PUBLIC CLICK DATA AND METADATA EXPLOITATION FOR INFERENCE ATTACKS ON TWITTER

Abigail Hill[1], Jack Adams[2], Samuel Thompson[3], and Daniel Robinson*[3]

[1] Department of Civil Engineering, University of Melbourne, Australia
[2] School of Engineering, University of Toronto, Canada
[3] Department of Chemical Engineering, University of Manchester, UK

ABSTRACT

Twitter is one of the most popular microblogging services, which is generally used to share news and updates through short messages restricted to 280 characters. However, its open nature and large user base are frequently exploited by automated spammers, content polluters, and other ill-intended users to commit various cyber crimes, such as cyberbullying, trolling, rumor dissemination, and stalking. Accordingly, a number of approaches have been proposed by researchers to Address these problems. However, most of these approaches are based on user characterization and completely disregarding mutual interactions. In this study, we present a hybrid approach for detecting automated spammers by amalgamating community based features with other feature categories, namely metadata content, and interaction-based features.

KEYWORDS: Social network analysis, Spammer detection, Spambot detection, Social network security.

1. INTRODUCTION

The field covers all the processes and mechanisms by which computer-based equipment, information and services are protected from unintended or unauthorized access, change or destruction. Computer security also includes protection from unplanned events and natural disasters. Otherwise, in the computer industry, the term security -- or the phrase computer security -- refers to techniques for ensuring that data stored in a computer cannot be read or compromised by any individuals without authorization. Most computer security measures involve data encryption and passwords. Data encryption is the translation of data into a form that is un intelligible without a deciphering mechanism. A password is a secret word or phrase that gives a user access to a particular program or system. 1.2 Working conditions and basic needs in the secure computing

If you don't take basic steps to protect your work computer, you put it and all the information on it at risk. You can potentially compromise the operation of other computers on your organization's network, or even the functioning of the network as a whole. Physical security Technical measures like login passwords, anti-virus are essential. (More about those below) However, a secure physical space is the first and more important line of defense. Is the place you keep your workplace computer secure enough to prevent theft or access to it while you are away? While the Security Department provides coverage across the Medical center, it only takes seconds to steal a computer, particularly a portable device like a laptop or a PDA. A computer should be secured like any other valuable possession when you are not present. Human threats are not the only concern. Computers can be compromised by environmental mishaps (e.g., water, coffee) or physical trauma. Make sure the physical location of your computer takes account of those risks as well. Access passwords The University's networks and shared information systems are protected in part by login credentials (user-IDs and passwords). Access passwords are also an essential protection for personal computers in most circumstances. Offices are usually open and shared spaces, so physical access to computers cannot be completely controlled. To protect your computer, you should consider setting passwords for particularly sensitive applications resident on the computer (e.g., data analysis software), if the software provides that capability. Prying eye protection Because we deal with all facets of clinical, research, educational and administrative data here on the medical campus, it is important to do everything possible to minimize exposure of data to unauthorized individuals. Anti-virus software Up-to-date, properly configured anti-virus software is essential. While we have server-side anti-virus software on our network computers, you still need it on the client side

(your computer).Anti-virus products inspect files on your computer and in email. Firewall software and hardware monitor communications between your computer and the outside world. That is essential for any networked computer. It is critical to keep software up to date, especially the operating system, anti-virus and anti-spyware, email and browser software. The newest versions will contain fixes for discovered vulnerabilities. Almost all anti-virus have automatic update features (including SAV). Keeping the "signatures" (digital patterns) of malicious software detectors up-to-date is essential for these products to be effective.Even if you take all these security steps, bad things can still happen. Be prepared for the worst by making backup copies of critical data, and keeping those backup copies in a separate, secure location. For example, use supplemental hard drives, CDs/DVDs, or flash drives to store critical, hard-to-replace data. If you believe that

your computer or any data on it has been compromised, your should make a information security incident report. That is required by University policy for all data on our systems, and legally required for health, education, financial and any other kind of record containing identifiable personal information.

