SURVEY ON PRIVACY PRESERVED METHODS FOR SOCIAL NETWORKING IN CLOUD COMPUTING

Sure Suresh[1], K Devika Rani [2]

[1]Pursuing M.tech (CSE), [2]Assistant Professor,
Nalanda Institute of Technology (NIT) Siddhartha Nagar, Kantepudi(v), Sattenapalli Guntur-522438.

ABSTRACT

Now a day companies would publish social networks to the third party for example the cloud service provider (CSP) for marketing reasons. The preserving privacy when publishing the social network data becomes a important issue. In this scenario we identify the novel type of the privacy attack termed 1-neighborhood attack. In this we assume that theattacker has knowledge about a degree of the target one hop neighbors in the addition to target's 1 neighborhood graph which consists of one-hop neighbors of a target and a relationships among these neighbors. With this information the attacker may re-identify a target from the k anonymity social network with the probability higher than 1/k where any node's 1-neighborhood graph is the isomorphic with k−1 other nodes' graphs. To the resist the 1*-neighborhood attack we define the key privacy property probability in the distinguishability for the outsourced social network and the proposed the heuristic indistinguishable group anonymization (HIGA) scheme to generate the anonymized social network with this privacy property. An empirical study indicates that an anonymized social networkcould still be used to the answer aggregate queries with high accuracy.*

Index Terms: The cloud computing, social networks, privacy and probability in distinguishability.

I INTRODUCTION

The social networks have developed rapidly recent research has been begun to the explored social networks to the recognized their structurethe advertising and marketing and the data mining. The cloud computing as the emerging computing paradigm is the expected to the reshape information technology processes in near future. The cloud services which are the available in the payas you-go manner promise ubiquitous 24/7 access at the low cost. Because of those overwhelming merits of the cloud computing, e.g. flexibility and the scalabilitymaximum organizations that host the social network data choose to the outsourced portion of their data to the cloud environment. The preserving privacy when publishing social network data becomes the important issue. The social networks model social relationships with the graph structure using the nodes and edges where nodes model individual social actors in the network, and the edges model relationships amongthe social actors. The relationships interpreted among social actors are often private and the directly outsourcing social networks to the cloud may result in the unacceptable disclosures. F.epublishingthe social network data that describes the set of the social actors related by the sexual contacts or the shared drug injections may compromise privacy of the social

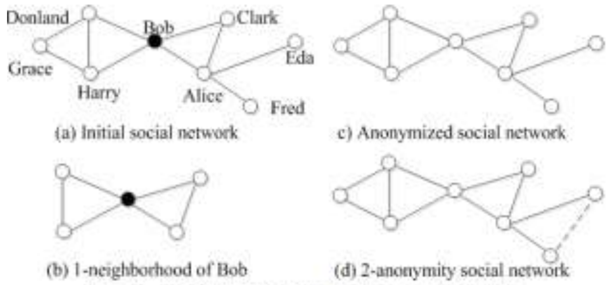
Donland
Bob
Clark
Grace
Eda
Harry
Alice
Fred
(a) Initial social network
c) Anonymized social network
(b) 1-neighborhood of Bob
(d) 2-anonymity social network

Fig. 1. 1-neighborhood attacks in a social network.

Actors involved. So that existing the research has proposed to the anonymized the social networks just before outsourcing. The natıve approach is to the simply anonymizedan identity of social actors just before outsourcing. However the attacker that has some knowledge about the targets neighborhood especially the one-hop the neighborhoodcould still re- identify target with the high confidence. This is attack termed 1-neighborhood attack is a proposed by Zhou et al.

1) In this we identify the novel attack 1*-neighborhood attack for outsourcing social networks to the cloud.

2) In this we define probabilistic in-distinguishability property for the outsourced social network and the propose the heuristic in-distinguishable group the anonymization scheme (HIGA) to the generate social networks with this is privacy property.

3) In this we conduct experiments on the both synthetic and real data sets to the verify effectiveness of the proposed scheme.

