



## PPKM: Preserving Privacy in Knowledge Management

N. Maheswari (Corresponding Author)

P.G. Department of Computer Science

Kongu Arts and Science College, Erode-638-107, Tamil Nadu, India

E-mail: mahii\_14@yahoo.com

Dr. K. Duraiswamy

K.S. R. College of Technology

Tiruchengode- 637-209, Tamil Nadu, India

E-mail: kduraiswamy@yahoo.co.in

### Abstract

This paper discusses the techniques that support the extraction, sharing, and utilization of knowledge for collaborative problem solving applications. A system framework is proposed for secure knowledge management, called PPKM, which in addition to provide standard security mechanisms such as access control, will possess crucial feature, namely privacy-preservation, where privacy-preservation means that the knowledge extraction process should not compromise the privacy of the source data. This framework is explained by elaborating on its components and their relationship to existing techniques such as database, data perturbation, rule hiding, data mining, and machine learning.

**Keywords:** Privacy, Knowledge management system, Data mining, Data perturbation

### 1. Introduction

The advancement in networking, storage, and processor technologies has brought in an unprecedented amount of digitalized information. In order to effectively utilize collected data in applications, organizations routinely use Database Management Systems (DBMSs) to store, manage, and use the collected data. While it is well-accepted that data has become vital assets of organizations, what many decision-making applications really need is the knowledge hidden in the raw data. For this reason, knowledge-extraction technologies such as data mining and machine learning have been developed in recent years to make it feasible to "refine" large volumes of raw data into succinct knowledge that can be directly utilized in decision-making applications. However, most current data mining based applications are designed to solve problems for the owners of the data, that is, the data mining is performed on the data of an organization to solve business problems of the same organization. Although still very popular, such use of data mining is limited and needs to be extended.

As the Internet quickly evolves into a global computational infrastructure, it provides a platform for new applications that allow autonomous organizations to collaboratively solve problems using data mining. Knowledge extracted from the data is often more abstract and less bulky than the raw data, the sharing of extracted knowledge may be much easier and more beneficial than sharing of data in many problem solving scenarios. One desirable feature of this "knowledge-sharing" paradigm is the distinction between the knowledge extraction process in which data mining algorithms are applied to discover the hidden knowledge in data, and the knowledge-dissemination process in which the discovered knowledge is utilized in applications to solve problems.

There is a need for a flexible *framework* for Knowledge Management Systems (KMSs) that provide the basic and common functionalities required to effectively coordinate the knowledge extraction with the knowledge dissemination. In this paper, a framework is proposed with an emphasis on security and privacy protection.

#### 1.1 Contributions

The need of a systematic investigation on Knowledge Management Systems is discussed. KMS has the functionalities in the flavor of the traditional Database Management Systems (DBMSs): (1) It facilitates the extraction of knowledge from existing traditional database and/or knowledge-base systems. The extraction of knowledge may be based on some data mining algorithms. (2) It facilitates storage, retrieval, integration, transformation, visualization, and analysis of extracted knowledge structures (e.g., decision trees, association rules, neural networks). Besides, it also supports construction of new knowledge structures from those existing ones to form knowledge that is deeper than the knowledge directly extracted from the raw data.

Besides traditional security goals a new security goal of secure knowledge management system is specified, namely *privacy-preservation*. A major feature of PPDM techniques is that they usually entail modifications to the data in order to sanitize them from sensitive information (both private data items and complex data correlations) or anonymize them with some uncertainty level. Thereby, a framework for secure knowledge management called Privacy-Preserving Knowledge Management (PPKM) is presented.

## 2. Related Work

On the evolution of service based computing paradigms (Sarawagi, Nagaralu, (2000) the relationship to privacy protection of data (Chaum, 1981), (Clifton, Marks,1996). On the relationship to data mining and machine learning (Agrawal, Srikant, 2000), (Evmievski, Gehrke, Srikant, 2003), (Lindell, Pinkas,2000). On the relationship to knowledge sharing (Data Mining Group, 2003). These efforts are focused either on the mechanisms that enable data mining systems to transfer discovered models to application programs, or on the types of services in which the discovered models can be useful. The PPKM framework described in this paper provides a general system framework that serves as a platform to integrate techniques of knowledge extraction, knowledge sharing, and knowledge utilization in a secure environment. Obviously, the data mining model as a service is a special instance of the framework, and the security requirements of PPKM greatly enrich the functionality of knowledge sharing systems.

