IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN : 2278-0661, p-ISSN : 2278-8727
PP 101-106
www.iosrjournals.org

Outsourcing Privacy-Protecting Social Networks to a Cloud

Harshit Varshney [1], Nital Adikane [2]
[1] *(Information Technology,MITCOE ,Pune,India MITCOE,Pune,India)*
[2] *(Information Technology,MITCOE ,Pune,India MITCOE,Pune,India)*

Abstract: *In the planet, corporations would publish social networks to a 3rd party, e.g., a cloud service supplier, for promoting reasons. Protective privacy once business enterprise social network information becomes a vital issue. During this paper, we tend to establish a unique sort of privacy attack, termed 1*-neighborhood attack. We tend to assume that associate aggressor has data regarding the degrees of a target's one-hop neighbors, additionally to the target's 1-neighborhood graph, that consists of the one-hop neighbors of the target and also the relationships among these neighbors With this data, Associate in Nursing wrong doer could re-identify the target from a k-anonymity social network with a likelihood beyond 1/k, wherever any node's one-neighborhood graph is isomorphism with k - 1 alternative nodes' graphs. To resist the 1*-neighborhood attack, we have a tendency to outline a key privacy property, likelihood identity, for Associate in Nursing outsourced social network, Associate in Nursing propose a heuristic indistinguishable cluster anonymization (HIGA) theme to come up with an anonymzed social network with this privacy property. The empirical study indicates that the anonymzed social networks will still be accustomed answer combination queries with high accuracy.*

Keywords*: Cloud computing, social networks, privacy, probability, indistinguishability*

I. INTRODUCTION

As social networks have developed speedily, recent analysis has begun to explore social networks to know their structure, advertising and selling, and data processing [1]. Cloud computing [2], [3], as associate degree rising computing paradigm, is expected to reshape the knowledge technology processes in the close to future. Cloud services, that are offered in an exceedingly pay as-you-go manner, promise omnipresent 24/7 access at a coffee cost. attributable to the overwhelming deserves of cloud computing, e.g., flexibility and measurability, additional and additional organizations that host social network information opt to source some of their information to a cloud surroundings [4].conserving privacy once publishing social network information becomes a very important issue. Social networks model social relationships with a graph structure exploitation nodes and edges, wherever nodes model individual social actors in a very network, and edges model relationships between social actors [5]. The relationships between social actor’s square measure usually non-public, and directly outsourcing the social networks to a cloud could lead to unacceptable disclosures. For example, commercial enterprise social network knowledge [6] that describes a set of social actors connected by sexual contacts or shared drug injections could compromise the privacy of the social actors concerned. Therefore, existing analysis has projected to anonymize social networks. Before out sourcing. A native approach is to easily anonymize the identity of the social actors before outsourcing. However, Associate in Nursing aggressor that has some information a few target’s neighborhood, especially a one-hop neighborhood, will still re-identify the target with high confidence. This attack, termed 1-neighborhood attack, is projected by Chow et al. [7].

In this paper, we have a tendency to establish a unique variety of privacy attack, termed 1*-neighborhood attack, wherever associate degree wrongdoer is assumed to know the degrees of the target’s one-hop neighbors, in addition to the structure of the 1-neighborhood graph. We call this sort of background the 1*-neighborhood IEEE TRANSACTIONS ON SOCIAL NETWORKING YEAR 2013 graph. This assumption is affordable, since once the wrongdoer knows the identities of the target’s one-hop neighbors; he will be terribly probably to gather additional data regarding the one-hop neighbors, instead of solely aggregation the association information between them. With this assumption, the wrongdoer may re-identify the target from a k-anonymity social network with a chance over 1/k. To illustrate, allow us to assume that the wrongdoer is aware of the degrees of Bob’s one-hop neighbors, Alice, Clark, Donland, and Harry, say 4, 2, 3, 3, severally. In Fig. 1-(d), the degrees of Alice’s one-hop neighbors, Bob, Clark, Eda, and Fred, are 4, 2, 2, 2, severally. Since Ref. [7] solely adds edges to form 1-neighborhood graphs similarity, Alice is often excluded from the target candidate set, and also the chance to re-identify Bob is 1. To handle the 1* neighborhood attack, Ref. [7] requires the addition of additional edges, so the degrees of the k similarity graphs square measure a similar. For instance, by adding edges between Grace and Fred, and between Grace and Eda, the degrees of Alice's one-hop neighbor's square measure a similar as that of Bob's. However, as additional edges square measure additional, the usage of the social networks is going to beanie compromised.

