

Knowledge Hiding in Databases and Its Relationship to KDD

Tom Johnsten
T-Johnsten@wiu.edu
Computer Science Department
Western Illinois University
Macomb, IL 61455, USA

Vijay V. Raghavan
raghavan@cacs.louisiana.edu
Center for Advanced Computer Studies
University of Louisiana Lafayette
Lafayette, LA 70504, USA

Abstract

In this paper, we present a methodology to analyze the impact of KDD technology on database security. Our methodology consists of several steps that make up a new data analysis process that we call *Knowledge Hiding in Databases* (KHD) that is analogous to KDD. The specific contributions of this paper include a description of the steps of the KHD process, and a comparison of KHD and KDD.

Keywords: Data Mining, Database Security, Classification, Association, Knowledge Hiding in Databases.

1 Introduction

Nowadays companies and organizations frequently use KDD technology to analyze their stored data to discover valuable patterns or rules that can help them maintain their competitive edge. Recently, researchers within the information security community have begun to examine the impact of this technology on database security [2,3,4,5,7,8].

Chris Clifton and Don Marks are two of a small number of researchers who have examined this issue [4]. In their paper, *Security and Privacy Implications of Data Mining*, Clifton and Marks outline several general strategies designed to eliminate or reduce the security threat presented by this new technology. Their pro-

posed strategies include allowing users access to only a subset of data, altering existing data or introducing additional (spurious) data. They contend that the application of such policies is most effective in the context of specific learning tasks. These tasks include classification, estimation, clustering, characterization, and association. Of special interest are the classification learning tasks, which have the potential to disclose sensitive information whenever a database contains both sensitive and non-sensitive data; and, association learning tasks, which have the potential to disclose sensitive associations among stored data items.

In this paper, we present a methodology to analyze the impact of KDD technology on database security. Our methodology consists of several steps that make up a new data analysis process that we call *Knowledge Hiding in Databases* (KHD) that is analogous to KDD. The goal of KHD, in contrast to KDD, is the non-trivial hiding of potentially sensitive knowledge in data. The rest of this paper is organized as follows. Section 2 presents two examples to illustrate the security threat presented by KDD technology. Section 3 presents a general overview of KHD, while Section 4 compares and contrasts KHD with KDD. Finally, Section 5 presents the conclusions and discusses some future research projects.

2 Motivating Examples

We present two small examples to illustrate the security threat posed by KDD technology. Specifically, the examples are designed to illustrate the security threat posed by classification and association mining algorithms.

2.1 Classification Mining

Suppose the car company that owns the data in Table-1 has implemented the following security policy: "junior engineers may not access the mileage class of pre-production cars". This policy might be the result of company officials attempting to reduce the chance that someone outside the company will learn the mileage class of a newly designed car. As a result, company officials have concealed from junior engineers the mileage value associated with tuples T15, T16 and T17. In this scenario, the *Mile* attribute is referred to as the *protected attribute* since it contains the protected data elements (or attribute values) in its domain; and, the attributes *Id*, *Fuel*, *Cyl*, *Pow*, *Tran*, and *Pro* are referred to as *non-protected attributes* since they contain no protected data elements in their domain. The tuples T15, T16 and T17 are referred to as *protected tuples* since they contain a protected data element. The security risk presented in this example is the extent to which the voluntarily released data facilitates the violation of a security policy relative to a protected mileage value. The disclosure of that information can be achieved through the process of solving a classification problem. In other words, a junior engineer may be able to correctly infer a protected mileage value through the application of a classification mining algorithm to the data shown in Table-2. Specifically, tuples T1 through T14 represent the training set available to a junior engineer to create a classification model that allows for the assignment of a *Mile* label to the protected tuples T15 through T17.

