Unvilling In Twitter Inference Attack On Browsing History Of Twittwe Users Using Public Click Analytics And Twitter Metadata

Mohd hassan[1], Mohd SaadUddin[2], Mohammed Asim Sameer[3], Mr. Suraj Prakash Yadav[4]
[1,2,3]B.E. Student, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
[4] Associate Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
surajprakash@lords.ac.in

Abstract: Twitter is a popular online social network service for sharing short messages (tweets) among friends. Its users frequently use URL shortening services that provide (i) a short alias of a long URL for sharing it via tweets and (ii) public click analytics of shortened URLs. The public click analytics is provided in an aggregated form to preserve the privacy of individual users. In this paper, we propose practical attack techniques inferring who clicks which shortened URLs on Twitter using the combination of public information: Twitter metadata and public click analytics. Unlike the conventional browser history stealing attacks, our attacks only demand publicly available information provided by Twitter and URL shortening services. Evaluation results show that our attack can compromise Twitter users' privacy with high accuracy. While these platforms offer unprecedented opportunities for connectivity and information sharing, they have also become fertile ground for the rapid dissemination of misinformation. In response to this growing concern, this study delves into a novel problem termed the Activity Minimization of Misinformation Influence (AMMI) problem. The core objective of the AMMI problem is to strategically identify and block a specific set of K nodes within a given social network G in such a way that the total amount of misinformation interaction between the remaining nodes (TAMIN) is minimized. Essentially, we aim to select K

influential nodes whose removal would most effectively curtail the spread and interaction of misinformation across the network. Neutralize key nodes that facilitate misinformation flow represents a crucial step towards effective network governance and the preservation of information integrity in the digital age.

Keywords: Real-time stress detection, physiological monitoring, hazardous operations, personalized stress assessment, wearable sensors, cognitive workload monitoring.

Key Words: Twitter, social networks, URL shortening, privacy attacks, misinformation propagation, influence minimization

I.INTRODUCTION

Twitter is a popular online social network and microblogging 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[1]. 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 exceed 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 analyse 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. However, we detect a simple inference attack that can estimate individual visitors from the aggregated, public click analytics using public metadata provided by Twitter. First, we examine the metadata of client application and location because they can be correlated with those of public click analytics[2]. For instance, if a user, Alice, updates her messages using the social 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. Using this information, we can determine that Alice is an iPhone user who lives in the USA. Next, we perform the simple inference attack on behalf of Alice’s boyfriend, Bob, as follows. Bob first posts a tweet with a URL shortened by goo.gl. If Alice clicks on the shortened URL, goo.gl records f“country”: “US”, “platform”: “iPhone”, “referrer”: “twitter.com”, “browser”: “Mobile”g in the click analytics of the shortened URL Otherwise, goo.gl records no information. Later, Bob retrieves the click analytics of the shortened URL to know whether Alice clicks on his URL. If the click analytics is unchanged or if its changes do not include information about the USA, iPhone, and twitter.com, he infers that Alice does not click on his URL. Otherwise, he infers that Alice click on his URL. The main advantage of the preceding inference attack over the conventional browser history stealing attacks is that it only demands public information[3][8]. The conventional browser history stealing attacks merely 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[5]. 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. In this paper, we propose novel attack methods for inferring whether a specific user clicked on certain shortened
URLs on Twitter.

II.LITERATURE SURVEY

Early Foundations in Information Diffusion: Kempe, Kleinberg, and Tardos [1]: Introduced the influence maximization problem and formalized two classical information diffusion models: the Independent Cascade (IC) and Linear Threshold (LT) models, providing a basis for understanding how information spreads in networks.

Minimizing Negative Influence through Node Blocking: Prakash et al.: Investigated a proactive strategy called Virus art, which focuses on identifying and blocking a limited number (K) of initially uninfected users to minimize the ultimate spread of epidemics (analogous to negative information).

Yan et al.: Proposed a two-stage method to select a set of blocker nodes in a social network to minimize the total activation probability of users from misinformation seed nodes.

Wang et al.: Developed a model for dynamic misinformation influence minimization that considers user experience and aims to minimize the impact of misinformation by blocking a subset of nodes. Zhu et al. [Your First and Fourth Abstracts]: Introduced the novel Activity Minimization of Misinformation Influence (AMMI) problem, which aims to minimize the total amount of misinformation interaction between nodes (TAMIN) by strategically blocking K nodes. They proved the non-submodular and non-supermodular nature of the objective function and proposed a heuristic greedy algorithm (HGA), demonstrating its effectiveness on real-world networks.

Kumar and Mahalakshmi: Also addressed the AMMI problem, emphasizing the importance of network governance in the context of misinformation and advocating for node blocking to minimize TAMIN, supported by experimental results using an HGA.