II PRELIMINARIES

The system Model: In this scenario we take asa system that consists of the publisher the cloud service provider, the attacker and more users. Publisher such as the Facebook or Twitter outsources the social network to thecloud. In our system the social network is modeled as the undirected and unlabeled graph

G = (V (G), E(G)), where V (G) is the set of nodes and E(G) ⊆ V (G) × V (G) is the set of the edges. Node identities are the assumed to be removed. Attacker has been certain background knowledge about target and he tries to re-identify target by the analyzing outsourced social network. The protect a privacy of a social actors in the network from attacker the publisher anonymized G to G = (V (G), E(G)) early outsourcing. As in we assume that every node in the G exists in the G and no fake nodes are added in G to the preserve global structure of the social network. As previous work we allow edges {(u, v)} ∈ E(G) to be removed from E (G).

Attack Model:Assume that the attacker is many interested in privacy of the social actors. Previously launching the attack the attacker needs to the collect some background knowledge about a target victim. Assume that the attacker may have background knowledge about the 1*-neighborhood graph of some targets. Informally the targets 1*-neighborhood graph consists of the both the 1-neighborhood graph of the target and the degrees of target’s one hop neighbors. The following the work in for each node u ∈ V (G) the related 1-neighborhood graphdesignated as Gu, is defined as follows: 1-Neighborhood Graph. Gu = (Vu, Eu) where Vu characterizes a set of the nodes {v|(u, v) ∈E(G)

∨ (v = u)} and Eu denotes the set of the edges {(w, v)|(w, v) ∈ E(G) ∧ {w, v} ∈ Vu}. Every node u ∈ V (G) the related 1*-neighborhood graph, represented as G∗ u, is defined as follow as 1*-Neighborhood Graph. G∗ u = (Gu, Du) where Gu is the 1-neighborhood graph of node u and Du is the sequence of the degrees of u one-hop neighbors.

The Random Walk (RW) based Approximate Matching:Stimulated by the work in We use random walk based approximate matching as building block of our HIGA scheme. The random walk (RW) is known as a beneficialdevice to achieve the steady state distribution for a graph mentioned to as topological signatures which provide the foundation for approximate matching. Specifically given graph G = (V (G), E(G))

where V (G) = {u1,...,u|V (G)|} A RW on G permits the probability puj (t) of the node uj∈ V (G) being located at the time t to be computed with Eq.1: puj (t) = ui∈V (G) 1 |V (G)| · (1 − d) · pui (t − 1) + ui∈N(uj) 1 |N(ui)| · d · pui(t − 1) (1) where |V (G)| is a number of the nodes in G, |N(uj)| is the number of one-hop neighbors for node uj , and d is the damping factor which defines the probability of straight jumping or negotiating.

Design Goals: The highest design goal of our work is to decrease the probability of a social actor being re-identified while publishing social networks to the cloud. Specifically given the social graph G we want to generate the anonymized graph G for that the following requirements are satisfied:

• Privacy Given any targets 1-*neighborhood graph, the attacker can't re-identify target from the anonymized social network with the confidence higher than the threshold. • Usabilitythe anonymized social networks can be used to the answer aggregate queries with the high accuracy.

III SCHEME OVERVIEW

Definitions: To preserve privacy previous research anticipated to make any node's 1-neighborhood graph be isomorphic with at least k−1 others. In much case isomorphism is the strong condition that is not necessary for the anonymzing the graph. We define the concept of the probabilistic in distinguishability which can preserve privacy at the lower anonymization cost.

Let G_u^* and $G_u^{'*}$ denote the 1*-neighborhood graph of node u in the unique social network G and in the anonymized social network $G^{'}$correspondingly. Probabilistic indistinguishability can be defined in a hierarchical way as tracks:Node In distinguishability. Nodes u and v are indistinguishable if an observer can't decide whether or not $G_u^* \neq G_v^*$ v in the original graph G, by comparing$G_u^*$and $G_v^{'*}$ v in an anonymized graph $G^{'}$. Group In distinguishability. For a group of nodes g = {v|v∈ V (G)} and |g| ≥ k if for each pair of nodes {u, v |u, v ∈ g}, u and v are indistinguishable in the published graph $G^{'}$, group g is an indistinguishable group. Probabilistic Indistinguishability. A published social network $G^{'}$ achieves probabilistic in distinguishability, if all{v|v∈ V (G)} can be classified into m ≥ 1 groups, where each group has the property of group in distinguishability.