Id	Fuel	Cyl	Pow	Tran	Pro	Mile
T1	efi	4	high	manu	y	med
T2	efi	6	high	manu	y	med
T3	2-bbl	6	high	auto	y	low
T4	efi	6	med	manu	y	med
T5	efi	4	high	manu	y	high
T6	2-bbl	4	med	manu	y	high
T7	efi	6	high	auto	y	low
T8	efi	6	med	manu	y	low
T9	efi	4	med	auto	y	med
T10	2-bbl	4	high	manu	y	high
T11	efi	4	med	manu	y	med
T12	efi	4	high	auto	y	high
T13	2-bbl	4	low	manu	y	high
T14	efi	6	high	auto	y	med
T15	2-bbl	4	high	auto	n	high
T16	efi	6	med	auto	n	low
T17	2-bbl	4	low	auto	n	med

Table 1: Car Data.

2.2 Association Mining

Association mining, like classification mining, also poses a threat to database security. To illustrate, consider the data shown in Tables -3 and -4. Suppose that each transaction in Table-3 corresponds to items purchased at a local health store. The *Cust-Id* data is obtained from loyalty cards that the health store offers to its customers. To acquire a loyalty card a customer fills out a form that includes such personal information as name and home address. A loyalty card benefits customers by making them eligible for a discount on purchased items. The security threat presented in this example is the extent to which the collected data facilitates the identification of individuals with specific health conditions. For example, a store employee could construct a set of syntactic constraints to identify transactions from Table-3 associated with a particular health condition. After which, the employee could apply an association mining algorithm, along with the constraints, to identify individual customers from Table-4 associated with the targeted health condition.

Id	Fuel	Cyl	Pow	Tran	Pro	Mile
T1	efi	4	high	manu	yes	med
T2	efi	6	high	manu	yes	med
T3	2-bbl	6	high	auto	yes	low
T4	efi	6	med	manu	yes	med
T5	efi	4	high	manu	yes	high
T6	2-bbl	4	med	manu	yes	high
T7	efi	6	high	auto	yes	low
T8	efi	6	med	manu	yes	low
T9	efi	4	med	auto	yes	med
T10	2-bbl	4	high	manu	yes	high
T11	efi	4	med	manu	yes	med
T12	efi	4	high	auto	yes	high
T13	2-bbl	4	low	manu	yes	high
T14	efi	6	high	auto	yes	med
T15	2-bbl	4	high	auto	no	null
T16	efi	6	med	auto	no	null
T17	2-bbl	4	low	auto	no	null

Table 2: Data Available to a Junior Engineer.

CustId	StJohnsWort	GinkgoBiloba	...
001	1	1	...
011	0	1	...
023	0	1	...
004	1	1	...
001	1	0	...
006	1	0	...
:	:	:	:

Table 3: Health Store Transaction Data.

The two examples in this section motivate the need to develop a general methodology to analyze the security threats presented by KDD technology.

3 KHD Methodology

As mentioned in the introduction, the goal of Knowledge Hiding in Databases (KHD) is the non-trivial hiding of potentially sensitive knowledge in data. We define the term 'non-trivial' to imply that knowledge is concealed in a manner that maximizes the amount of released data and maintains, to the greatest extent possible, the

Cust-Id	Name	...
001	Chris Bennet	...
004	Linda Jones	...
006	Steve Anderson	...
011	Anne Richardson	...
023	Jack VanBrooker	...
:	:	:

Table 4: Health Store Customer Data.

integrity of the data. In this context, data integrity ensures the extraction of accurate knowledge from parts of the data that are legitimately available. The inputs into the KHD process are a set of security constraints that must be satisfied and a collection of data D ; and, the output is a collection of data D' , derived from D , that satisfies the given security constraints.

3.1 Risk Assessment of Unauthorized Inference

It is possible to view the KDD security threat in terms of the expected occurrence of an unauthorized inference. In general, an unauthorized inference occurs when a set of data items, X , can be used to obtain information, Y , through the application of a function f [1]. With respect to KDD, the function f represents the decision rule(s) derived through the application of a data mining algorithm to the items X . We can categorize such functions, or inferences, into two board groups. Those that produce directed inferences and those that produce undirected inferences. From an adversary's point of view, the former type of inference results in the discovery of a specifically targeted collection of sensitive knowledge; while, the latter type of inference results in the discovery of an arbitrary collection of sensitive knowledge. Thus, an adversary may use KDD technology to obtain either a specifically targeted or arbitrary collection of sensitive knowledge.