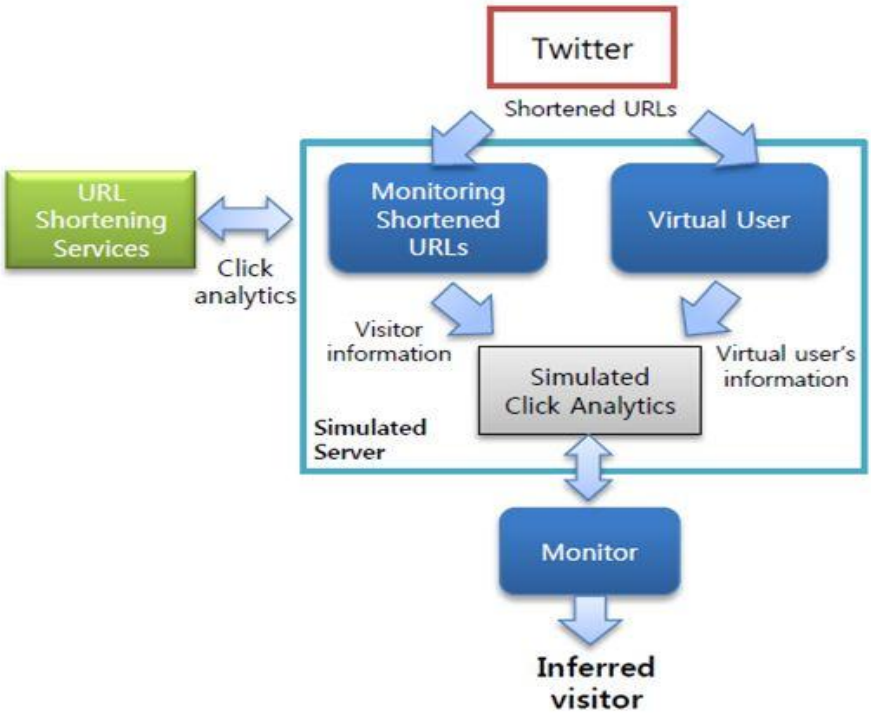
2. MATERIALS AND METHODS

Public Click Analytics Method

Some researchers propose attack methods to steal browsing history using user interactions and side-channels. Weinberg et al. exploit CAPTCHA to deceive users and to induce user's interaction. They also use a webcam to detect the light of the screen reflected at the user's face, which can be used to distinguish the colors of visited from those of unvisited links. He et al. and Linda mood et al.build a Bayesian network to prundisclosed personal attributes. Zheleva and Getoor show how an attacker can exploit a mixture of private and public data to predict private attributes of a target user. Similarly, Mnislove 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. Calandrino et al. propose algorithms inferring customer's transactions in the recommender systems, such as Amazon and Hunch. Previous studies have considered attack techniques that cause privacy leaks in social networks, such as inferring private attributes and de-anonymizing users. Most of them combine public information from several different data sets to infer hidden information. Need complicated techniques or assumptions In this paper, we propose novel attack methods for inferring whether a specific user clicked on certain shortened URLs on Twitter. Our attacks rely on the combination of publicly available information: click analytics from URL shortening services and metadata from Twitter. The goal of the attacks is to know which URLs are clicked on by target users. We 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.

To perform the first attack, we find a number of Twitter users who frequently distribute shortened URLs, and investigate the click analytics of the distributed shortened URLs and the metadata of the followers of the Twitter users. To perform the second attack, we create monitoring accounts that monitor messages from all followings of target users to collect all shortened URLs that the target users may click on. We then monitor the click analytics of those shortened URLs and compare them with the metadata of the target user. Furthermore, we 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 use Twitter.

Figure:



Modules descsription:

A. System construction and Data Collection

In this module, we develop the system with the entities to implement and evaluate our proposed model. In this module, we implement URL shortening services. Profiling module obtains the information of the target user from the target user's profile and timeline. We collected data by crawling the click analytics of the shortened URLs, using the API methods offered by goo. Gland bit.ly. goo.gl APIs have a rate limit of 1,000,000 queries per day. Similarly, bit.ly allows users to create no more than five concurrent connections from one IP address. bit.lyalso enforces per-hour limits, per-minute limits, and per-IPrate limits for each API method. However, bit.ly does not publish the exact number of allowed requests on each limit.

B. Periodic Monitoring and Matching

The monitoring module extracts the shortened URLs from the tweets posted by the followings of the target user and monitors the changes in the click analytics of the shortened URLs. We create a Twitter user (monitoring user) who follows all the followings of the target user in order to access all tweets that the target user may view. In this module, We periodically monitor click analytics of shortened URLs to observe its instant changes made by a new visitor. Whenever we notice that there is a new visitor, we match his or her information with each of our target users to know whether the new visitor is one of our target users. We can estimate information about visitors by checking the differences between the new and the old click analytics. The matching module compares the information about the new visitor with the information about the target user when the monitoring module notices the changes in the click analytics. If the matching module infers that the new visitor is the target user, it includes the corresponding shortened URL in a candidate URL set.

C. Referrers

In this module, we determine whether a new visitor comes from Twitter by using the changed referrer information of public click analytics. The click analytics of goo.gl only records the hostname of the referrer site. If a visitor comes from Twitter, "t.co" or "twitter.com" is recorded in the Referrers field. In most cases, "t.co" is recorded because all links shared on Twitter are automatically shortened to t.co links.t.co handles redirections by context and user agents so that the Referrer information varies according to the source of a click. In some cases, "twitter.com" is recorded because some Twitter applications directly use original links instead of t.co