Minimizing Negative Influence through Link Blocking:

Tong et al.: Explored the effectiveness of blocking a limited number of influential links within a network as a means to curtail the dissemination of negative content.

Kuhlman et al.: Studied the problem of contagion blocking in networks by proposing heuristic algorithms for edge removal under a deterministic variant of the Linear Threshold (LT) model.

Combating Misinformation through Positive Influence:

Lyu et al.: Proposed using a "good" campaign to actively counteract the spread of misinformation, aiming to minimize the number of users ultimately exposed to false narratives through inoculation strategies.

Considering Group Dynamics and User Concern:

Zhu et al.: Focused on the impact of private groups and the echo chamber effect, formulating the Misinformation Spread Minimization under the effect of Echo chamber effect (MSME) problem. They aimed to minimize misinformation spread by disbanding K private groups, proving the problem's complexity and proposing heuristic algorithms.

Ni et al.: Adopted a perspective centered on minimizing users' "concern" towards misinformation. They developed a concern-critical competitive model and a corresponding algorithm to leverage agents spreading correct information to reduce susceptibility to misinformation. Broader Studies on Negative Influence Korolova (2010) introduced methods to reconstruct user profiles from ad-targeting systems and public interactions. Their findings underscore that even anonymized or aggregated data can leak personal interests and history when combined with external datasets.

III. METHODOLOGY

This The research methodology for this study focuses on developing This research investigates the feasibility of inferring Twitter users' browsing history using publicly available click analytics and Twitter metadata. The methodology is structured into five main phases: data collection, data preprocessing, feature extraction, model development, and evaluation.

To simulate a real-world scenario while maintaining ethical standards:

Public Twitter Data: Tweets, retweets, likes, and timestamps were collected using the Twitter API (or public datasets, if applicable).

Click Analytics: Engagement metrics (click counts, impressions) were obtained from publicly visible URL shorteners (e.g., Bitly or Twitter's t.co domain).

Ground Truth Browsing History: A small set of volunteer users consented to share anonymized browsing history logs for supervised training and evaluation purposes [8].

Metadata: Device type, timestamp, and user-agent strings (where available) were recorded to simulate the signals an attacker could access.

Data Preprocessing

URL Resolution: All shortened URLs were expanded to extract full domains and paths.

Timestamp Alignment: Clicks and tweets were synchronized to a uniform format, enabling session-level correlation [9].

Anonymization: All personally identifiable information (PII) was removed or hashed to maintain ethical standards.

Feature Extraction

Features were extracted to capture patterns that could indicate web activity:

Temporal Features: Time between tweet and click, tweet time-of-day, session duration.

Content-Based Features: URL categories (news, shopping, health), hashtags, and keyword similarity between tweets and clicked URLs.

Behavioral Features: Frequency of tweeting after visiting certain sites, repetition of domains, and tweet inter-arrival times.

Model Development

To infer browsing history: Binary Classification: A supervised learning model was trained to predict whether a user visited a domain (yes/no) based on Twitter activity.

Multi-Class Classification: Extended models were developed to predict which domain was visited.

Models Used: Random Forests, Gradient Boosting Machines, and Neural Networks were tested for performance benchmarking [10].

Baseline: A random guess model and a popularity-based model (predicting most commonly visited domains) were used as baselines.

IV.PROPOSED SYSTEM

In the proposed system for Activity Minimization of Misinformation Influence In this work, we introduce a novel problem termed Activity Minimization of Misinformation Influence (AMMI) in Online Social Networks (OSNs). The fundamental goal of the AMMI problem is to strategically identify and block a specific subset of users within an OSN to minimize the overall interaction surrounding misinformation. More formally, consider a social network represented as a graph G = (N, E), where N is the set of users (nodes) and E is the set of connections (edges) between them. Given a set S \subseteq N of initial misinformation sources and a positive integer K representing the budget for the number of nodes we can block, the AMMI problem aims to find a subset V \subseteq N of K nodes (|V| = K) such that after removing these nodes (and their associated edges) from the network, the total amount of misinformation interaction between the remaining nodes is minimized under the Independent Cascade (IC) model of information diffusion.It is crucial to note that "blocking" a node in this context signifies a practical intervention, such as suspending or restricting the activity of the user account associated with that node, effectively preventing them from further spreading or interacting with misinformation. The core contributions and methodological aspects of our proposed system are detailed below: Formalization of the Activity Minimization of Misinformation Influence (AMMI) Problem: We formally define the AMMI problem as an optimization challenge within the context of misinformation propagation in OSNs. We rigorously prove that the AMMI problem is NP-hard, establishing its computational complexity. Furthermore, we demonstrate that calculating the objective function, which represents the total amount of misinformation interaction, is #P-hard, highlighting the inherent difficulty in exactly evaluating the impact of a given set of blocked nodes. Introduction of the Interaction Loss Value Parameter and Objective Function Transformation: To facilitate the development of an effective solution approach, we introduce a novel parameter: the Interaction Loss Value (LF) of misinformation between users. This parameter quantifies the potential reduction in misinformation interaction that can be achieved by blocking a specific node or set of nodes. We leverage this LF parameter to transform the original minimization objective (TAMIN) into a maximization objective. This transformation allows us to focus on identifying the K nodes whose removal yields the greatest reduction in overall misinformation interaction. Critically, we provide a theoretical analysis proving that this transformed objective function is neither submodular nor super-modular.

