ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Protecting Privacy on Online Social Networks Based on Anonymization and Graph Structure.

M.Bhargav Kumar
B.Tech Student,
Department of CSE,
MLRIT.

Sheikh Gouse
Associate Professor,
Department of CSE,
MLRIT.

K.Sumathi
B.Tech Student,
Department of CSE,
MLRIT.

D.Shravya
B.Tech Student,
Department of CSE,
MLRIT.

S.Manideep
B.Tech Student,
Department of CSE
MLRIT.

U.Chandra Devi
B.Tech Student,
Department of CSE,
MLRIT.

Abstract:

Online Social networks are web-based services that allow individuals to create a public profile, to create a list of users with whom to share connections, and view and cross the connections within the system. Social networking systems may record activities of individuals, with data becoming a life stream. Such usage of social media and roaming services allow digital tracing data to include individual interests, social groups, behaviours, and location.

Such data can be gathered from sensors within devices, and collected and analyzed. Digital footprints are not a digital identity or passport, but the content and meta data collected impacts upon internet privacy, trust, security, digital reputation, and recommendation. In this paper we address Protecting Privacy on Online Social Networks Based on anonymization and Graph structure.

The algorithm mainly recognizes a seed sub-graph, either planted by an attacker or revealed by a involvement of a small group of users, and then breeds the seed larger grounded on the attacker's prevailing knowledge of the users' social relations. In this paper we identify and relaxes inherent norms taken by earlier works, removes random parameters, and advances identification efficiency and precision.

Keywords:

OSN, Anonymization, Grapgh Structure, User Protection, Privacy.

Introduction:

A social networking service is a platform to build social networks or social relations among people who share interests, activities, backgrounds or real-life connections. A social network service consists of a representation of each user (often a profile), his or her social links, and a variety of additional services. Social networks are web-based services that allow individuals to create a public profile, to create a list of users with whom to share connections, and view and cross the connections within the system. Most social network services are web-based and provide means for users to interact over the Internet, such as e-mail and instant messaging. Social network sites are varied and they incorporate new information and communication tools such as mobile connectivity, photo/video/sharing and blogging.

On the Internet a digital footprint is the word used to describe the trail, traces or "footprints" that people leave online. This is information transmitted online, such as forum registration, e-mails and attachments, uploading videos or digital images and any other form of transmission of information — all of which leaves traces of personal information about yourself available to others online.There are two main classifications for digital footprints: passive and active. A passive digital footprint is created when data is collected without the owner knowing, whereas active digital footprints are created when personal data is released deliberately by a user for the purpose of sharing information about

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

oneself by means of websites or social media. Passive digital footprints can be stored in many ways depending on the situation. In an online environment a footprint may be stored in an online data base as a "hit". This footprint may track the user IP address, when it was created, and where they came from; with the footprint later being analyzed. In an offline environment, a footprint may be stored in files, which can be accessed by administrators to view the actions performed on the machine, without being able to see who performed them.

Active digital footprints can also be stored in many ways depending on the situation. In an online environment, a footprint can be stored by a user being logged into a site when making a post or change, with the registered name being connected to the edit. In an off line environment a footprint may be stored in files, when the owner of the computer uses a keylogger, so logs can show the actions performed on the machine, and who performed them. One of the features of keylogger is to monitor the clipboard for any changes as the user will sometimes have a very good habit of copying and pasting passwords and taking screenshots.

Existing System:

Digital traces left by users of online social networking services, even after anonymization, are susceptible to privacy breaches. This is exacerbated by the increasing overlap in user-bases among various services. To alert fellow researchers in both the academia and the industry to the feasibility of such an attack.

Disadvantages of Existing System:

1. Although a trade-off between utility and privacy is necessary, it is hard, if not impossible, to find a proper balance overall. Besides, it is hard to prevent attackers from proactively collecting intelligence on the social network.

2. It is especially relevant today as major online social networking services provide APIs to facilitate third party application development. These programming interfaces can be abused by a malicious party to gather information about the network.

Proposed System:

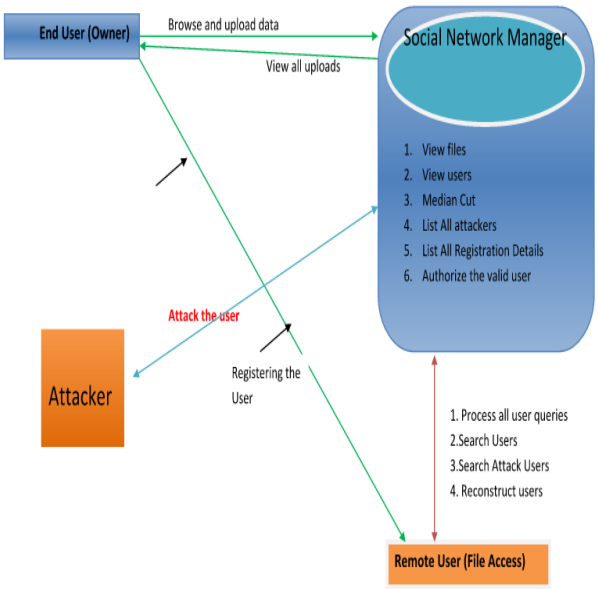
We propose an algorithm, Seed-and-Grow, to identify users from an anonymized social graph, based solely on graph structure. The algorithm first identifies a seed sub-graph, either planted by an attacker or divulged by a collusion of a small group of users, and then grows the seed larger based on the attacker's existing knowledge of the users' social relations. Our work identifies and relaxes implicit assumptions taken by previous works, eliminates arbitrary parameters, and improves identification effectiveness and accuracy. Simulations on realworld collected datasets verify our claim.