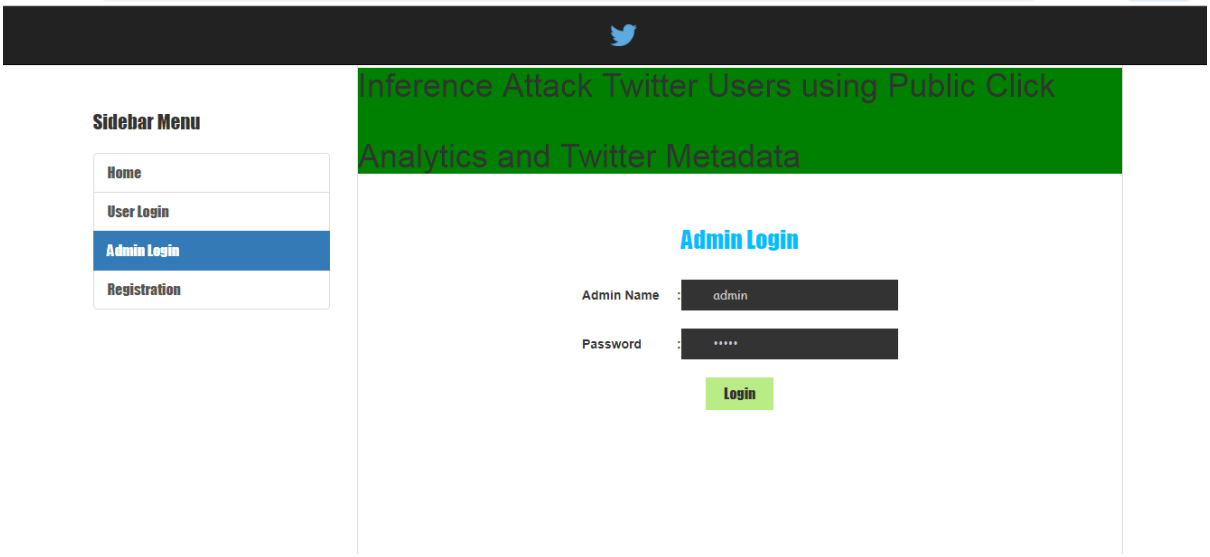
links. Consequently, if the Referrers information of the visitor is "t.co" or "twitter.com", we regard the visitor as coming from Twitter.

