can without much of a stretch simpleton Geographic data instruments. At the point when clients do enter their genuine areas, they have a tendency to be not any more exact than city scale. In [12] the authors announce a new kind of de-anonymization attack that exploits group membership information that is accessible on social networking sites. More precisely, they show that information about the group memberships of a user is sufficient to uniquely identify this person, at least reduce the set of possible candidates. In [13] an attack, history sniffing, which showed up as an unintended consequences of the mix of three independently attractive features: visited-link indication to the clients, CSS control of all aspects of page appearance, and JavaScript monitoring and checking of page rendering. Automated history sniffing attacks, including timing attacks. However, attacks that involve the client remain possible, as do attacks via side channels outside of the browser's control. In [14] An Anonymity is not adequate for security when managing with social networks. We built up a nonspecific re-identification algorithm and demonstrate that it can effectively de-anonymize few thousand clients in the anonymous graph of a popular micro blogging service, using a totally different social network (Flickr) as the source of small amount of helping information. In [15] they demonstrate that an adversary who knows just a little bit about an every subscriber can easily recognize this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, so effectively recognized the Netflix records of known clients, uncovering their apparent political inclinations and other possibly sensitive information. In [16] author presents techniques to predict private user attributes using only music interest that are disclosed by users, they separate ephemeral interest topics by analyzing the collection of interest. They infer private attribute using this technique using user's interest to predict hidden information. In [17] author presents the inference attack technique using the social network data to predict the undisclosed private information. They addressed issues like private information leakage in social networks. In [18] the authors present the model that use friendship and group membership information to infer the sensitive information using mixture of private and public use profiles. They uses link based and group based classification to study privacy implication in social networking.

3. GAP ANALYSIS

In the above survey the attacker exploit CSS visited styles and utilizes the browsers display visited links in contrast to unvisited links. Attacker misuse the browser DNS cache to conduct history stealing attacks. Some researcher propose attack strategies to steal browsing history utilizing clients interaction also use webcam to identify the light of screen reflected at users face, which can be used to recognize the colors of unvisited links to visited links. All the history stealing attacks accept that victims visit a malicious website page or victims are contaminated by malware. Our inference attacks do not need to make these assumptions. Our inference attacks just utilize the combination of freely accessible data, so anybody can be an attacker or a victim.

4. EXISTING SYSTEM

In the paper, they proposed a derivation attack that induces shortened URLs that are tapped on by the objective client. All the data required in our attack is open data; that is, the click investigation of URL shortening services and Twitter metadata. Both information is open and can be accessed by anybody. We joined two bits of open data with inferred candidates. To assess our system, we checked the click analytics of URL shortening services and Twitter information. All through the tests, we have demonstrate that our attack can inferring the candidates in the larger part of cases. To the best of our insight, this is the primary study that infers the URL visiting history on Twitter. We additionally proved that if an attacker knows some data about the objective client, he could figure out if the objective client taps on the shortened URL [20].

5. DRAWBACKS OF EXISTING SYSTEM

1) The periodic checking, monitoring and coordinating have a restriction since Twitter does not authoritatively give individual data about client such as country, browsers, and platforms.
2) URL is a fundamental service for Twitter users who need to share long URLs by means of tweets having length limitation

6. USE CASE DIAGRAMS

6.1. For User

Fig. 2. shows the use case diagram for the user of the twitter system which provide input as no. of users action such as Registration, Login, Get URL shortening, Post a message, Identify the attack details and identify the users. When user login to twitter system hit and distributes the URLs and checks the click analytics.

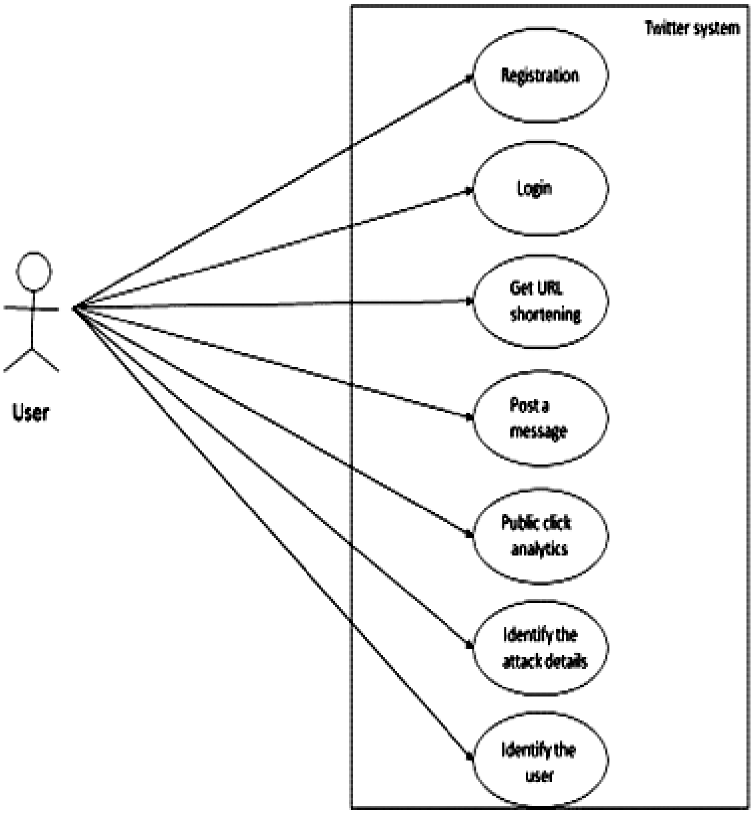

Figure 2: Use case diagram for user.