## 3. The PPKM Framework

In PPKM knowledge management mean the methodology for systematically extracting and utilizing the knowledge. A *Knowledge Management System* (KMS) is a collection of collaborative software systems that collectively provide the functionality needed to perform the tasks of knowledge management. The purpose of PPKM framework is to define various roles that are played by participating systems, the relationships among different roles, and how they are related to the two key functionalities of a KMS, namely knowledge-extraction and Knowledge-dissemination and the security goals.

### 3.1 Model

As shown in Figure 1, at the heart of the PPKM is a Knowledge Management System (KMS), which can be thought of abstractly as a system that takes data and rules as input, extracts knowledge from the data (possibly with the help of the input rules), manages the extracted knowledge, and provides knowledge based services to knowledge customers. In the following, the PPKM framework is explored by describing the roles and their relationships in more details.

Input to KMS - The input to a KMS is a set of *datasets* and optionally a set of data, rules from databases and rule bases. In Fig. 1, the input to KMS includes data and rules. We stress that the access to the datasets and rules are protected by their respective sources through appropriate security policies (e.g., Mandatory Access Control, Discretionary Access Control, Role-Based Access Control), and that the *controlled access* may be enforced, for instance, by a security mechanism implemented in a DBMS. Moreover, the data and rules may be owned by different parties that are presumably prohibited from sharing, or not willing to share, their data/rules, although they are allowed to take advantage of the data in their own decision-making applications.

KMS - From a functionality perspective, a KMS is analogous to a traditional Database Management System (DBMS). However, there are some fundamental differences: (1) The objects managed by a KMS are knowledge models such as decision trees or association rules. Whereas, the objects managed by a DBMS are raw data. (2) Parties are autonomous and a party may play one or more roles. (3) A KMS is strictly more powerful than a DBMS, because it must ensure the property, namely *privacy-preserving knowledge extraction*. Whereas, no such requirement is specified in a traditional DBMS. The KMS consists of components that play three types of roles: *knowledge miner*, *knowledge provider*, and *knowledge manager*. For example, in the specific instance of KMS in Fig. 1 there are one knowledge miner and two knowledge providers, and each one of them also is a knowledge manager. The functionality of these three roles is as follows.

Knowledge Miner - A knowledge miner provides supports for knowledge extraction tasks which for example may include the preparation of data, the specification of extraction tasks, and the execution of extraction algorithms. A knowledge miner may be fully automated or interactive. Knowledge can be extracted from database using an appropriate method such as data mining (example: association rule mining). A key feature of knowledge miners is that they must guarantee that the extraction of knowledge will not compromise individual privacy. This feature can be ensured by the so-called privacy preserving data mining techniques.

Knowledge Provider - A knowledge provider provides services to knowledge customers. The simplest form of the service is to deliver an extracted knowledge model to a knowledge customer. However, more sophisticated and value-added services may require a nontrivial utilization of extracted knowledge. For example, a knowledge provider may provide a service by using the extracted knowledge to answer queries posted by a decision-making application of a knowledge customer. Such services may be implemented through a variety of techniques, such as web services and software agents.

Knowledge Manager - A knowledge manager provides supports for storage, retrieval, analysis, integration, visualization, and transformation of extracted knowledge. In other words, a knowledge manager is to knowledge what a database management system is to data. In a KMS, knowledge managers are often not separable from other roles of the system since they provide a set of functionality that is fundamental to both knowledge miners and knowledge providers. Extracted knowledge may be expressed in various representation languages.