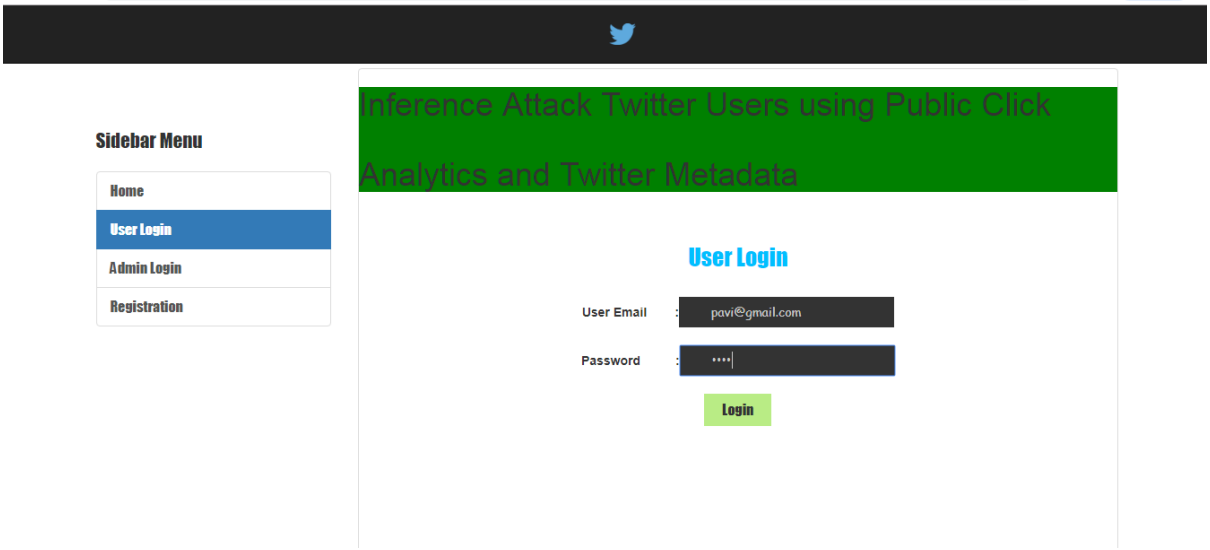
D. Inference Attack

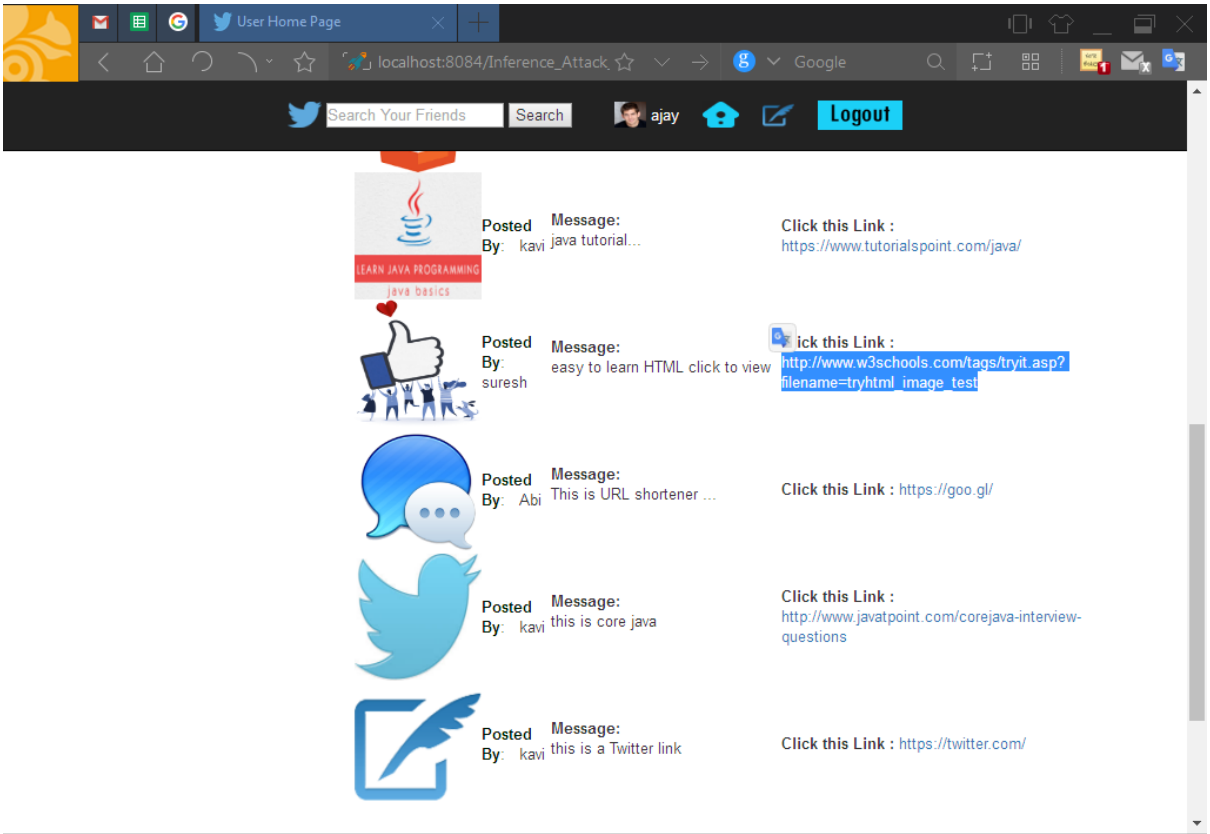
In this module, we evaluate our inference attack in the simulated environment to correctly identify its accuracy The definite ways to exactly evaluate our attacks are (i)asking the target users whether they really visited the

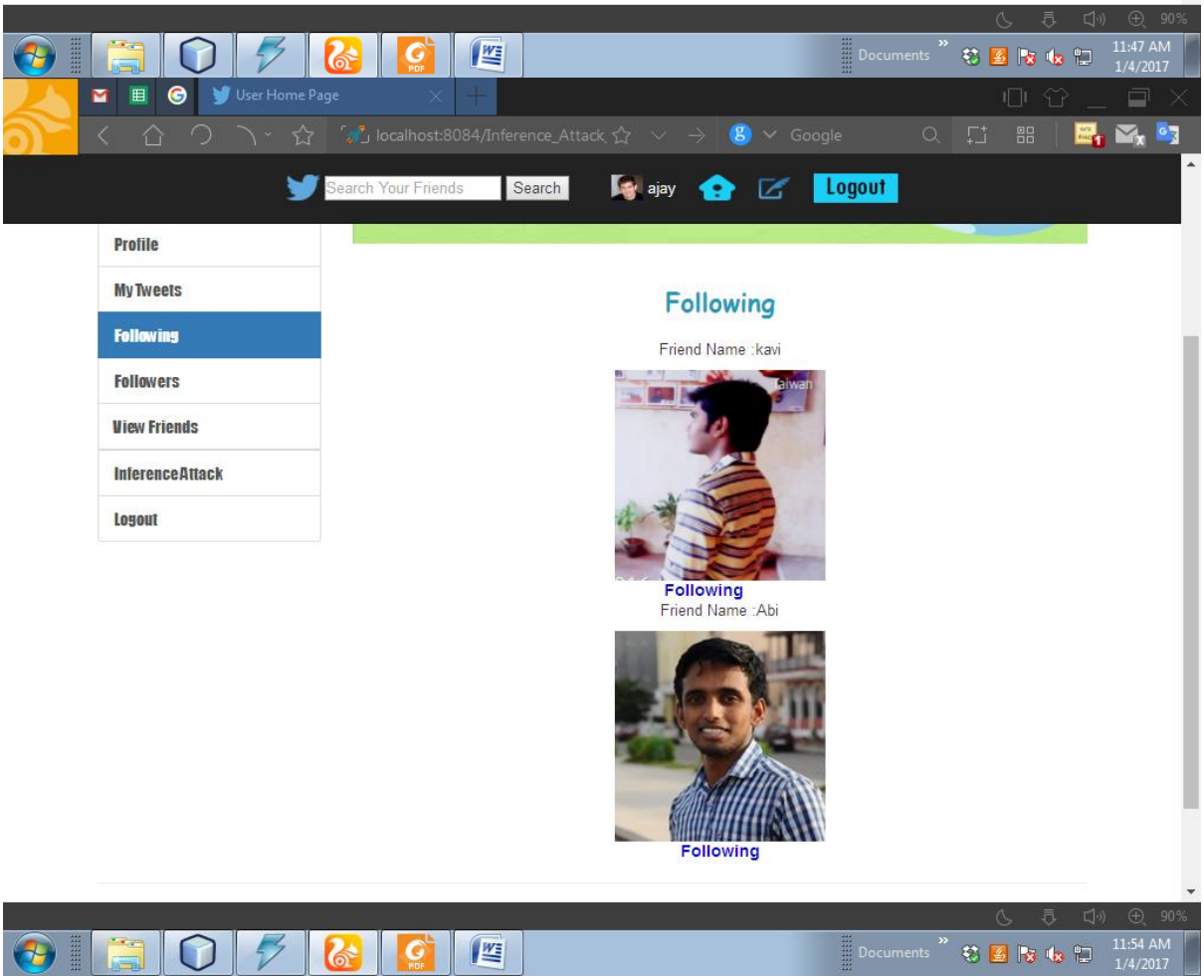
shortened URLs or (ii) monitoring their browsing activities by using logging software. But, both approaches are restrictive because we cannot survey all of them or require them to install the logging software. Therefore, we construct a simulated environment supported by real click analytics to perform our attacks in almost real environment.