6.2. For Admin

Fig. 3. Shows that the use case diagram for admin which takes actions like Login, Access the public clicks analytics and when attacker is find then block the attack details. Admin monitors the instant changes periodically and find attacker details and block it.

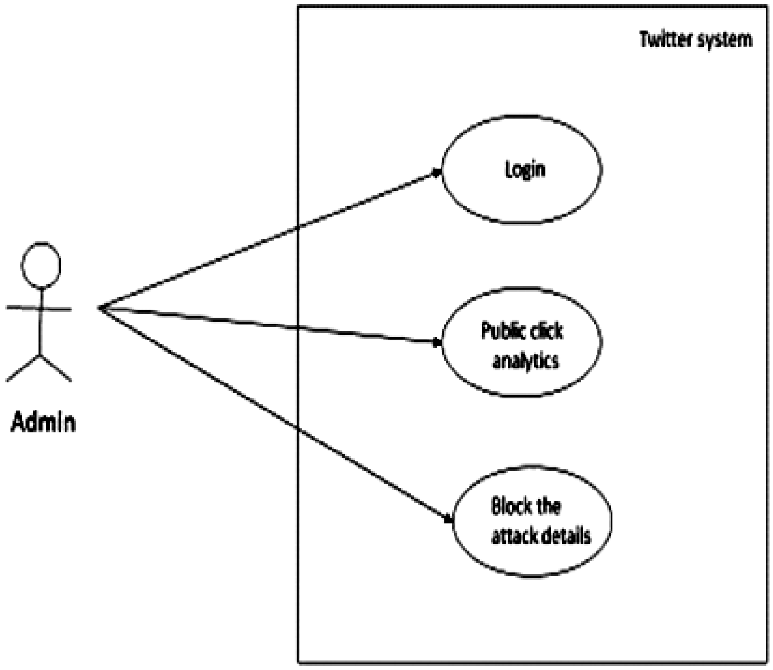

Figure 3: Use case diagram for admin

7. PROPOSED SYSTEM

Fig. 4. Shows that when the user logs in to the system, the user will distribute the URLs and Get or click on the URLs also accessing public clicks. If an unknown person clicks on the user's shortened URL then the admin checks and monitors continuously public clicks and instant changes in the click by means of analytics. Afterwards, Identify user will find the attack details and block the details of the attacker. A good strategy for gathering when a particular client tapped on a fixed shortened URLs on Twitter is to use freely accessible data to perform a click investigation from the URL shortening services and metadata from Twitter. Two

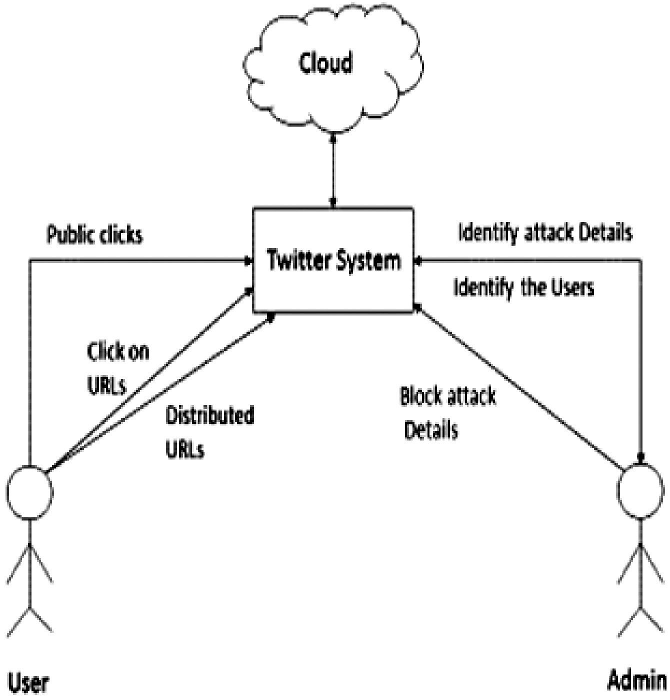

Figure 4: System Architecture.

distinctive attack strategies are possible: (1) an attack to know who tapped on the URLs conveyed by target clients, and (2) an attack to know, which URLs are tapped on by target clients. To play out a primary attack, the various Twitter clients who most of the time convey abbreviated URLs, can be located and then the click analytics of the distributed shortened URLs can be examined together with the metadata of the supporters of the Twitter clients. To play out a secondary attack, we make checking records that monitor the messages from all followers of the target clients to assemble all shortened URLs that the objective clients click on. At that point, we may monitor the click analytics of those abbreviated URLs and contrast all of them and the metadata of target client. With the scheme in Fig.4.,we are able to identify the number of attack and number of attacks user details. In the existing systems, it is not possible to find the exact location of the attacker because some twitter users have similar information location and there are no other ways to identify the attacker.

