International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

Outsourcing Cloud Data Privacy-Preserving Based On Over Encryption

Student: Thati Praveen (13H61D5829)
Guide: Jayendar Kumar
(Assistant Professor) CVSR ENGINEERING COLLEGE

Abstract—
In the real world, companies would publish social networks to a third party, e.g., a cloud service provider, for marketing reasons. Preserving privacy when publishing social network data becomes an important issue. In this paper, we identify a novel type of privacy attack, termed 1-neighborhood attack. We assume that an attacker has knowledge about the degrees of a target's one-hop neighbors, in addition to the target's 1-neighborhood graph, which consists of the one-hop neighbors of the target and the relationships among these neighbors. With this information, an attacker may re-identify the target from a k-anonymity social network with a probability higher than 1/k, where any node's 1-neighborhood graph is isomorphic with k−1 other nodes' graphs. To resist the 1*-neighborhood attack, we define a key privacy property, probability in distinguish ability, for an outsourced social network, and propose a heuristic indistinguishable group anonymization (HIGA) scheme to generate an anonymized social network with this privacy property.*
Index Terms—Cloud computing; social networks; privacy; probability in distinguish ability

INTRODUCTION

Social networks have developed rapidly, recent research has begun to explore social networks to understand their structure, advertising and marketing, and data mining. Cloud computing as an emerging computing paradigm, is expected to reshape the information technology processes in the near future. Cloud services, which are available in a pay as-you-go manner, promise ubiquitous 24/7 access at a low cost. Due to the overwhelming merits of cloud computing, e.g., flexibility and scalability, more and more organizations that host social network data choose to outsource a portion of their data to a cloud environment. Preserving privacy when publishing social network data becomes an important issue.

Social networks model social relationships with a graph structure using nodes and edges, where nodes model individual social actors in a network, and edges model relationships between social actors. The relationships between social actors are often private, and directly outsourcing the social networks to a cloud may result in unacceptable disclosures. For example, publishing social network data that describes a set of social actors related by sexual contacts or shared drug injections may compromise the privacy of the social actors involved. Therefore, existing research has proposed to anonymize social networks before outsourcing.

A naıve approach is to simply anonymize the identity of the social actors before outsourcing. However, an attacker that has some knowledge about a target's neighborhood, especially a one-hop neighborhood, can still re-identify the target with high confidence. This attack, termed *1*-neighborhood attack, is proposed by Zhou etc.

Consider a synthetic social network of "co-authors", as shown in Fig. (a), where a node denotes an author and an edge that links two authors denotes that they previously cooperated on a paper. In the neighborhood attack, an attacker, who knows Mani's one-hop neighbors and the connections between them, i.e., Mani's *1-neighborhood graph*, as shown in Fig. (b), can still re-identify Mani from an anonymized graph, Fig.(c), where all user identities are removed.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

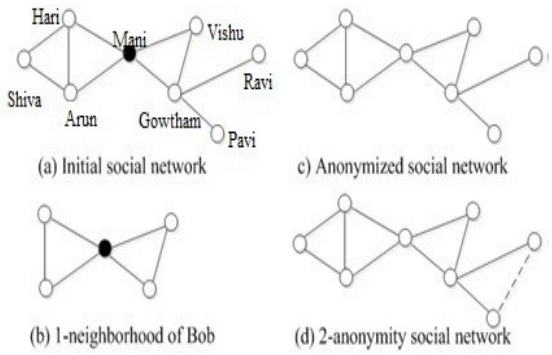


Fig. 1-neighborhood attacks in a social network.

Assume that the attacker knows the degrees of Mani's one-hop neighbors, Gowtham, Vishu, Hari, and Arun, say 4, 2, 3, 3, respectively. In Fig. (d), the degrees of Gowtham's one-hop neighbors, Mani, Vishu, Ravi, and Pavi, are 4, 2, 2, 2, respectively. Since Ref only adds edges to make 1-neighborhood graphs isomorphic, Gowtham can be excluded from the target candidate set, and the probability to re-identify Mani is 1. To deal with the 1*-neighborhood attack, Ref. requires the addition of more edges, so that the degrees of the *k* isomorphic graphs are the same. For example, by adding edges between Shiva and Pavi, and between Shiva and Ravi, the degrees of Gowtham's one-hop neighbors are the same as that of Mani's. However, as more edges are added, the usage of the social networks will be further compromised.

ENCRYPTION TO CLOUD STORAGE

The emergence of sharing confidential corporate data on cloud servers, it is imperative to adopt an efficient encryption system with a fine-grained access control to encrypt outsources data. Cipher text-policy attribute-based encryption (CP-ABE), as one of the most promising encryption systems in this field, allows the encryption of data by specifying an access control policy over attributes, so that only users with a set of attributes satisfying this policy can decrypt the corresponding data. However, a CP-ABE system may not work well when enterprise users outsource their data for sharing on cloud servers, due to the following reasons:

First, one of the biggest merits of cloud computing is that users can access data stored in the cloud anytime and anywhere using any device, such as thin clients with limited bandwidth, CPU, and memory capabilities. Therefore, the encryption system should provide high performance.