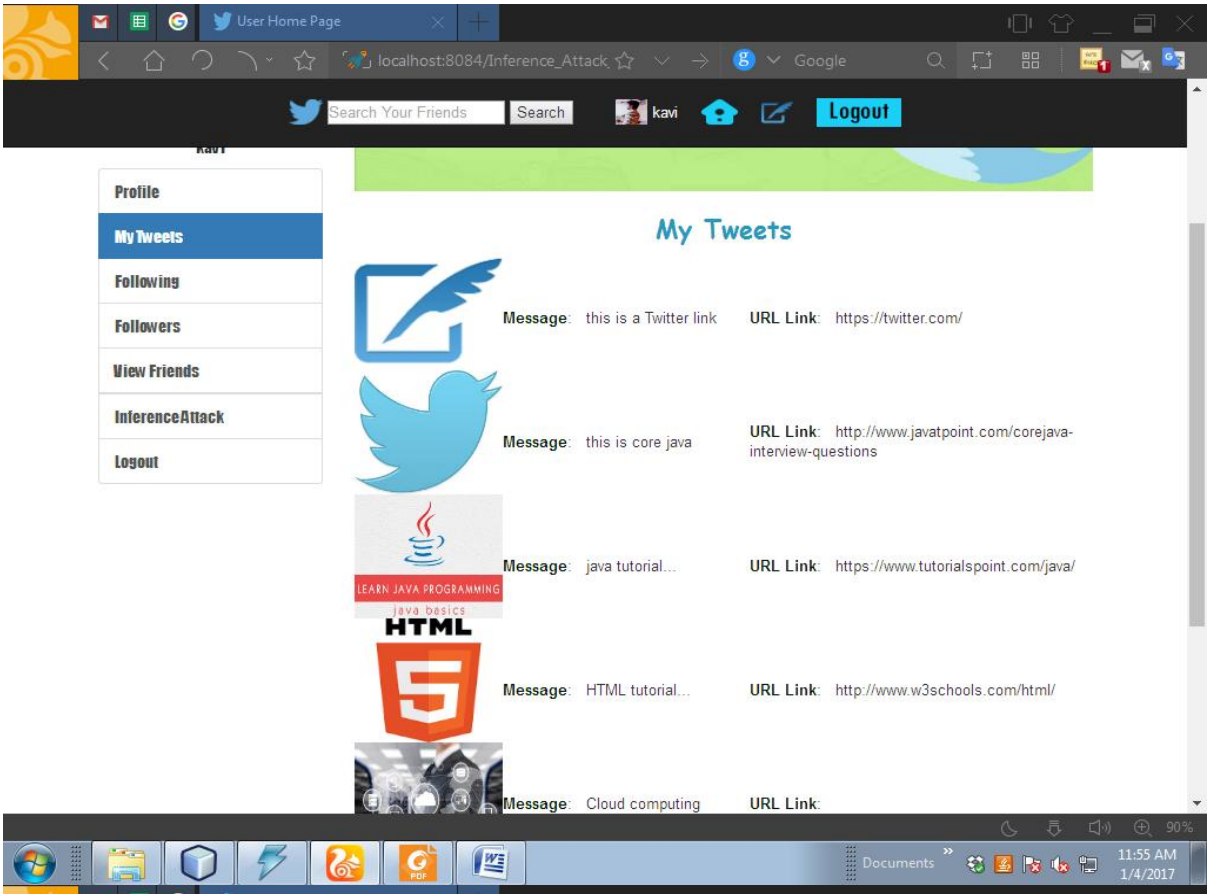
3. RESULTS AND DISCUSSION

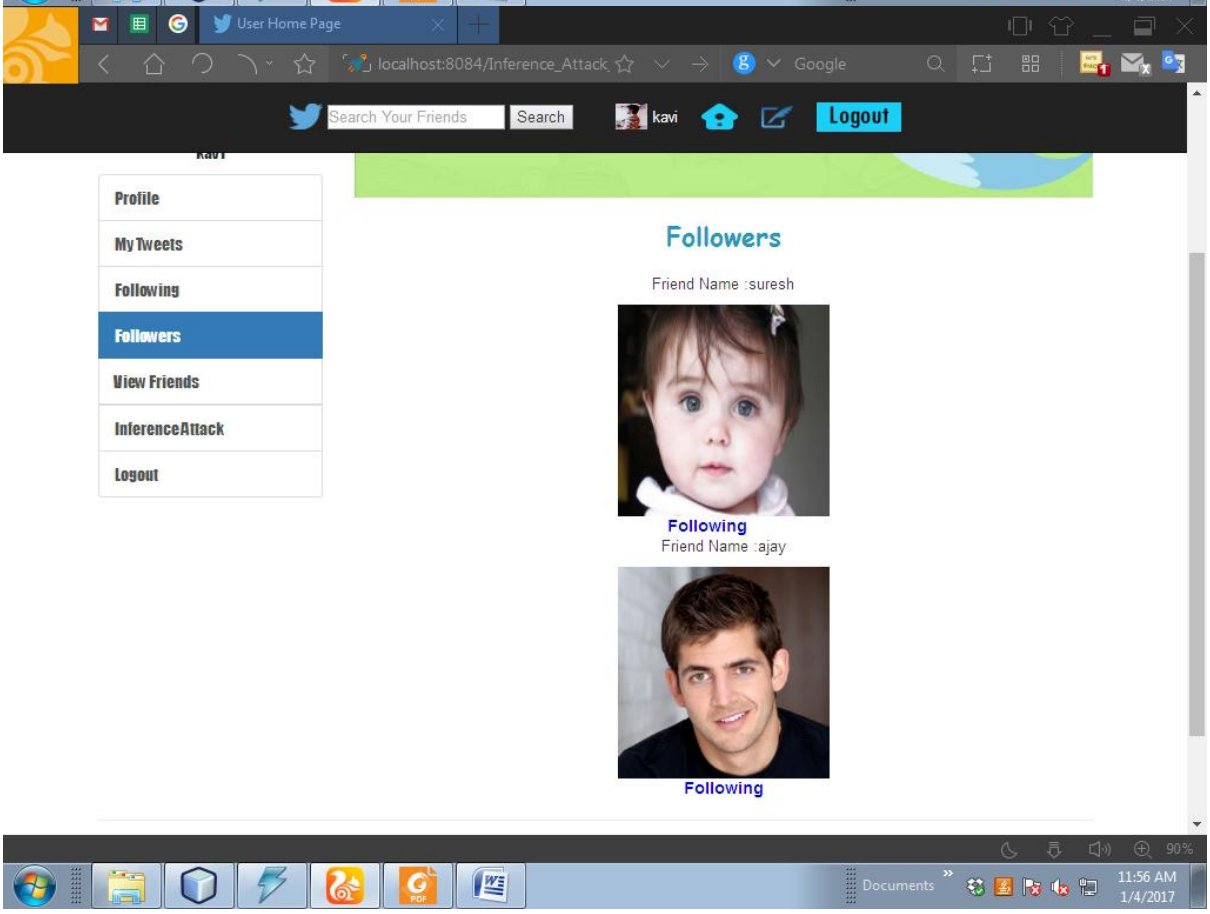


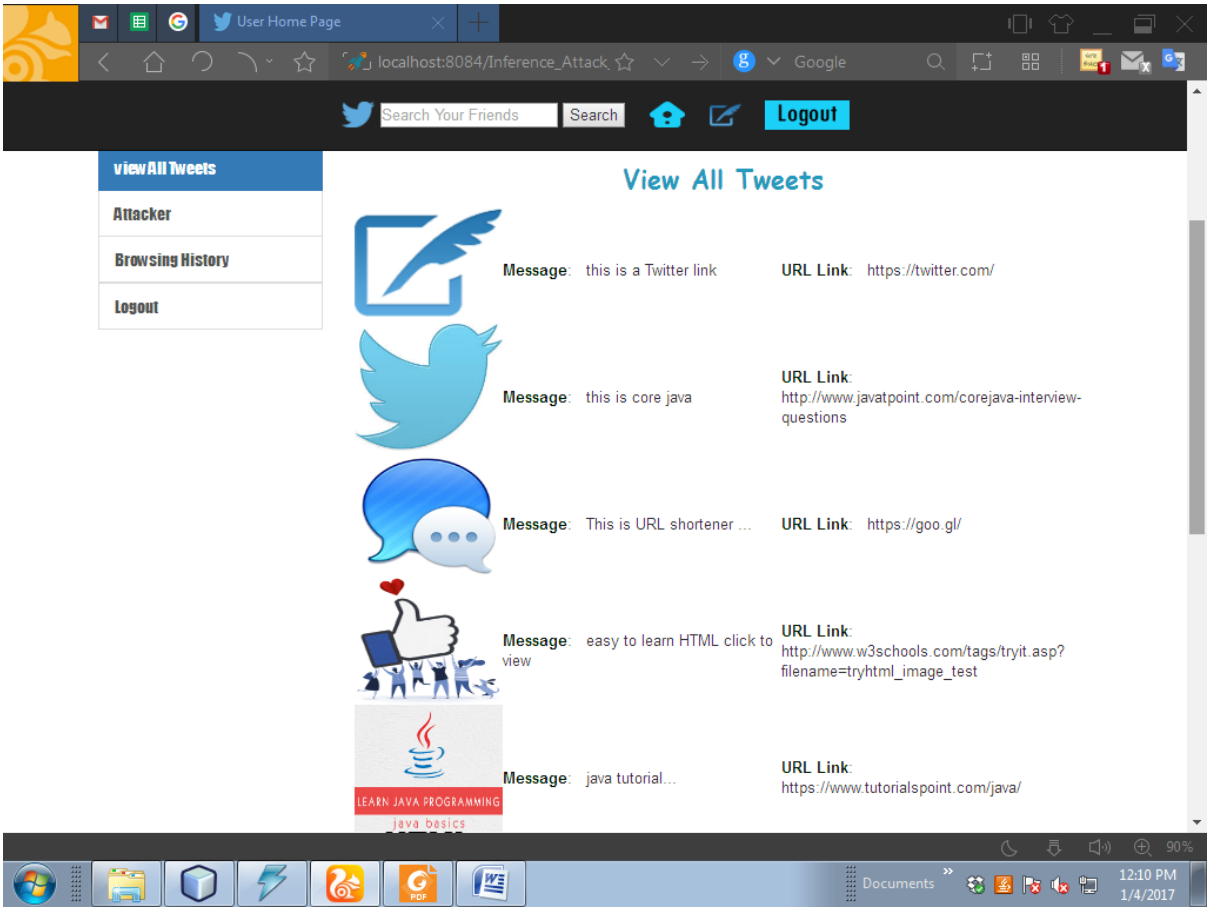


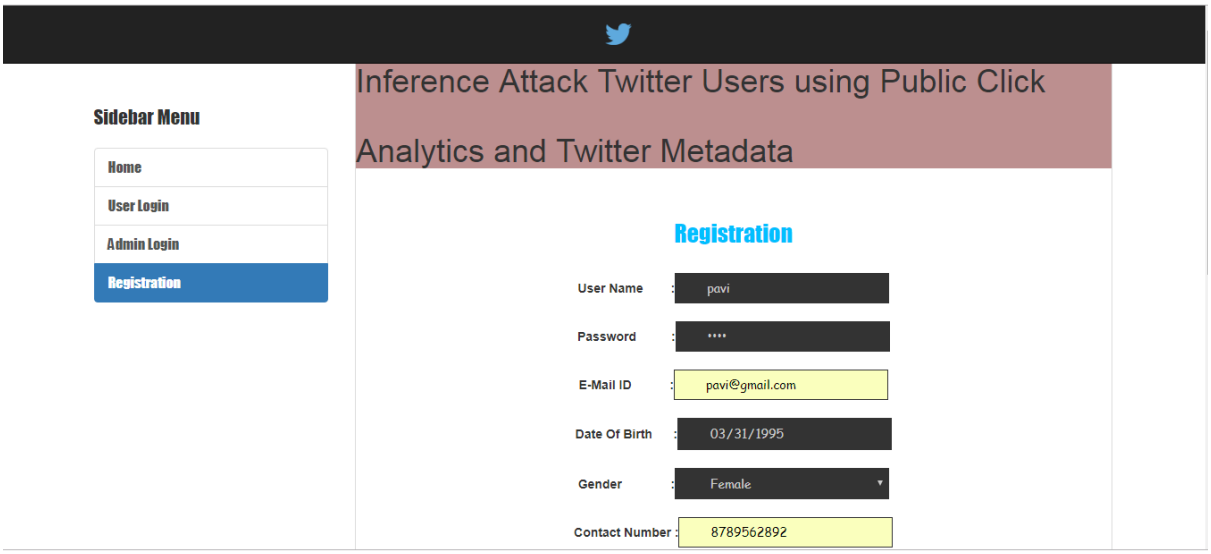

User Home Page
localhost:8084/Inference_Attack
Google
Search Your Friends
Search
ajay
Logout
Posted By: kavi
Message: java tutorial...
Click this Link : https://www.tutorialspoint.com/java/
Posted By: suresh
Message: easy to learn HTML click to view
ick this Link : http://www.w3schools.com/tags/tryit.asp?filename=tryhtml_image_test
Posted By: Abi
Message: This is URL shortener ...
Click this Link : https://goo.gl/
Posted By: kavi
Message: this is core java
Click this Link : http://www.javatpoint.com/corejava-interview-questions
Posted By: kavi
Message: this is a Twitter link
Click this Link : https://twitter.com/