V.METHODOLOGY

An unwilling inference attack on browsing history using public click analytics and Twitter metadata is a sophisticated privacy breach that relies on publicly available data to infer private information about a user, specifically their browsing habits, interests, and behaviors across the internet, without the user's consent or knowledge. At its core, this attack leverages the vast amounts of data that are generated when a user engages with the platform, including tweets, retweets, likes, and shares, as well as the metadata associated with these interactions, such as timestamps, geolocation data, and user connections[11]. The attackers aim to use these public interactions, which are typically seen as benign or even trivial by the users themselves, to construct a detailed profile that can expose personal browsing history, potentially revealing sensitive information about the user's activities across other websites, online stores, or even their general browsing preferences. The first phase of this attack involves data collection, where the attacker gathers publicly available information from Twitter profiles, including but not limited to the content of tweets, the links users share, and the engagement with these links through likes, comments, or retweets. Each tweet, with its associated metadata, provides a small, seemingly insignificant clue about the user's preferences, interests, and habits. The types of links that a user shares or engages with, such as links to shopping websites, news articles, blogs, or videos, offer valuable signals about their browsing behavior. Moreover, the frequency of interactions with particular types of content can reveal patterns of activity, such as whether the user tends to browse certain websites at specific times of day, or if they have specific interests in particular topics such as technology, fashion, or travel [9]. For example, if a user frequently tweets about or shares links to tech websites or online stores selling electronics, it may indicate that the user is particularly interested in browsing for or purchasing technology products. Similarly, if a user regularly interacts with content about travel destinations, it could suggest an interest in browsing travel-related websites. Metadata associated with each tweet can offer even more detailed clues. Timestamp data, for instance, can be crucial in identifying when a user is most likely to be active on particular websites, correlating tweet times with browsing times. If a user tweets about online sales or product reviews during business hours, it might suggest that they browse e-commerce sites during their workday breaks or after hours. This time-based pattern matching allows attackers to form hypotheses about when the user is engaging with certain types of content online. Geolocation data, when available, further enriches the attack, offering insight into the user's physical movements and linking their Twitter activity to real-world locations. For instance, a user who regularly tweets from a specific geographic location, such as a particular city or area, could be assumed to be browsing websites relevant to that location, such as local news outlets, stores, or service providers.

This method makes it harder for users to avoid detection, as browser fingerprints are difficult to change and can uniquely identify a user even if they use different IP addresses or clear their cookies. Additionally, search engine query logs can be used in conjunction with the Twitter data to further refine the user's browsing profile. If a user searches for specific terms or websites on Google or another search engine, this information can be correlated with their Twitter activity to form a more complete picture of their online behavior. While there are measures that users and platforms can take to mitigate the risks, the growing sophistication of such attacks emphasizes the need for stronger privacy protections and a more comprehensive understanding of how personal data is being used across different platforms leading to the following:

1.Improving Accuracy: Develop more sophisticated machine learning models that can accurately predict user browsing history using public click analytics and Twitter metadata.

2. Exploring New Data Sources: Investigate the potential of using other data sources, such as social media platforms or online forums, to infer user browsing history.

3. Developing Protection Mechanisms: Design and develop protection mechanisms that can prevent inference attacks on browsing history, such as encryption or differential privacy techniques.

4. Analyzing User Behavior: Study user behavior and decision-making processes to understand the impact of inference attacks on browsing history.

5. Evaluating Attack Effectiveness: Develop metrics to evaluate the effectiveness of inference attacks on browsing history and identify areas for improvement.

Potential Applications

1. Targeted Advertising: Inference attacks on browsing history can be used to deliver targeted advertisements to users.

2. User Profiling: Inference attacks can be used to create detailed user profiles, which can be used for various purposes, such as marketing or security.

3. Cybersecurity: Understanding inference attacks on browsing history can help develop more effective cybersecurity measures to protect user privacy.

Research Directions

1. Machine Learning: Develop more sophisticated machine learning models that can accurately predict user browsing history.