Second, in the case of a large-scale industry, a delegation mechanism in the generation of keys inside enterprises needed. Although some CP-ABE schemes support delegation between users, which enables a user to generate attribute secret keys containing a subset of his own attribute secret keys for other users, hope to achieve a full delegation, that is, a delegation mechanism between attribute authorities (AAs), which independently make decisions on the structure and semantics of their attributes.

Third, in case of a large-scale industry with a high turnover rate, a scalable revocation mechanism is a must. The existing CP-ABE schemes usually demand users to heavily depend on AAs and maintain a large amount of secret keys storage, which lacks flexibility and scalability.

Motivation

A hierarchical attribute-based encryption (HABE) model by combining a HIBE system and a CP-ABE system, to provide fine-grained access control and full delegation. Based on he HABE model, construct a HABE scheme by making a performance-expressivity tradeoff, to achieve high performance. Finally, a scalable revocation scheme by delegating to the CSP most of the computing tasks in revocation, to achieve a dynamic set of users efficiently

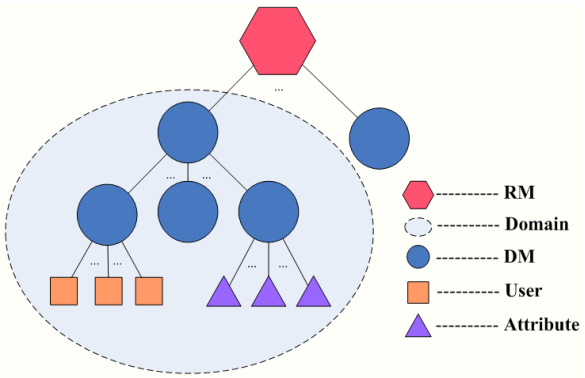
The habe model

The HABE model (see Figure) consists of a root master (RM) that corresponds to the third trusted party (TTP), multiple domain masters (DMs) in which the top-level DMs correspond to multiple enterprise users, and numerous users that correspond to all personal in an enterprise.

The RM, whose role closely follows the root private key generator (PKG) in a HIBE system, is responsible for the generation and distribution of system parameters and domain keys. The DM, whose role integrates both the properties of the domain PKG in a HIBE system and AA in a CP-ABE system, is responsible for delegating keys to DMs at the next level and

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

distributing keys to users. Specifically, enable the leftmost DM at the second level to administer



all the users in a domain, just as the personnel office administers all personnel in an enterprise, and not to administer any
attribute. Notice that other DMs administer an arbitrary number of disjoint attributes, and have full control over the structure and semantics of their attributes.

In the HABE model, first mark each DM and attribute with a unique identifier (ID), but mark each user with bothan ID and a set of descriptive attributes. Then, as Gentry et al [1],enable an entity's secret key to be extracted from the DM administering itself, and an entity's public key, which denotes its position in the HABE model, to be an IDtuple consisting of the public key of the DM administering itself and its ID, e.g., the public key of DM*i* with ID*i* is in the form of (PK*i*−1*;* ID*i*), the public key of user *U* with ID*u* is in the form of (PK◇*;* ID*u*), and the public key of attribute *a* with ID*a* is in the form of (PK*i;* ID*a*), where PK*i*−1, PK◇, and PK*i* are assumed to be the public keys of the DMs that administer DM*i*, *U*, and *a*, respectively.

NEIGHBORHOOD-PRIVACY PROTECTED

Graph structured data are used in numerous applications, *e.g.,* web graphs, social networks, ontology graphs, biological and chemical pathways, transportation networks. High efficiency is essential for frequent and basic graph operations. However, even basic operations on a graph can be very time-consuming due to the complexity of structural connectivity's and graph size. Moreover, real graph datasets are growing rapidly in size, making the attainment of high efficiency even harder.

The paradigm shift of cloud computing offers a new approach for storage- and compute-intensive tasks allowing users to migrate their burden (*e.g.,* data maintenance and computing utilities) to an outsourced server (or cloud server). The outsourced server typically has sufficient resources to maintain very large datasets and provides quick response to users' requests with its powerful distributed and parallel architecture.