8. CONCLUSION

This paper investigated the vulnerabilities of social media users when clicking on various links posted by their social friends. Inference attacks are a strong tool to derive, which shortened or abbreviated URLs have been tapped on by target clients are an effective strategy to protect social media users. The entire data required in these attacks is open data: the click examination of URL shortening services and Twitter metadata. For assessing these attacks, one may observe the click investigation or analytics of URL shortening services and Twitter data. Furthermore by utilizing this we discover the attacker details and block that attacker details.

REFERENCES

[1] J. A. Calandrino, A. Kilzer, A. Narayanan E. W. Felten, “You might also like: Privacy risks of collaborative filtering” in Proc, IEEE Symp. Secure. Privacy-11.

[2] E. W. Felten and M. A. Schneider, “Timing attacks on web privacy” in Proc. 7th ACM Conference. Comput. Communication. (css), 2000.

[3] D. Boyd and G. Lotan, “Tweet, tweet, retweet: Conversational aspects of retweeting on twitter,” in Proc. 43rd Hawaii International Conference System Science.

[4] Jonghyuk Song, Sangho , Jong Kim, “I Know the Shortened URLs You Clicked on Twitter: Inference Attack using Public Click Analytics and TwitterMetadata”.

[5] Liang Zhao 1, Ting Hua1, Chang-Tien Lu, “A Topic-focused Trust Model for Twitter”.

[6] A. Janc “Web browser history detection as a real world privacy threat” in Proc. 15th Eur. Conference Res. Computing. Secure.-10.

[7] Lin-Shung Huang and Chris Evans, “Protecting Browsers from Cross-Origin CSS Attacks”.

[8] Liang, Wei You, Liangkun and Wenchang Shi, “Script less Timing Attacks on Web Browser Privacy”.

[9] C. Jackson, A. Bortz and J. C. Mitchell, “Protecting browser state from web privacy attacks” in Proc., 15th International World Wide Web Conference-06.

[10] J. He, W. W. Chu, and Z. V. Liu, “Inferring privacy information from social networks” in Proc.,4th IEEE International Conference Informatics-06.

[11] B. Hecht, L. Hong and E. H. Chi, “Tweets from justinbieber’s heart: The dynamics of the location field in user profiles” in Proc., SIGCHI Conference Human Factors Comput. System-11.

[12] G. Wondracek, T. Holzand and C. Kruegel, “A practical attack to de-anonymize social network users,” in Proc., IEEE. Secur. Privacy-10.

[13] Z. Weinberg, E. Y. Chen, C. Jackson, “I still know what you visited last summer: Leaking browsing history via user interaction and side channel attacks,” in Proc. IEEE Symp. Secur. Privacy-11.

[14] A. Narayanan, V. Shmatikov, “De-anonymizing social networks” in Proc. 30th IEEE Symp. Secur. Privacy-09, pp. 173–187.

[15] A. Narayanan, V. Shmatikov, “Robust de-anonymization of large sparse data set,” in Proc. IEEE Symp. Secur. Privacy-08, pp. 111–125.

[16] A. Chaabane, G. Acs, M. A. Kaafar, “You are what you like! Information leakage through user’s interests,” in Proc. 19th Network and Distributed System Security Symp.-12.

[17] J. Lindamood, R. Heatherly, M. Kantarcioglu, B. Thuraisingham, “Inferring private information using social network data” in Proc. 18th International World Wide Web Conference (www)-09.

[18] E. Zheleva , L. Getoor, “To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles, in Proc. 18th International World Wide Web Conference-09, pp. 531–540.

[19] L. Backstrom, C. Dwork, J. Kleinberg, “Wherefore art thou r3579x? anonymized social networks, hidden patterns, and structural steganography, in Proc. 16th International World Wide Web Conference-07, pp. 181–190.

[20] Jonghyuk Song, Sangho Lee, Member, IEEE, Jong Kim, Member, IEEE “Inference Attack on Browsing History of Twitter Users Using Public Click Analytics and Twitter Metadata”.

[21] Twitter blog. (2012). Shutting down spammers. [Online]. Available:http://blog.twitter.com/2012/04/shutting-down-spammers.html

[22] Twitter blog. (2011). One million registered twitter apps. [Online]. Available: http://blog.twitter.com/2011/07/ one-million-registered-twitter-apps.htm

[23] www.support.twitter.com.