Output of KMS - The KMS disseminates knowledge to knowledge customers through an appropriate interface (e.g., web services). For example, a knowledge customer may ask one or more knowledge providers certain questions, so that the answer(s) will be utilized in the knowledge customer's decision making procedure. The access to the knowledge may be controlled via an appropriate policy, and enforced via an appropriate system.

### 3.2 Privacy Preserving

Besides traditional security requirements such as access control, authorization, and authentication, a KMS should satisfy the new security requirement such as privacy-preservation. By privacy-preservation we mean that the knowledge-extraction procedure must protect individual privacy in the input datasets. This may be crucial to certain knowledge management systems (e.g., the systems coordinating government agencies counter-terror activities). Privacy preserving algorithms (Elisa Bertino, IgornaiFovino, Loredana, 2005), are applied for the inputs (data, rules). The algorithms help the knowledge provider in transferring the information required by the knowledge customers and the information about the other customers stored in the database will be hidden.

## 4. Example

In the following, we demonstrate the generality of the PPKM framework by describing two specific instances of the framework.

### 4.1 General Category

In the general category, we only consider the setting where there is a single organization that owns the data, buys data mining software, and runs the software to extract knowledge that will be exclusively used by the organization itself. As a further step towards what we called KMS, the organization may not have to buy the data mining software. Instead, it can use that software through "application as a service". In this case, the issues of privacy-preserving emerge: the application server (i.e., data mining software owner) should not learn any information about the organization's datasets, while allowing the organization to obtain the extracted knowledge. In principle, privacy preserving technique can solve this problem.

### 4.2 Business Category

We now consider a scenario arising in an emerging computing paradigm called "knowledge-as-a-service" which is a natural extension of service oriented computing, such as "application as a service" and "database as a service" (Hacigümüs, Mehrotra, Iyer, 2002). These service oriented paradigms emerge as cost-effective business models in response to increasing business competition and to the cost of keeping the desired computational, data management, and knowledge discovery capabilities that has become too high to be justified for many organizations. By delegating computational, data management, and knowledge discovery tasks to appropriate service providers, organizations can better satisfy their information processing needs with much lower costs. The .knowledge-as-a-service. serves an example of the separation of knowledge extraction and knowledge utilization, and is justified for the following reasons.

1). The rising costs of knowledge extraction. Data mining is a specialized and complex task that involves many steps and requires well trained personnel. Despite the tremendous advance in hardware, software, and networking technologies, the costs associated with knowledge extraction is still on the rise. These costs are for the acquisition of software, hardware, datasets, and the maintenance and management of systems. The situation is further complicated if one needs accurate knowledge and strict privacy in the knowledge extraction procedure.

2). Restricted access to data. Although knowledge models are often extracted by an organization from its own datasets, much of useful knowledge may be in data owned by other organizations. Access to data of another organization may be prohibited by law or policies.

For example, the national criminal databases can only be accessed by law enforcement organizations. Likewise, hospital patient's data is only accessible to relevant health-care organizations. Yet another typical scenario is that competition rivals would never share their data, but would benefit from knowledge extracted from each other's datasets. As a consequence, just like that data are valuable assets of today's organizations; knowledge models will be valuable assets of tomorrow's organizations.

3). Limited choice of technology that addresses privacy concerns. The emerging of the data mining industry has inspired a lot of concerns on individual privacy (Clifton, Marks, 1996). To relieve these concerns, privacy-preserving data mining techniques have been proposed. (Agrawal, Srikant, 2000), (Evmievski, Gehrke, Srikant, 2003), (Goldreich,

Micali, Wigderson, 1987). As a consequence, one who is interested in accurate knowledge and strong privacy guarantee may be forced to conduct computation- and communication-extensive tasks, which may incur significant investment.

#### 4. Different needs of knowledge by different applications.

There are many ways that knowledge models can be utilized. Two extreme examples are: (1) An application needs to own an entire knowledge model. (2) An application only needs to apply (rather than to own) a knowledge model to certain instance data. The difference between these two types of utilization is comparable to the difference between buying a car and taking a taxi, or to the difference between purchasing an expensive full-fledged software system and paying only for some of the needed functionality.

