Project Proposal - Detecting Privacy Leakage

YounSun Cho

CS590BB, Fall 2010

1 Statement of Problem

The recent emergence of cloud computing has showed an enormous potential for a considerable impact on every aspect of our daily lives. This fad of technological advance, however, raises new privacy concerns which make people hesitate to adopt cloud computing. On the other hand, some new technologies give us a way to preserve privacy. Motivated by this dual role of privacy and the necessity of privacy preserving mechanism, I will investigate the problem of detecting privacy leakage in a cloud computing environment.

2 What needs to be done to solve this problem?

First, I need to define the scope of problem to investigate by surveying related work and categorizing the types of privacy violation/leakage detections. Once I decide the scope of problem, I need to define a model and make assumptions for the model and adversaries. This might require for me to study background knowledge of statistical inference in data minding or intrusion detection systems.

3 What has been done?

Krekke [5] identified two types of privacy violation detections and discussed a strawman architecture by adopting the intrusion detection system without implementing detail. Bhattacharya *et al.* [2] proposed a privacy violation detection and monitoring system by using intelligent data mining techniques in conjunction with network features. Bottcher and Steinmetz [3] proposed an algorithm for identifying a set of suspicious queries for an XML database by comparing submitted queries with the exposed data. This algorithm, however, limited on a single suspicious query to determine the leakage of privacy rather than the combination of related suspicious queries. Kruger *et al.* [6] proposed a method for detecting privacy violation malicious code (e.g., spyware) by using dynamic slicing to discover dependencies between events in an execution trace in the view of compiler/programming language.

Interesting research which are more relevant to my project have been done by Ahmed *et al.* [1] and Chow *et al.* [4]. In [1], they proposed an audit mechanism for detecting the leakages of attributes and identifying adversaries by counting

the frequency of an attribute with a disclosure threshold and computing the probability of exposure of a single attribute by a host. The recent study [4] described a theoretical framework for inference detection using a reference corpus and measured its strength through web-mining algorithms. They demonstrated the performance of their schemes by identifying all the keywords which allow for inference of HIV with confidence above a certain threshold.

4 What can I do?

Although I may or may not change the direction of this project by slightly changing models or assumptions as I proceed the project, a current tentative model which I have briefly in mind is reactive privacy leakage detection after the fact based on a transaction analysis by adopting the statistical inference. In a cloud computing environment, a group of people could work on the same set of data as a collaborative teamwork, and this group of people can be viewed as a clique. Furthermore, additional useful information can be obtained by combining and analyzing a series of related transactions by using conjunctive or disjunctive operations of transactions. Thus I will investigate an algorithm (or method) for the detection of violating k-clique privacy as well as individual privacy with an analysis of a single and/or the combination of suspicious multiple transactions. Note that I will not address the problem of the access control or authorization in a multiuser environment since it is orthogonal problem and out of scope of this project. If time permits, I will perform a theoretical analysis to demonstrate the feasibility of my algorithm.

5 What I have done?

I have read related papers and tried to come up with defining the scope of problem and a model.

References

1. M. Ahmed, D. Quercia, and S. Hailes. A statistical matching approach to detect privacy violation for trust-based collaborations. In *Proceedings of the First International IEEE WoWMoM Workshop on Trust, Security and Privacy for Ubiquitous Computing*, 2005.
2. J. Bhattacharya, R. Dass, V. Kapoor, and S. k. Gupta. Utilizing network features for privacy violation detection. In *First International Conference on Communication System Software and Middleware (Comsware'06)*, 2006.
3. S. Bottcher and R. Steinmetz. Detecting privacy violations in sensitive xml databases. In *Secure Data Management (SDM'05)*, pages 143–154, 2005.
4. R. Chow, P. Golle, and J. Staddon. Detecting privacy leaks using corpus-based association rules. In *Proceedings of KDD*, 2008.
5. T. H. Krekke. Privacy violation detection. Master's thesis, Norwegian University of Science and Technology (NTNU), 2004.
6. L. Kruger, H. Wang, S. Jha, P. D. McDaniel, and W. Lee. Towards discoverying and containing privacy violations in software. Technical report, Sept. 2004.