Clifton [5] has proposed an approach to prevent an accurate undirected inference. His approach is based on the relationship between the sample size of the released (non-protected) data and the likelihood that the discovered rules from the sample are correct. The size of the released

sample is dependent upon two parameters, ϵ and δ , as defined in the following security policy, *here is a sample you may mine, but you can expect any result you get will be wrong $\epsilon\%$ of the time with probability δ , no matter how good your data mining is.* The strength of this approach is that the security guarantee applies broadly to any type of knowledge that can be derived. However, this strength is also a disadvantage in that it prevents accurate mining of legitimately available data.

There are two general strategies for assessing the security threat of a directed inference. One strategy is to assess the threat based on individual data mining algorithms. Then, based on the results from several selected algorithms, a decision can be made with regards to the threat. An alternative strategy is to make a generic assessment that is independent of a specific mining algorithm. This particular strategy has a number of potential advantages over the former. These include:

- Producing security policies that are applicable to a general set of mining algorithms.
- Providing insight on how to close an unauthorized inference channel.
- Reducing the overall time complexity of the assessment process.

Unfortunately a completely generic assessment especially in the context of classification mining algorithms is, in all likelihood, an impossibility as a result of variations among algorithms. However, such an assessment becomes feasible when the scope of the evaluation is limited to a specific group of mining algorithms or certain restrictions are placed on the domain of the given attributes. The realization of the former condition requires partitioning algorithms into groups whose members possess a common set of properties related to their security assessment.

3.2 The KHD Process

In general, KHD is an iterative process consisting of the following five steps:

- Identify Sensitive Information
- Identify Data Mining Algorithms
- Formulate Security Policies
- Risk Assessment
- Sanitize Data

We define the KHD process in terms of both a directed and undirected inference assessment. That is, we mean an assessment of the security risk assuming a directed and an undirected inference is to be performed, respectively. The process of performing an undirected inference assessment consists of only the last two steps, risk assessment and sanitizing data. There is no need to identify sensitive knowledge nor to formulate security policies since the objective is to prevent all accurate knowledge discovery tasks. In contrast, a directed inference assessment requires successful completion of all five steps.

The first step is to identify the sensitive knowledge that needs to be concealed within the data. For example, the statements "junior engineers may not have knowledge of pre-production cars" and "store employees may not have knowledge of individual customers suffering from sensitive health conditions", represent knowledge that an organization may wish to conceal. The identification of sensitive knowledge is an important task since the resulting analysis will only be as complete as the identified knowledge. In general, such knowledge is derived from an organization's stated security policies and its collected data. The second step is to match the knowledge identified in the first step with data mining algorithms that are capable of discovering it. For example, the directed inferences related to "pre-production mileage" and "sensitive health conditions" could be discovered through the application of a classification and association mining algorithm, respectively. Of course, if there

does not exist a data mining algorithm that is capable of discovering some piece of knowledge K , then K is concealed within the data.

The third step is to translate the sensitive knowledge identified in the first step into security policies that can be evaluated against the collected data. The structure of the policies will depend, in large part, on the classes of data mining algorithms identified in the second step. For example, the security policy corresponding to "pre-production mileage" might be defined in terms of an upper bound on the predicted accuracy of assigning the correct class label to tuples T15, T16, and T17 in Table-2. Similarly, the security policy corresponding to "sensitive health conditions" might be defined in terms of an upper bound on the confidence level of association rules of the form "*Cust-Id* \Rightarrow *Sensitive Health Condition*" derived from Table-3.

The fourth step is to evaluate the formulated policies against the collected data to discover if there are security policy violations. If it is determined that there will be violations, then it is necessary to sanitize the data in order to conceal, or hide, the sensitive knowledge. In general, sanitizing the data involves altering existing data, concealing existing data, or introducing additional (spurious) data. The specific manner in which data are modified is dependent upon several factors including, the data mining algorithms identified in the second step, the need to maximize the amount of released data, and the need to maintain the integrity of the data. A subtle, but significant, issue is that in some instances the act of sanitizing the data may affect the fundamental behavior of the targeted data mining algorithms.

