From Itemsets Through Trajectories to Location Based Services: A Knowledge Hiding Privacy Approach

Aris Gkoulalas–Divanis
Computer & Communication Engineering Department
University of Thessaly, Volos 38222, Greece.
aris.gkoulalas-divanis@vanderbilt.edu

Overview

Since its inception in 2000, privacy preserving data mining has gained increasing popularity in the data mining community. This line of research can be attributed to the growing concern of individuals and organizations for the violation of privacy in the mining of their data by the existing data mining technology. Consequently, a new body of research emerged, providing novel approaches for the mining of data, while prohibiting the leakage of sensitive information.

My dissertation studies methodologies for the preservation of privacy in different contexts, data domains, and application scenarios. It consists of three parts. The first part investigates methodologies for the hiding of sensitive knowledge in the form of association rules, extracted from large transactional databases. Our research led to the proposal of a new direction of approaches that guarantee optimal hiding by introducing the least amount of side effects, while causing minimal distortion to the original data.

The second part of the dissertation extends the applicability area of association rule hiding by applying similar techniques for the hiding of temporally and spatially annotated data. Our contribution is a privacy aware trajectory query engine, which enables untrusted users to query trajectory data that reside in a database. The engine guarantees that the answers that are returned to the end users do not violate the privacy of the users, whose movement is recorded in the database.

The last part of the dissertation studies trajectory hiding in a real–time environment where users, equipped with mobile devices, request services that depend on their location. The goal is to deliver methodologies that offer such services in a way that protects the identity of the requester. We considered services that require multiple location transmissions to be offered and contributed methodologies that protect the identity of the requester from the time of request, until the service provision. Moreover, we contributed PLOT; the first toolbox for the offering of privacy in location based services.

Part I: Association Rule Hiding

Association rule hiding aims at sanitizing a database in a way that (i) no sensitive association rule can be revealed when the sanitized database is mined at certain thresholds of confidence and support (or higher), (ii) all the nonsensitive rules can be successfully mined from the sanitized database, and (iii) no rule that was non-existent in the original database can be generated in its sanitized counterpart. Ideally, the sanitization process has to be accomplished in a way that minimally affects the original database, preserves the general patterns and trends, and conceals all the sensitive knowledge. A solution that addresses the three aforementioned goals is called *exact*. An exact solution that minimally distorts the original database is called *optimal*.

Prior to our work in [2], the approaches that had been proposed for association rule hiding were based on heuristics to sanitize the original database [16]. The benefit of using heuristics to guide the knowledge hiding process has to do with the efficiency and the scalability of these methodologies to very large problem sizes, rather than with the actual quality of the produced solutions. Indeed, after extensive experimentation with state–of–the–art heuristic methodologies for association rule hiding, we found that in many cases these methodologies were unable to identify optimal hiding solutions that could be derived from the original database.

In order to identify optimal solutions to association rule hiding problems, we formulate the hiding process as a Constraints Satisfaction Problem (CSP), where the holding of the constraints (effectively controlling the status – frequent vs infrequent – of a small portion of itemsets in the sanitized database) guarantees the lack of side effects in the hiding solution. Among the potentially many solutions of the CSP, we select the one that minimizes the distortion of the original database, thus guarantees optimal hiding. My dissertation proposed three algorithms that facilitate optimal association rule hiding [2, 3, 10, 11] as well as a parallelization framework that improves their scalability without compromising the quality of the hiding solution [8]. Moreover, we proposed a methodology for the quantification of the privacy that is offered by the exact hiding algorithms [6].

Part II: Hiding Trajectory Patterns

In the second part of the dissertation, we extended the application domain of privacy aware methodologies for knowledge hiding from traditional transactional data to historical data of user mobility. The trajectories of the users, generated in the course of their movement and recorded in a dataset, lead to a far more powerful type of knowledge than the one that is offered by frequent itemsets. First of all, in mobility datasets we acknowledge an ordering of the elements per user trajectory, defining the sequence of traces that represents the trajectory that was followed by the user. Furthermore, unlike traditional data, mobility data is referenced both in space and time, which significantly adds to the complexity of the privacy aware mining of this data [1, 15].

The contribution in this part of my dissertation is a privacy

aware trajectory tracking query engine [9, 17] that provides on–site, restricted access to in–house data to facilitate privacy aware data publishing. The methodologies that had been proposed in the past were based solely on count and/or sum queries in statistical databases, since no other information was made available to the inquirer. On the other hand, the proposed query engine supports a large variety of queries, involving both trajectory and non–trajectory data. Furthermore, it provides the necessary mechanisms to secure the database against disclosure of confidential information, offering strict guarantees about what can be observed by untrusted third parties. Thus, it effectively blocks certain types of attacks that aim at utilizing the trajectories of the users in the database to reveal user identity.

Part III: Privacy in Location Based Services

The last part of my dissertation proposes methodologies for the offering of privacy in services that depend on users' location (known as Location-Based Services – LBSs) to be successfully provided. We assume users, each equipped with a mobile device, who request LBSs when on the move. A service request includes, among others, the location coordinates of the requester. Thus, with untrustworthy service providers, the identity of the requester can be easily found if the request reaches the service provider in its original form. To prohibit identification of the requester when he or she requests an LBS, the request has to be transformed to a privacy aware counterpart. Among the different research directions that have been investigated to tackle this problem, $\mathcal{K}$–anonymity has received most of the attention, requiring that each service request that reaches the service provider points to at least K potential issuers (including the requester).