To permit helpful analysis on the social networks, while preserving the privacy of the social actors concerned, we define a key privacy property, probabilistic sameness, for an outsourced social network. To come up with Associate in nursing anonymized social network with such a property, we tend to propose a heuristic indistinguishable cluster anonymization (HIGA) theme. Our basic plan consists of 4 key steps: Grouping, we group nodes whose 1* neighborhood graphs satisfy sure metrics together, and supply a mix and cacophonous mechanism to make every cluster size a minimum of capable k; Testing, in a group, we tend to use stochastic process (RW) [8], [9] to check whether or not the 1-neighborhood graphs of any try of nodes close to match or not; Anonymization, we tend to propose a heuristic anonymization algorithmic program to form any node's 1-neighborhood graph close to match those of alternative nodes in a very cluster, by either adding or removing edges [10], [11]; organization, we randomly modify the graph structure with an explicit likelihood to make positive every 1*-neighborhood graph encompasses assure probability of being totally different from the first one.

II. LITERATURE SURVEY

1. L. Getoor and C. Diehl, "Link mining: A survey," ACM SIGKDD Explorations Newsletter, 2005.

Cloud computing brings significant benefits for service providers and users because of its characteristics: e.g., on demand, pay for use, scalable computing. Virtualization management is a critical task to accomplish effective haring of physical resources and scalability. Existing research focuses on live Virtual Machine (VM) migration as a workload consolidation strategy. However, the impact of other virtual network configuration strategies, such as optimizing total number of VMs for a given workload, the number of virtual CPUs (vCPUs) per VM, and the memory size of each VM has been less studied. This paper presents specific performance patterns on different workloads for various virtual network configuration strategies. For loosely coupled CPU-intensive workloads, on an 8-CPU machine, with memory size varying from 512MB to 4096MB and vCPUs ranging from 1 to 16 per VM; 1, 2, 4, 8 and 16VMs configurations have similar running time. The prerequisite of this conclusion is that all 8 physical processors are occupied by vCPUs. For tightly coupled CPU-intensive workloads, the total number of VMs, vCPUs per VM, and memory allocated per VM, become critical for performance. We obtained the best performance when the ratio of the total number of vCPUs to processors is 2. Doubling the memory size on each VM, for example from 1024MB to 2048MB, gave us at most 15% improvement of performance when the ratio of total vCPUs to physical processors is 2. This research will help private cloud administrators decide how to configure virtual resources for given workloads to optimize performance. It will also help public cloud providers know where to place VMs and when to consolidate workloads to be able to turn on/off Physical Machines (PMs), thereby saving energy and associated cost. Finally it helps cloud service users decide what kind of and how many VM instances to allocate for a given workload and a given budget.

2.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.

Cloud computing, as an emerging computing paradigm, enables users to remotely store their data into a cloud so as to enjoy scalable services on-demand. Especially for small and medium-sized enterprises with limited budgets, they can achieve cost savings and productivity enhancements by using cloud-based services to manage projects, to make collaborations, and the like. However, allowing cloud service providers (CSPs), which are not in the same trusted domains as enterprise users, to take care of confidential data, may raise potential security and privacy issues. To keep the sensitive user data confidential against untrusted CSPs, a natural way is to apply cryptographic approaches, by disclosing decryption keys only to authorized users. However, when enterprise users outsource confidential data for sharing on cloud servers, the adopted encryption system should not only support fine-grained access control, but also provide high performance, full delegation, and scalability, so as to best serve the needs of accessing data anytime and anywhere, delegating within enterprises, and achieving a dynamic set of users. In this paper, we propose a scheme to help enterprises to efficiently share confidential data on cloud servers. We achieve this goal by first combining the hierarchical identity-based encryption (HIBE) system and the cipher text-policy attribute-based encryption (CP-ABE) system, and then making a performance-expressivity tradeoff, finally applying proxy re-encryption and lazy re-encryption to our scheme

3. 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.

Development of online social networks and publication of social network data has led to the risk of leakage of confidential information of individuals. This requires the preservation of privacy before such network data is published by service providers. Privacy in online social networks data has been of utmost concern in recent years. Hence, the research in this field is still in its early years. Several published academic studies have proposed solutions for providing privacy of tabular micro-data. But those techniques cannot be straight forwardly applied to social network data as social network is a complex graphical structure of vertices and edges. Techniques like k-anonymity, its variants, L-diversity have been applied to social network data. Integrated technique of K-anonymity & L-diversity has also been developed to secure privacy of social network data in a better way.

4**.** M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava, “Anonymizing social networks,” Tech. Rep., 2007.

