Secure Knowledge Discovery in Databases

Rick L. Wilson
Oklahoma State University, USA

Peter A. Rosen
University of Evansville, USA

Mohammad Saad Al-Ahmadi
Oklahoma State University, USA

INTRODUCTION AND BACKGROUND

Knowledge management (KM) systems are quite diverse, but all provide increased access to organizational knowledge, which helps the enterprise to be more connected, agile, and effective. The dilemma faced when using a KM system is to balance the goal of being knowledge-enabled while being knowledge-secure (Cohen, 2003; Lee & Rosenbaum, 2003).

A recent survey of IT security professions found that over 50% of respondents indicated an increase in the security budgets of their organizations since September 11, 2001, and projected that 2004 IT security budgets would be larger than ever (Briney & Prince, 2003).

The need for increased security is driven by both monetary concerns and legal/regulatory requirements. The goal of any security architecture, and specifically for KM systems, is to reduce the potential loss caused by intrusion, system misuse, privilege abuse, tampering, and so forth. Protection must be provided against external threats and from internal abuse and must include components that address the requirements for preserving the confidentiality of data where appropriate.

A 2002 Jupiter Research Consumer Survey estimates that as much as $24.5 billion in online sales will be lost by 2006 due to consumers' lack of confidence in the privacy of online transactions (*E-Compliance Advisor,* 2002). While lack of trust is an opportunity cost, security breaches can causes real losses. One study found firms with publicly announced security breaches lose an average of 2% of market capitalization within two days of attack, for an average of $1.65 billion dollars per breach (Cavusoglu, Mishra, & Raghunathan, 2002). On the regulatory side, legislation like the Health Insurance Portability & Accountability Act (HIPAA) and the Gramm-Leach-Bliley Act (GLBA) have forced companies in health care and financial services fields to improve their security measures (Briney & Prince, 2003; Ingrian Networks, 2004). Table 1 summarizes some common security threats.

While most of the major news stories about security breaches involve hackers who steal or access confidential information, infect systems with viruses, and cause trouble with worms or spam, an equally important threat comes from inside organizations. A report from Ingrian Networks (2004) indicated that 50% of security breaches are perpetrated by internal staff (see Lee & Rosenbaum, 2004). Internal threats represent a bigger risk than those from outsiders due to the difficulty in quantifying and counteracting the attacks. But while the risk of insider intrusions looms large, many IT security professionals still seem to be externally focused (Briney & Prince, 2003).

With the increased focus on security, both internally and externally, a method that seems to be gaining popularity is a layered security approach (e.g., Kolluru & Meredith, 2001; Clark, Croson, & Schiano, 2001). The layered approach proposes using multiple, overlapping forms of security measures. A representative list of such security measures is summarized in Table 2. The layered security approach is a good way to prevent breaches, because if one measure fails, it is possible that other measures employed can stop the attack.

While many network security texts discuss network related hardware and software that are relevant for protecting the IT and KM system infrastructure (e.g., Panko, 2004) from external threats, this article illustrates a complimentary approach. *Data perturbation* focuses on protecting confidential data primarily from unauthorized internal data snoopers. This approach can be used alone or in conjunction with other methods.

Data perturbation involves modifying confidential attributes using random noise, with the objective being to prevent disclosure while maximizing access to accurate information (Muralidhar, Parsa, & Sarathy, 1999). Thus, a KM system can maintain and allow access to masked representative confidential data while preventing exact data disclosure.

To illustrate the different ways that perturbation techniques can be utilized, consider an example scenario

Table 1. Security threats

Information Source	Ingrian, 2004	Briney, 2000	Boren, 2003
General	Poor security policies, human error, dishonesty, abuse of privileges, introduction of unauthorized software	Viruses, malicious code, executables, electronic theft, disclosure of proprietary data, use of resources for illegal / illicit activities	Storage threats: theft of servers, desktops, hard drives, tape backups, information, malicious software installed on server
Identification / Authorization	Internal / external attackers posing as valid users / customers		
Reliability of Service	Natural disasters, equipment failures, denial of service	Denial of service, buffer overflows	
Privacy	Eavesdropping, unauthorized monitoring of sensitive data		
Integrity / Accuracy	Modification or damaging of information		
Access Control	Password cracking, backdoors, security holes	Protocol weakness, insecure passwords, attacks on bugs in servers	Authentication credentials stolen / not properly managed, users given access to unnecessary information

Table 2. Selected security measures

Information Source	Ingrian, 2004	Briney, 2000	Boren, 2003
Security Policies / Security Education Programs	X	X	X
Identification/ Classification of Sensitive Data	X		
Determination of acceptable threat level	X		X
Passwords	X	X	X
Firewalls	X	X	X
Encryption	X	X	X
Backup / Recovery	X		X

where two divisions of a company are sharing information, some of which is considered confidential, and the sharing of data is done electronically. A layered approach to security would be the use of both data perturbation and encryption to secure the data. Data perturbation can be done before the transfer to mask or hide confidential attributes. During the transfer of data, encryption can be used to prevent attackers from accessing the data. By using the layered approach, the sending division protects its confidential data, obtains security during the transfer, and gives the receiving division full access to the perturbed data on the back end.

Data perturbation also can be used as a stand-alone technique to prevent unauthorized access from snoopers and hackers. If the data that users have access to is masked, then the impact of either an internal or external security breach is minimized. All attributes that are confidential (and numerical) can be masked, so their true values are hidden. In this way, even if there is a breach of security, no confidential information will be exposed.

Of all the different hardware and software security measures that are well documented for use in all information systems, data perturbation techniques are uniquely equipped to be one of the most useful and specifically applicable security techniques for knowledge management systems due to their focus on the data (and, therefore, on knowledge contained in the data). Thus, understanding how such perturbation techniques work and the