Consider a Company's database consists of the suppliers and the items details. The database consists of three suppliers A, B, and C and their corresponding items. If the supplier A wishes to know the items stored in the database other than his supply. But the supplier A does not have the rights to access the database. Using the privacy preserving algorithm, Knowledge Provider provides the items details to the Knowledge Customer (Supplier A), only the item names can be provided, but the cost and supplier details of the corresponding items are hidden. Knowledge Miner extracts the information with privacy. Knowledge Provider helps the Knowledge Customer in providing essential information based on the queries of the Knowledge Customer.

#### 5. Conclusion

In this paper, we present a system framework for secure knowledge management, which provides privacy-preservation which is ensured in the knowledge extraction procedure. The framework is explored by describing the roles played by system components, the relationships among various roles, and how these roles are related to existing technology, such as databases, data mining, privacy preserving and data perturbation. The framework helps in securing the information in Knowledge Management System.

#### References

- Agrawal, R. & Srikant, R. (2000). Privacy-preserving data mining. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD 2000)*, pages 439-450.
- Chaum, D. (1981). Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM*, 24:84-88.
- Clifton, C., Marks, D. (1996). Security and privacy implications of data mining. In *Workshop on Research Issues in Data Mining and Knowledge Discovery*.
- Data Mining Group. (2003). PMML version 2.1. <http://www.dmg.org>.
- Elisa Bertino, IgornaiFovino, Loredana. (2005). A Frame work for Evaluating Privacy Preserving Data Mining Algorithms, *Springer Science*, 121-154.
- Evmievski, A., Gehrke,J., Srikant, R. (2003). Limiting privacy breaching in privacy preserving data mining. In *Proceedings of the 2000 Symposium on Principles of Database Systems (PODS 2003)*, pages 211-222.
- Farkas.C, Jajodia. (2003). The inference problem: A survey. *SIGKDD Explorations*, 4(2):6.11.
- Goldreich .O, Micali .S, Wigderson .A, (1987). How to playany mental game or a completeness theorem for protocols with honest majority. In *Proc. 19th ACM Symp. on Theory of Computing*, pages 218.229.
- Hacigümüs .H, Mehrotra. S, Iyer.B. (2002). Providing database as a service. In *Proceedings of the 8th International Conference on Data Engineering (ICDE 2002)*, pages 29.38. IEEE Computer Society.
- Lindell, Y., Pinkas, B. (2000). Privacy preserving data mining. In M. Bellare, editor, *Advances in Cryptology . Crypto 2000*, pages 36.54. *Springer*, 2000. Lecture Notes in Computer Science No. 1880.
- Sarawagi,S., Nagaralu,S.H. (2000). Data mining models as services on the Internet. *ACM SIGKDD Explorations*, 2(1):24. 28.

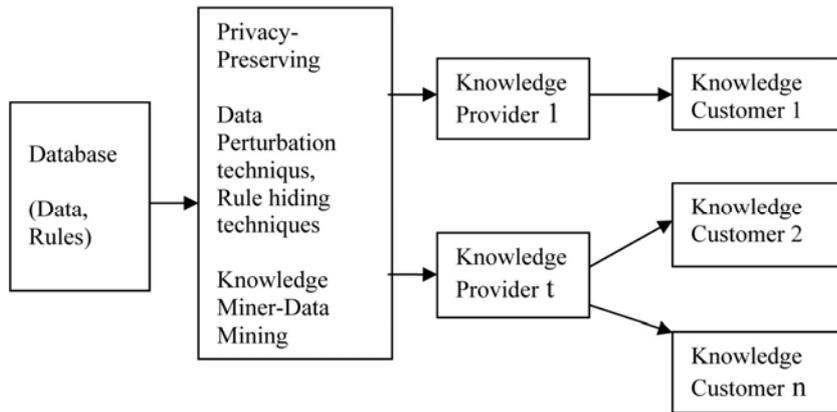


Figure 1. PPKM Framework