Problem Definition: Specified a network graph G = (V (G), E(G)) and a positive integer k, derive an anonymized graph $G^{'}$ = (V ($G^{'}$) E($G^{'}$)) to be published such that V 1)($G^{'}$) = V (G);(2) G is probabilistic indistinguishable with respect to G; (3) the anonymization from G to $G^{'}$ has minimal anonymization cost.Intuition:The uncomplicated idea of the heuristic indistinguishable group anonymization (HIGA) scheme consists of four steps:

Step 1: The grouping. We classify nodes whose 1*- neighborhood graphs satisfy certain metrics into groups, where each group size is at least equal to k. • Step 2: Testing. We use random walk (RW) to test whether the 1-neighborhood graphs of nodes in a group approximately match or not. • Step 3: Anonymization. We use a heuristic anonymization algorithm to make the 1-neighborhood graphs of nodes in each group approximately match. • Step 4: Randomization. We randomly modify the graph with certain probability to make each node's 1*- neighborhood graph be changed with certain probability.

IV EXISTING SYSTEM

In social network models social relationship between graphs structured using the node and the edges where nodes model individual social actors in the network and the edges model relationships among social actors. The relationships among social actors are often the private and the directly outsourcing social networks to the cloud may result in the unacceptable disclosures. F.e publishing the social network data that describes the set of the social actors related by the sexual contacts or shared drug injections may compromise a privacy of asocial actor involved. Hencethe existing research has been proposed to the anonymizedthe social network previously outsourcing.

Disadvantages:

1) The user can only explicitly specify the group of user who can or can't access location information.

2) To access control policy supportedthe binary choices only which means users can only choose to allow or restrictinformation disclosures. The existing control of the strategies also suffered from the privacy leakage in terms of server storage.

V PROPOSE SYSTEM

To the permit useful analysis on a social networks while preserving privacy of social actors involved we can define the key privacy property probabilistic in distinguishability for the outsourced social networks to the generate an anonymizedsocial network with such the property we propose the heuristic indistinguishable group anonymization(HIGA) scheme. The basic idea consider as four key steps.One is that Grouping we group nodes whose I*- neighborhood graph of whichever pair of nodeapproximately match or not. Anonymization we propose the heuristic anonymization algorithm to make whichever nodes 1-neighborhood graph approximately match those of other nodes in the group by either adding or the removing edges. The randomization we randomly modify graph structurewiththe certain probability to make sure each 1* neighborhood graph has the certain probability of being different from original one.

Advantages:In this scenario we can identify the novel 1*-neighborhood attack. To resist this attack we define the key property probabilistic indistingguishability for the outsourcedsocial network sand we propose the heuristic anonymization scheme to the anonymized social network with this property.

VI RELATED WORK

The outsourcing privacy-preserving social networks are to the cloud environment. The research in this area is still in its infancy. The best of our knowledge and work by is the first to address this problem. The work that is closest to our work can be found in the publishing privacy-preserving social networks. As the pioneering work Backstromdiscussed two the re-identification attacks in natıve anonymized social networks. In the active attacker the attacker not unintentionally embeds thesub graph into the social network before publishing and the uses such kinds of the background knowledge to the re-identify nodes and the edges in the published network. In passive attacks the attacker with known ledged of the targetsub graph can infer identity of the nodes in a published network. Hence they don't provide the solution to the counter these attacks. To defend the re-identification attacks the work in advocated k-anonymity model where each node should be indistinguishable with at least k other nodes in the terms of both the attributes and associated structural information such as the neighborhood and node degree. To preserve scale and local structures of original graph, existing the anonymization approachestry to locally modify the graph structure to achieve privacy preservation requirement, for example the work inthe proposed guarantee of the k-anonymity on the node degrees so that for the every node v there are at least k other nodes that have the same node degrees as v. The work in provided the heuristic solution against the 1-neighborhood attacked. Worked in quantified privacy risks associated with the different kinds of attacks on the social networks. Work in anonymized data graph by the adding edges and the nodes for a resulting graph is kautomorphic. Work in the identified two realistic targets of the attacks the Node Info and the Link Info and the proposed the solution to form k pairwise isomorphic sub graph.