Advances in technology have made it possible to collect data about individuals and the connections between them, such as email correspondence and friendships. Agencies and researchers who have collected such social network data often have a compelling interest in allowing others to analyze the data. However, in many cases the data describes relationships that are private (e.g., email correspondence) and sharing the data in full can result in unacceptable disclosures. In this paper, we present a framework for assessing the privacy risk of sharing anonymized network data. This includes a model of adversary knowledge, for which we consider several variants and make connections to known graph theoretical results. On several real-world social networks, we show that simple anonymization techniques are inadequate, resulting in substantial breaches of privacy for even modestly informed adversaries. We propose a novel anonymization technique based on perturbing the network and demonstrate empirically that it leads to substantial reduction of the privacy threat. We also analyze the effect that anonymizing the network has on the utility of the data for social network analysis

III. PRELIMINARIES

3.1 System Model

We think about a system that consists of a publisher, a cloud service supplier, Associate in Nursing assailant, and plenty of users. The publisher, such as Facebook or Twitter, outsources a social network to a cloud. In our system, a social network is sculptured as Associate in Nursing gun directed and unlabelled graph G = (V (G),E(G)), where V (G) may be a set of nodes, and E(G) ⊆ V (G)×V (G) may be a set of edges. The node identities square measure assumed to be removed. The assailant has bound information regarding the target and he tries to re-identify the target by analyzing the outsourced social network. To safeguard the privacy of the social actors within the network from the assailant, the publisher anonymizes G to G = (V (GE (G)) before outsourcing. As in [7], we have a tendency to assume that every node in G exists in G and no pretend nodes square measure superimposed in G to preserve the worldwide structure of the social network. As previous work [10], [11], we have a tendency to permit edges ∈ E (G) to be off from E (G).

3.2Attacker Model:

1-Neighborhood Graph Gu = (Vu, Eu), where Vu denotes a set of nodes {v| (u, v) ∈ E (G) ∨ (v = u)}, and Eu denotes

a set of edges {(w, v)| (w, v) ∈ E (G) ∧ {w, v} ∈ Vu}. For each node u ∈ V (G), the related 1*-neighborhood graph, denoted as Gu* is defined as “1*-Neighborhood Graph. Gu*= (Gu, Du)”, where Gu is the1-neighborhood graph of node u, and Du is a sequence of degrees of u’s one-hop neighbors.

3.2 Random Walk (RW) based Approximate Matching

We use random-walk-based approximate matching because the building block of our HIGA scheme. The stochastic process (RW) [8] is understood as a great tool to obtain the steady state distribution for a graph, referred to because the topological signatures, which give the inspiration for the approximate matching

3.3 Design Goals:

The main style goal of our work is to scale back the chance of a social actor being re-identified where as business enterprise social networks to a cloud. Specifically, given a social graph G, we wish to come up with AN anonymized graph G, in order that the subsequent needs square measure satisfied:

- Privacy. Given any target's 1-*neighborhood graph, the aggressor cannot re-identify the target from AN anonymized social network confidently on top of a threshold.
- Usability. The anonymized social networks may be wont to answer combination queries with high accuracy.

3.4 Algorithm: Heuristic Anonymization Algorithm

{Given m groups g1, . . . , gm as CGS}

```
Sort CGS in descending order of the number of neighbors
While CGS is not empty
Do
    Choose the first group in CGS as the processing group
    g*and remove g*from CGS
    for each node u in g*
    Do
        Construct 1-neighborhood graph Gu
        Use Eq. 3 to calculate Gu's topological signatures
    for each pair of nodes (u, v) in g*
    Do
        Use Eq. 5 to calculate cost of matching Gu and Gv
    While exists a cost larger than $\alpha$
    Do
        Randomly choose a node $u \in g^*$ as the group seed
        For each node $v \in g^*$
        Do
            if cost(Gu,Gv) > $\alpha$ then
                Approach Gu to Gv with probability q
                Approach Gv to Gu with probability 1 – q
```

Suppose that there square measure m teams, g1, . . . , gm, where each group size is assumed to be a minimum of up to k. for every cluster, if any try of nodes don't seem to be close to matching, we use a heuristic anonymization algorithmic program (Alg. 1) to create the 1-neighborhood graphs close to match as follows:

Initially, the candidate cluster set (CGS) consists of m groups. we tend to kind teams in falling order of the amount of neighbors, decide the primary one because the process cluster, and remove it from cgs system. for every node within the process group, we tend to construct its 1-neighborhood graph, and use RW to calculate connected topological signatures. Then, for any try of nodes u and v, we use Eq. five to calculate the value of matching their 1-neighborhood graphs. For any try of nodes, if this cost is smaller than a threshold price α, we elect successive grouping in cgs system because the process cluster and fuck once more.