User Home Page
localhost:8084/Inference_Attack
Google
Search Your Friends
Search
ajay
Logout
Profile
My Tweets
Following
Followers
View Friends
InferenceAttack
Logout
Following
Friend Name :kavi
Following
Friend Name :Abi
Following


User Home Page
localhost:8084/Inference_Attack
Google
Search Your Friends
Search
kavi
Logout
Profile
My Tweets
Following
Followers
View Friends
InferenceAttack
Logout
My Tweets
Message: this is a Twitter link
URL Link: https://twitter.com/
Message: this is core java
URL Link: http://www.javatpoint.com/corejava-interview-questions
LEARN JAVA PROGRAMMING
java basics
Message: java tutorial...
URL Link: https://www.tutorialspoint.com/java/
HTML
Message: HTML tutorial...
URL Link: http://www.w3schools.com/html/
Message: Cloud computing
URL Link:
Documents
11:55 AM
1/4/2017


User Home Page
localhost:8084/Inference_Attack
Google
Search Your Friends
Search
kavi
Logout
Profile
My Tweets
Following
Followers
View Friends
InferenceAttack
Logout
Followers
Friend Name :suresh
Following
Friend Name :ajay
Following
Documents
11:56 AM
1/4/2017





4. CONCLUSION

Inference attacks to infer which shortened URLs clicked on by a target user. All the information needed in our attacks is public information: the click analytics of URL shortening services and Twitter meta data. 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.

5. FUTURE WORK

A hybrid approach exploiting community based feature with Metadata content and interaction based feature for detecting automated spammers in twitter . Generally planted in OSNs for varied purposes but absence of real

life identity hinders them to join the trust network of benign users. Attaining perfect accuracy in spammer detection is extremely difficult and accordingly any feature set can never be considered as complete and sound as spammers keep on changing their operating behavior to evade detection mechanism. Unlike existing approaches of characterizing spammers based on their own profiles, the proposed approach lies in the characterization of a spammer based on its neighboring nodes and their interaction network.

REFERENCES

[1] L. Backstrom, C. Dwork, and J. Kleinberg. Wherefore art thour3579x? anonymized social networks, hidden patterns, and structural steganography. In Proc. 16th Int'l World Wide Web Conf. (WWW),2007.

[2] D. boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In Proc. 43rd Hawaii International Conference on System Sciences (HICSS), 2010.

[3] Bugzilla. Bug 57351: css on a:visited can load animage and/or reveal if visitor been to a site, 2000.https://bugzilla.mozilla.org/show bug.cgi?id=57351.

[4] Bugzilla. Bug 147777: visited support allows queries into global history, 2002. https://bugzilla.mozilla.org/show bug.cgi?id=147777.

[5] J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten, andV. Shmatikov. "you might also like:" privacy risks of collaborative filtering. In Proc. IEEE Symp. Security and Privacy (S&P), 2011.

[6] A. Chaabane, G. Acs, and M. A. Kaafar. You are what you like! information leakage through users' interests. In Proc. 19th Network and Distributed System Security Symp. (NDSS), 2012.

[7] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet:A content-based approach to geo-locating twitter users. In Proc.19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.

[8] A. Clover. Css visited pages disclosure, 2002.http://seclists.org/bugtraq/2002/Feb/271.

[9] C. Dwork. Di_erential privacy. In Proc. 33rd International Colloquium on Automata, Languages and Programming (ICALP),2006.

[10] E. W. Felten and M. A. Schneider. Timing attacks on web privacy. InProc. 7th ACM Conf. Computer and Comm. Security (CCS), 2000.

[11] L. Grangeia. Dns cache snooping or snooping the cache for fun and profit. In Side Step Seguranca Digitial, Technical Report, 2004.

[12] J. He, W. W. Chu, and Z. V. Liu. Inferring privacy information from social networks. In Proc.4th IEEE international conference on Intelligence and Security Informatics (ISI), 2006.

[13] B. Hecht, L. Hong, B. Suh, and E. H. Chi. Tweets from justinbieber's heart: The dynamics of the location field in user profiles. In Proc. SIGCHI Conference on Human Factors in Computing Systems(CHI), 2011.

[14] C. Jackson, A. Bortz, D. Boneh, and J. C. Mitchell. Protecting browser state from web privacy attacks. In Proc. 15th Int'l WorldWide Web Conf. (WWW), 2006.

[15] M. Jakobsson and S. Stamm. Invasive browser sni_ng and countermeasures. In Proc. 15th Int'l World Wide Web Conf. (WWW),2006.

[16] A. Janc and L. Olejnik. Web browser history detection as a real world privacy threat. In Proc. 15th European conference on Research in computer security (ESORICS), 2010..