4 KHD: The Other Side of KDD

Despite having different goals, KHD and KDD share some basic properties. Both processes analyze a collection of data for its information content. KDD tries to discover knowledge of interest by incorporating various robust heuristics; while, KHD tries to discover knowledge of

interest and at the same time enforcing security policies for sensitive knowledge (in spite of the robustness of the data mining algorithm).

Both KDD and KHD represent iterative processes in which individual steps may be repeated multiple times throughout the process. In addition, there exist a correspondence between individual steps. For instance, both processes begin with an information requirement, include a discovery phase, and terminate with a reporting phase. The reporting phase, in the context of KHD, includes communicating to the security analyst the results from the risk assessment and sanitization phases. Finally, with respect to directed KDD, the user may provide "inclusive" templates to influence the knowledge discovery process [9]. The concept of templates also exists in the context of KHD. In the latter case, the templates represent knowledge for which the reported accuracy should be relatively low. We are currently developing a security assessment tool for use with association mining algorithms that utilizes inclusive templates.

5 Conclusion and Future Work

In this paper, we proposed a new process of data analysis called Knowledge Hiding in Databases (KHD) that is analogous to KDD. The goal of KHD is the non-trivial hiding of knowledge in data to prevent the occurrence of unauthorized directed or undirected inferences. From an adversary's point of view, the former type of inference results in the discovery of a specifically targeted collection of sensitive knowledge and the latter type of inference results in the discovery of an arbitrary collection of sensitive knowledge. Despite having different goals, KHD and KDD share some basic properties. Both processes analyze a collection of data for its information content. KDD tries to discover knowledge of interest by incorporating various robust heuristics; while, KHD tries to discover knowledge of interest and at the same time enforcing security policies for sensitive knowledge. KHD, like KDD, is an iterative process that consist of

a sequence of steps. These steps include, identification of sensitive knowledge and data mining algorithms, formulation of security policies, risk assessment, and sanitizing data.

There are many potential research issues with respect to this challenging research area. The issues include development of formal methods to identify sensitive knowledge, and the development of additional security policies and risk assessment procedures for both classification and association mining algorithms.

References

- [1] Castano, S., Fugini, M., Martella, G., and Samarati, P. (1994). *Database Security*, Addison-Wesley.
- [2] Chang, L. and Moskowitz, I. (1998). Parsimonious Downgrading and Decision Trees Applied to the Inference Problem. *Proceedings of New Security Paradigms*, pp. 82-89.
- [3] Chang, L. and Moskowitz, I. (2000). An Integrated Framework for Database Privacy Protection. *Proceedings of the fourteenth Annual IFIP WG 11.3 Working Conference on Database Security*.
- [4] Clifton, C. and Marks, D. (1996). Security and Privacy Implications of Data Mining. *1996 SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, pp. 15-19.
- [5] Clifton, C. (1999). Protecting Against Data Mining Through Samples. *Proceedings of the Thirteenth Annual IFIP WG 11.3 Working Conference on Database Security*, pp. 193-207.
- [6] Holsheimer, M. and Siebes, A. (1994). Data Mining: The Search for Knowledge in Databases. *Report CS-R9406*. CWI. Amsterdam, The Netherlands.
- [7] Johnsten, T. and Raghavan, V. (1999). Impact of Decision-Region Based Classification Mining Algorithms on Database Security. *Proceedings of the Thirteenth Annual IFIP WG 11.3 Working Conference on Database Security*, pp. 177-191.
- [8] Johnsten, T. and Raghavan, V. (2001). Security Procedures for Classification Mining Algorithms. *Proceedings of the Fifteenth Annual IFIP WG 11.3 Working Conference on Database Security*, pp. 293-309.
- [9] Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., and Verkamo, A. (1994). Finding Interesting Rules from Large Sets of Discovered Association Rules. *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94)*, pp. 401-407.