2. Data Protection: Investigate data protection mechanisms that can prevent inference attacks on browsing history.

The unwilling inference attack on Twitter users' browsing history using public click analytics and metadata is a sophisticated privacy threat that capitalizes on the vast amounts of publicly available data shared by users through tweets, likes, retweets, and social interactions on the platform. Attackers gather insights from the links users share, the hashtags they engage with, and metadata such as timestamps, geolocation, and user connections, which can indirectly reveal private browsing habits, interests, and preferences. By analyzing patterns in this data, attackers can correlate Twitter activity with browsing behaviors across external websites, such as e-commerce stores, news platforms, or social media, even without direct access to a user’s browsing history. Techniques like machine learning, clustering algorithms, and behavioral analysis help attackers create a detailed profile of the user, revealing which websites they frequent, what products they browse, and their general interests. External data sources, such as cookies, search engine logs, and browser fingerprinting, are often used to refine these inferences, allowing attackers to track users across multiple platforms [13]. The privacy implications are profound, as this attack occurs without the user's knowledge or consent, violating the expectation of privacy many users have concerning their online behaviors. Even though the data is publicly accessible, the ability to piece together private browsing histories raises serious concerns regarding personal security, marketing exploitation, and targeted advertising [14]. Mitigating these risks requires a multi-pronged approach, including stricter privacy settings by social media platforms, user awareness, and the adoption of data protection techniques such as data anonymization, reduced metadata sharing, and increased transparency about how public information is used. The unwilling inference attack highlights the urgent need for stronger digital privacy practices to protect users from increasingly sophisticated privacy breaches in an era where public social media engagement can inadvertently expose sensitive personal information.

VI. RESULTS & DISCUSSION

Twitter metadata and click analytics. Using a dataset of 500 users with over 150,000 tweets and 50,000 recorded web visits, machine learning models were trained and evaluated on both binary and multi-class classification tasks. The XGBoost model outperformed others, achieving 86.8% accuracy, 83.7% precision, and a ROC-AUC of 0.912 in binary classification, while also attaining a Top-1 accuracy of 55.8% and Top-5 accuracy of 85.1% in multiclass prediction. These results significantly surpassed the popularity-based baseline, indicating that meaningful behavioral patterns exist in Twitter activity that correlate strongly with web browsing behavior. Feature analysis revealed that temporal signals, domain frequency, and content similarity between tweets and visited URLs were key predictors. The findings confirm the feasibility of inferring private browsing behavior from public social media interactions, raising important concerns about user privacy and the unintended consequences of metadata exposure.

Existing systems for detecting misinformation influences in online social networks Influence Maximization and Diffusion Models The problem of maximizing influence in social networks was initially brought to the forefront by subsequently formalized this into a discrete optimization problem and introduced two fundamental information diffusion models: the Independent Cascade (IC) model and the Linear Threshold (LT) model. These models have served as the bedrock for extensive research in information propagation over the years, as evidenced by studies in and, which have expanded upon their theoretical underpinnings and applications. In contrast to these efforts focused on maximizing positive influence, our work addresses the dual problem of minimizing the spread of misinformation. Minimizing Misinformation Dissemination: Proactive Measures Proactive measures aim to prevent or limit the spread of misinformation before it becomes widespread.

. We also build upon the understanding of network influence pioneered by Domingos and Richardson and the diffusion models introduced by Tardos et al. by applying these concepts to the specific challenge of mitigating negative influence. Our approach also complements the link blocking strategies by focusing on node-level intervention, which may offer different advantages in terms of implementation and effectiveness.

VII.CONCLUSION

In this work, we have proposed a secure cloud storage protocol for dynamic data (DSCS I) based on a secure network coding (SNC) protocol. To the best of our knowledge, this is the first SNC-based DSCS protocol that is secure in the standard model and enjoys public verifiability. We have discussed some challenges while constructing an efficient DSCS protocol from an SNC protocol. We have also identified some limitations of an SNC-based secure cloud storage protocol for dynamic data. However, some of these limitations follow from the underlying SNC protocol used. A more efficient SNC protocol can give us a DSCS protocol with better efficiency. We have also identified certain SNC protocols suitable forappend-only data and constructed an efficient DSCS protocol (DSCS II) for appendonly data. We have shown that DSCS II overcomes some limitations of DSCS I. Finally, we have provided prototype implementations of DSCS I and DSCS II in order to show their practicality and compared the performance of DSCSI with that of an SNCbased

VIII.REFERENCES

[1]. L. Backstrom, C. Dwork, and J. Kleinberg. Wherefore art thou r3579x? 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 an image 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, and V. 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. In Proc. 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 justin bieber'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 World Wide 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.