Otherwise, we tend to modify 1-neighborhood graphs of the nodes in the process cluster as follows: we tend to initial opt for a random node u within the cluster because the cluster seed. For the other node v during this cluster, if the connected price cost(Gu,Gv) is larger than α, we tend to approach the structure of Gu thereto of Gv with probability Q, and approach the structure of Gv thereto Gu with likelihood 1−q. This method can continue till, for any pair of nodes within the process cluster, the price for matching their 1-neighborhood graphs is up to or smaller than α. The anonymization method is also algorithmic, since some changes may impact the teams that are processed antecedently. However, owing to the power-law node distribution, and the small world development [14], this method can quickly stop. To approach the structure of Gv = (Vv,Ev) thereto of Gu = (Vu,Eu), we tend to initial acquire the best matching of nodes in 2 graphs with the strategy mentioned in Section IV-(B). In the optimal matching, for any try of nodes $x \in Vv$ and $w \in Vu$, if cost(x,w) > α, we tend to create u's connections an equivalent as those of v. throughout the approaching method, we tend to ensure that the structure of Gu won't be changed. For example, in Fig. 3-(B), the best matching for graphs G1 and G2 is A1, B2, C3, D4, E5. Suppose α = zero, we tend to approachG1 to G2. For all pairs of nodes, solely the price of matching D ∈ G1 and four∈ G2 is larger than zero. Therefore, we change the association of node four to an equivalent as that of node D. In the first spherical, node D solely connects to node C in G1, but node four connects to each nodes three and five. Therefore, we remove the edge between nodes four and five, and check the price once more. In this round, cost(E, 5) > α, and node E connects to C in G1, but node five has no connections. Therefore, we tend to add a position between nodes three and five, and check the price once more. during this spherical, the connected price of matching *G*1 and *G*2 is equal to *α*, and the anonymization completes.

IV. EVALUATION

4.1 Anonymization Strength

Here, we assume an intelligent attacker who knows the uniform random noise probability *p*. We also assume that the intelligent attacker will not give up even if the exact match cannot be found.

4.2 Anonymization Cost

Our solution anonymizes a graph by adding and removing edges, which will lead to some information loss.

4.3 Data Sets

Synthetic Dataset

We use the Barab´*a*i-Albert algorithm [16] (B-A algorithm for short) to generate synthetic data sets. The basic idea of the B-A algorithm is to first generate a network of a small size, and then use that network as a *seed* to build a larger sized network, continuing this process until the actual desired network size is reached. The node degree follows the power law distribution. In our experiments, the initial seed size contains 5 interconnected nodes, and the generated networks contain 1,000, 2,000, 3,000, 4,000, and 5,000 nodes.

Real Dataset

To validate the effectiveness of our anonymization method, we conduct experiments on a real social network, Arxiv ASTRO-PH (Astro Physics) collaboration network1, which is from the e-print arXiv and covers scientific collaborations Between authors, submitted papers to Astro Physics category, in the period from January 1993 to April 2003, and contains 18,772 nodes and 396,160 edges.

V. CONCLUSION

In this paper, we have a tendency to determine a completely unique 1*-neighborhood attack. To resist this attack, we have a tendency to outline a key property, probabilistic indistinguishability for outsourced social networks, and we propose a heuristic anonymization theme to anonymize social networks with this property. The empirical study indicates that the anonymized social networks will still be wont to answer mixture queries with high accuracy. For our future work, we'll conduct an intensive theoretical study of risks on the outsourcing social networks to a cloud, and try to introduce alternative privacy mechanisms to our theme, e.g., by combining with l-diversity, we have a tendency to change the nodes in a very cluster to be related to a minimum of l totally different attributes. what is more, the average node degree is twenty two within the analysis. However, in several social networks, the common node degree is far higher which can build the projected anonymization theme inefficient. Therefore, we'll conduct additional experiments on larger social graphs with higher 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, “Neighbor hood privacy 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.

[11] M. Hay, C. Li, G. Miklau, and D. Jensen, “Accurate estimation of the degree distribution of private networks,” in Proc. of IEEE ICDM, 2009.

[12] M. Gori, M. Maggini, and L. Sarti, “Exact and approximate graph matching using random walks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005.

[13] K. Liu and E. Terzi, “Towards identity anonymization on graphs,” in Proc. of ACM SIGMOD, 2008.

[14] J. Scott, “Social network analysis,” Sociology, 1988.

[15] M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava, “Anonymizing social networks,” Tech. Rep., 2007.

[16] A. Barab´asi and R. Albert, “Emergence of scaling in random networks,” Science, 1999.

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

[18] A. Campan and T. Truta, “A clustering approach for data and structural anonymity in social networks,” in Proc. of PinKDD, 2008.

[19] L. Zou, L. Chen, and M. O¨ zsu, “K-automorphism: A general framework for privacy preserving network publication,” in Proc. of the VLDB, 2009.

[20] J. Cheng, A. Fu, and J. Liu, “K-isomorphism: privacy preserving network publication against structural attacks,” in Proc. of ACM COMAD, 2010.

[21] X. Ying and X. Wu, “On link privacy in randomizing social networks,” Knowledge and Information Systems, 2011.

[22] C. Tai, P. Yu, D. Yang, and M. Chen, “Privacy-preserving social network publication against friendship attacks,” in Proc. of ACM KDD, 2011.