Advantages of Proposed System:

1. This algorithm automatically finds a good balance between identification effectiveness and accuracy.

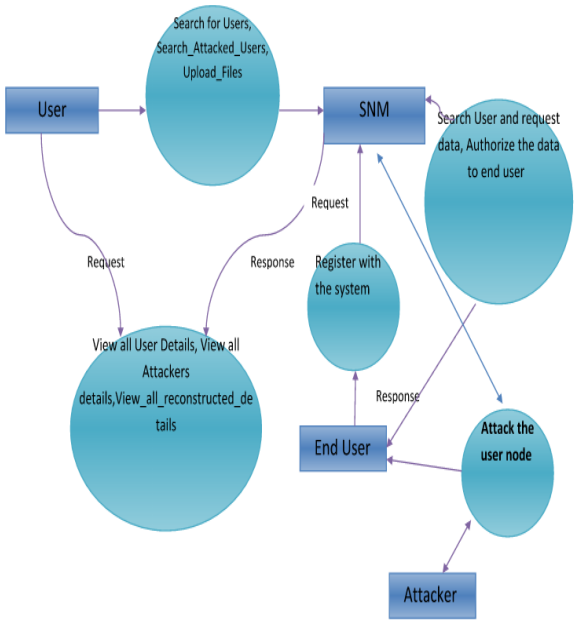
2. Although a trade-off between utility and privacy is necessary, it is hard, if not impossible, to find a proper balance overall. Besides, it is hard to prevent attackers from proactively collecting intelligence on the social network.

Architecture:



ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Data Flow:



Implementation:

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

1.User Module:

In this module, Users are having authentication and security to access the detail whoch is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first.

2.Initial Speed Size:

Recent literature on interaction-based social graphs(e.g., the social graph in the motivating scenario) singles out the attacker's interaction budget as the major limitation to attack effectiveness. The limitation translates to 1) the initial seed size and 2) the number links between the fingerprint graph and the initial seed. Our seed algorithm resolves the later issue by guaranteeing unambiguous identification of the initial seed, regardless of link numbers. As showm below, our grow algorithm resolves the former issue by working with a small initial seed.

3.Grow Algorithm:

At the core of the grow algorithm is a family of related metrics, collectively known as the dissimilarity between a pair of vertices from the target and the background graph, respectively. In order to enhance the identification accuracy and the reduce the computation complexity and the false-positive rate, we introduce a greedyheuristic with revisiting into the algorithm. It is natural to start with those vertices in GT which connect to the initial seed VS because thet are more close to the certain information, i.e., the already identified vertices VS. For these vertices, their neighboring vertices can be divided into two groups.

4.Re-Visiting:

The dissimilarity metric and the greedy search algorithm for optimal combination are heuristic in nature. At an early stage with only a few seeds, there might be quite a few mapping candidates for a particular vertex in the background graph; we are very likely to pick a wrong mapping no matter which strategy is used in resolving the ambiguity. If left uncorrected, the incorrect mappings propagate through the grow process and lead to large-scale mismatch. We address this problem by providing a way to reexamine previous mapping decisions, given new evidences in the growth algorithm; we call this revisiting. More concretely, for each iteration , we consider all vertices which have at least one seed neighbor, i.e., those pairs of vertices on which the dissimilarity metrics in are well-defined. We expect that the revisiting technique will increase the accuracy of the algorithm. The greddy heuristic with revisiting is summarized in Algorithm.

Conclusion:

In this paper, we implement an algorithm, Seed-and-Grow, to recognize users from an anonymized social graph.

Our algorithm exploits the growing overlapping userbases among services and is based solely on social graph structure. The algorithm first identifies a seed sub-graph, either planted by an attacker or divulged by collusion of a small group of users, and then grows the seed larger based on the attacker's existing knowledge of the users' social relations. We classify and relax implicit assump-tions for unmistakable seed identification taken by pre-vious works, eliminate arbitrary parameters in grow algorithm, and demonstrate the superior performance over previous works in terms of identification effective-ness and accuracy by simulations on realworld-collected social-network datasets.

References:

[1] Wei Peng ; Feng Li ; XukaiZou ; Jie Wu, A Two-Stage Deanonymization Attack against Anonymized Social Networks, IEEE Transactions on Computers (Volume:63 , Issue: 2)

[2] A. Narayanan and V. Shmatikov, "De-anonym zing social networks," in Proc. IEEE S&P, 2009.

[3] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography," in Proc. ACM WWW, 2007.

[4] M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava, "Anonymizing social networks," Univ. Massachusetts, Amherst, Tech. Rep., 2007.

[5] E. Zheleva and L. Getoor, "Preserving the privacy of sensitive relationships in graph data," in Proc. ACM SIGKDD, 2007.

[6] A. Korolova, R. Motwani, S. Nabar, and Y. Xu, "Link privacy in social networks," in Proc. ACM CIKM, 2008.

[7] B. Zhou and J. Pei, "Preserving privacy in social networks against neighborhood attacks," in Proc. Intl. Conf. on Data Engineering(ICDE). IEEE, 2008.

[8] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis, "Resisting structural re-identification in anonymized social networks," VLDBEndowment, vol. 1, no. 1, pp. 102–114, 2008.

[9] J. Scott, Social network analysis: a handbook. SAGE Publications, 2000.

[10] K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Incognito: efficient full-domain k-anonymity," in Proc. ACM ICMD, 2005.