VII EVALUATION

We evaluate our anonymization method on together synthetic and the real data sets. The experiments are conducted with the MATLAB R2010a running on the local machine with the Intel Core 2 Duo E8400 3.0 GHz CPU and 8 GB Linux RAM. The parameter Setting following parameters need to be the determined before the conducting experiments:

- The threshold value α that is determines if or not two 1-neighbor graphs approximately match.
- The probability p for the randomizing graph. To obtain the reasonable threshold value we conduct experiments with respect to the different sizes of the 1-neighbor graphs as follows: The Given N we first randomly generate the 1- neighborhood graph Gv with the N nodes and then generate the similar graph G v by the randomly modifying $p\ast$ percentage of the edges. Lastly with n different damping factors D = [d1,...,dn], we can calculate the cost for the optimal matching Gv and G vdenoted as cost(G v ,G v). Process above mention will be conducted for the multiple rounds and the average value c is used the threshold value α. In this experiment we set $p\ast = [0.1, 0.05]$ and $\beta = 0.1/N$ for the matching nodes in Virtual node set and choose damping factors

D = [0.7, 0.8, 0.9]. Threshold value α with different $p\ast$ values, while N ranges from 4 to 100, α will decrease as N increases, e.g. when $N = 4$, $\alpha = 0.22$, and when $N = 100$, $\alpha = 0.0158$.

VIII CONCLUSION

In this scenario Novel 1*-neighborhood attack. To resist this attack we can define the key property probabilistic in distinguishability for the outsourced social networks and we propose the heuristic anonymization scheme to the anonymized social networks with this property. Empirical study indicates that the anonymized social networks can quiet to be used to the answer aggregate queries with the high accuracy. In this we will conduct the thorough theoretical study of the risks onoutsourcing social networks to the cloud and try to introduce another privacy mechanism to our scheme for example by the combining with the l-diversity we enable the nodes in the group to be associated with at least l different attributes. For that the average node degree is 22 in evaluation. Hence in many social networksand average node degree is higher which may take the proposed anonymization scheme incompetent. So that we will conduct much experiments on the larger social graphs with sophisticated node density.

REFERENCES

[1] L. Getoor and C. Diehl, “Link mining: A survey,” ACM SIGKDD Explorations Newsletter, 2005.

[2] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica et al., “A view of cloud computing,” Communications of the ACM, 2010.

[3] G. Wang, Q. Liu, and J. Wu, “Hierarchical attribute-based encryption for fine-grained access control in cloud storage services,” in Proceedings of ACM CCS, 2010.

[4] J. Gao, J. Yu, R. Jin, J. Zhou, T. Wang, and D. Yang, “Neighborhoodprivacy protected shortest distance computing in cloud,” in Proc. of ACM COMAD, 2011.

[5] B. Zhou, J. Pei, and W. Luk, “A brief survey on anonymization techniques for privacy preserving publishing of social network data,” ACM SIGKDD Explorations Newsletter, 2008.

[6] J. Potterat, L. Phillips-Plummer, S. Muth, R. Rothenberg, D. Woodhouse, T. Maldonado-Long, H. Zimmerman, and J. Muth, “Risk network structure in the early epidemic phase of HIV transmission in Colorado springs,” Sexually Transmitted Infections, 2002.

[7] B. Zhou and J. Pei, “Preserving privacy in social networks against neighborhood attacks,” in Proc. of IEEE ICDE, 2008.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank citation ranking: Bringing order to the web,” Stanford InfoLab, Tech. Rep., 1999.

[9] M. Diligenti, M. Gori, and M. Maggini, “A unified probabilistic framework for web page scoring systems,” IEEE Transactions on Knowledge and Data Engineering, 2004.

[10] E. Zheleva and L. Getoor, “Preserving the privacy of sensitive relationships in graph data,” in Proc. of ACM PinKDD, 2007

AUTHOR PROFILE

Sure Suresh is currently pursuing M.Tech in the Department of Computer Science, from Nalanda Institute of Engineering & Technology(NIET), siddharth Nagar, Kantepudi(V),Sattenapalli(M),Guntur (D), Andhra Pradesh , Affiliated to JNTU-KAKINADA.

K Devika Rani working as Professor at Nalanda Institute of Engineering & Technology (NIET), siddharth Nagar, Kantepudi(V), Sattenapalli (M), Guntur (D), Andhra Pradesh , Affiliated to JNTU-KAKINADA.