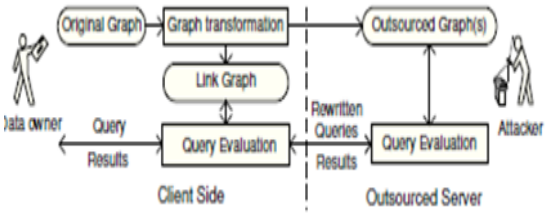
Privacy-Preserving Graph Publishing

Privacy protection for graph publishing has been studied recently. Most of the existing work on graph publishing focuses on certain structural anonymizations, such as 1-neighborhood, k-degrees, k-automorphism, k-isomorphism, cluster based vertex anonymity, as well as many others. These techniques typically focus on using the least amount of modifications of the original graph (minimal information loss) to make it satisfy the targeted security requirement. Unfortunately, the anonymized graphs produced from these privacy protection techniques generally do not necessarily maintain the statistical and graph theoretical characteristics of the original. In particular, for any pair of vertices, there is no guarantee of the degree of similarity or preservation of shortest distances between the anonymized graph and the original graph. For example, k-isomorphism is proposed recently to partition a graph into k disjoint, isomorphic sub-graphs, which unfortunately cannot be used to compute shortest distances in the original graph. In addition, most of the exiting works deal with privacy on un weighted graphs, and do not consider the impact of edge weights.

Security Issues in Outsourced Server

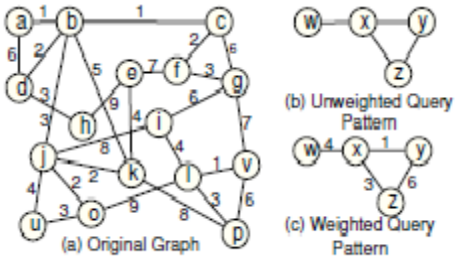
Sensitive data protection and verification of query results in the outsourced server have attracted much attention recently. A work closely related to this studies the verification issue in outsourcing graphs for shortest path discovery. In their solution, the original graph data are outsourced along with verification objects, and the client side will validate the correctness of the results with the verification objects. However, they do not consider how to protect the sensitive information of the original graph.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org



The main contributions are summarized below.

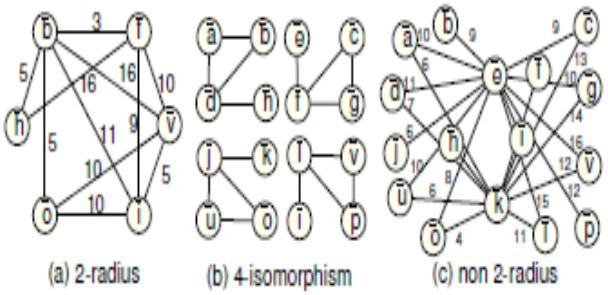
- Formulate a new graph transformation problem as minimizing the size of the link graph Gl on the client side on the condition that the privacy of outsourced graphs Go is protected and shortest distances can be answered using Go and Gl. A new security model, named 1-neighborhoodd- radius, which hides local details in direct edges or a d radius for each vertex to counter neighborhood attacks.
- propose a greedy approach to generate outsourced graphs and a link graph for exact shortest distance answering with protected neighborhood privacy.
- study how to answer approximate shortest distances in the same context with an average additive distance error bound



Protecting Neighborhood Privacy

Protecting the sensitive local neighborhood information. Basically, the information of how an individual vertex links to its neighbors and what are the edge weights for these links are deemed sensitive and need to be protected. In particular, focus on 1-neighborhood attacks since it tends to be more difficult to collect information beyond that . In addition, when two vertices are very close to each other (within a threshold d), even without a direct link, their relationship can also be considered important. Formally, the outsourced graph should meet the requirement of

An interesting and important question is whether enforcing the d-radius property on each outsourced graph is too strict. For instance, can simply remove all edges in the original graph and then only connect those pairs of vertices whose shortest distance is no smaller than d? The answer is negative.



It demonstrate a successful attack on such as graph. Suppose an outsourced graph Go is constructed by the removal of all direct edges and addition of edges from vertex u to v if _G(u, v) _ 2. it the edges related with vertices e and k in Go in Fig.3(c). This graph is a non- 2-radius graph since two adjacent vertices e and k in the original graph co-exist in the same d-radius outsourced graph, which violates the first condition of a d-radius graph. Attackers can observe that the graph is strongly connected, thus, they can know that there is a direct edge (with any weight) or a path with cost no larger than 2 between e and k in the original graph. Based on the triangle inequality over Go, attackers even infer that the edge weight is no smaller than 4!

A Naïve Approach

The two optimization targets (in Objective 3) are along the same line with one another and not in conflict. For instance, if minimize the space cost Gl, the computational cost over Gl tends to be minimized. In the following, will focus on minimizing the space cost of Gl. In addition, as discussed above, for any pair of vertices (u, v) with distance less than d, its distance can be discovered on the original graph. Thus, only those vertex pairs whose distances are no less than d need to be considered in the transformation. Formally, our graph

transformation can be reduced to a problem on minimizing Gl as follows:

CONCLUSION

A novel 1*-neighborhood attack. To resist this attack, a key property, probabilistic in distinguish ability for outsourced social networks, and by using heuristic anonymization scheme to anonymize social networks with this property. It states that the anonymized social networks can still be used to answer aggregate queries with high accuracy.

REFFERENCES

[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.