My dissertation contributed along two principal lines of research within the $\mathcal{K}$–anonymity direction of privacy in LBSs: *historical $\mathcal{K}$–anonymity*, and *trajectory $\mathcal{K}$–anonymity* approaches. To support historical $\mathcal{K}$–anonymity, we proposed two methodologies, which use the historical movement of the users in the system to automatically derive movement patterns that signify locations and times where the privacy of each user is under threat and use this information to adequately cover up the requesters of LBSs. The first approach [18] assumes unconstrained user movement, while the second approach [4] assumes that the movement of the users is constrained to a series of allowable routes.

Trajectory $\mathcal{K}$–anonymity approaches protect the privacy of the requesters of LBSs by using the current location of all users in the system as well as their future locations to cover up the requester until the service completion. To support trajectory $\mathcal{K}$–anonymity, we proposed a new line of research: *personalized approaches to trajectory $\mathcal{K}$–anonymity*, along with two novel approaches: a free terrain approach [5] that operates on the privacy model of [7], and a network aware approach [14] that operates on the privacy model of [12]. Both approaches consider an attacker who has knowledge of the user movement statistics for each user in the system and can use his/her knowledge to breach user privacy.

Last, we proposed PLOT [13], the first open–ended toolbox that integrates centralized approaches for location privacy in location based services, in a common framework. PLOT offers the essential functionality for the implementation of novel location and trajectory privacy approaches, while serving as a testbed for the comparison of different location privacy algorithms. It supports both networkless and network aware solutions to location privacy, it implements strategies for the identification of the frequent movement patterns of the users in the system, and it offers tools for the qualitative and quantitative evaluation of the implemented solutions.

Ph.D. Dissertation Committee

Prof. V. S. Verykios, Prof. E. N. Houstis, Prof. P. Bozanis, Prof. A. Daskalopulu, Prof. C. E. Houstis, Prof. Y. Manolopoulos, and Prof. Y. Theodoridis.

1. REFERENCES

[1] F. Bonchi, Y. Saygin, V. S. Verykios, M. Atzori, A. Gkoulalas-Divanis, S. V. Kaya, and E. Savas. *Privacy in Spatiotemporal Data Mining*, chapter 11, pages 297–333. Mobility, Data Mining and Privacy: Geographic Knowledge Discovery. Springer, 2008.

[2] A. Gkoulalas-Divanis and V. S. Verykios. An integer programming approach for frequent itemset hiding. In *CIKM*, pages 748–757, 2006.

[3] A. Gkoulalas-Divanis and V. S. Verykios. A hybrid approach to frequent itemset hiding. In *ICTAI*, pages 297–304, 2007.

[4] A. Gkoulalas-Divanis and V. S. Verykios. Concealing the position of individuals in location based services. In *EEEE*, pages 973–984, 2008.

[5] A. Gkoulalas-Divanis and V. S. Verykios. A free terrain model for trajectory $\mathcal{K}$–anonymity. In *DEXA*, pages 49–56, 2008.

[6] A. Gkoulalas-Divanis and V. S. Verykios. Hiding sensitive knowledge without side effects. *KAIS*, 2008. Accepted.

[7] A. Gkoulalas-Divanis and V. S. Verykios. *HESTIA: Historically–Enabled Spatio–Temporal Information Anonymity*, chapter 11, pages 1–28. Data Mining and Management. Nova Science Publishers, Inc., 2008.

[8] A. Gkoulalas-Divanis and V. S. Verykios. A parallelization framework for exact knowledge hiding in transactional databases. In *SEC*, pages 349–363, 2008.

[9] A. Gkoulalas-Divanis and V. S. Verykios. A privacy–aware trajectory tracking query engine. *SIGKDD Expl.*, 10(1):40–49, 2008.

[10] A. Gkoulalas-Divanis and V. S. Verykios. Exact knowledge hiding in transactional databases. *IJAIT*, 18(1):17–37, 2009.

[11] A. Gkoulalas-Divanis and V. S. Verykios. Exact knowledge hiding through database extension. *TKDE*, 21(5):699–713, 2009.

[12] A. Gkoulalas-Divanis, V. S. Verykios, and P. Bozanis. A network aware privacy model for online requests in trajectory data. *DKE*, 68(4):431–452, 2009.

[13] A. Gkoulalas-Divanis, V. S. Verykios, and D. Eleftheriou. PLOT: Privacy in location–based services: an open–ended toolbox. In *MDM*, 2009.

[14] A. Gkoulalas-Divanis, V. S. Verykios, and M. F. Mokbel. Identifying unsafe routes for network–based trajectory privacy. In *SDM*, 2009.

[15] V. S. Verykios, M. L. Damiani, and A. Gkoulalas-Divanis. *Privacy and Security in Spatiotemporal Data and Trajectories*, chapter 8, pages 213–240. Mobility, Data Mining and Privacy: Geographic Knowledge Discovery. Springer, 2008.

[16] V. S. Verykios and A. Gkoulalas-Divanis. *A Survey of Association Rule Hiding Methods for Privacy*, chapter 11, pages 267–289. Privacy Preserving Data Mining: Models and Algorithms. Springer, 2008.

[17] V. S. Verykios, A. Gkoulalas-Divanis, Y. Theodoridis, and N. Pelekis. *Privacy Preservation in Trajectories of Moving Objects*, chapter 15, pages 1–21. Privacy Preservation on Computer and Communication Technologies: Technical and Legal Issues. Papasotiriou, 2009.

[18] P. Zacharouli, A. Gkoulalas-Divanis, and V. S. Verykios. A $\mathcal{K}$–anonymity model for spatiotemporal data. In *IEEE STDM*, pages 555–564, 2007.