ON THE DAMAGE AND COMPENSATION OF PRIVACY LEAKAGE

Da-Wei Wang, Churn-Jung Liau, Tsan-sheng Hsu, and Jeremy K.-P. Chen

Abstract A query on the distribution of a sensitive field within a selected population in a database can be submitted to the data center, and the answer to this query can leak private information, even though no identification information is provided. Inspired by decision theory, we present a quantitative model of the privacy protection problem in such a database query environment. In our model, the user information states are defined as classes of probability distributions on the set of possible confidential values. These states can be modified and refined by knowledge acquisition actions. The data confidentiality is guaranteed by ensuring that misusing private information is more costly than any possible gain.

1. INTRODUCTION

Through computer and communication technology, it has become popular to store massive amounts of data in a central databank and distribute them to the end users via Internet. Appropriately used, a databank can be a valuable information source for scientists, analysts, and policy makers. However, a weighty breach of privacy occurs if it can be accessed without restriction. As noted in [1], "in the past, most individuals lacked the time and resources to conduct surveillance required to invade an individual's privacy, as well as the means to disseminate the information uncovered, so privacy violations were restricted to those who did, mainly the government and the press." Internet technology has radically changed the situation. Nowadays, any individual Internet user can easily spread information worldwide within seconds. As such, revealing private information to unauthorized users, even if unintentionally, may cause a serious invasion of privacy.

Preventing unauthorized access to confidential information is the most basic technical problem. The medical history of a potential customer would be valuable for an insurance company. However, disseminating an individual's health information without prior consent is definitely an invasion of privacy. Thus the value of confidential information affects the incentive of invading privacy. Information brokers may try to collect and sell personal information for profit, and it is usually difficult to estimate damage caused by privacy leakage. To

discourage privacy invasion, victim compensation must be awarded by the one who gave out the information. The evaluation of gain and loss in relation to privacy leakage is crucial in designing privacy protection laws and systems.

In this paper, we try to tackle the problem from this aspect of information value versus the damage caused by privacy leakage. We focus on the following database query environment. , In each record in our database, there are private and sensitive fields as well as identification fields. Answers to queries about the distribution of a sensitive field within a selected population in the database can leak private information of individuals though no identification information is given.

We study a quantitative model of the privacy protection problem by weighing damage and compensation of privacy leakage. Safety of data is guaranteed by enforcing anyone disseminating private information must pay more than his gain for doing so. In the model, we need to represent the knowledge states of users receiving information. These knowledge states are adaptable to represent newly received information. We also need a formalism to represent the data to be protected, and a language to describe allowable queries. We adopt the data table and decision logic proposed in [12] for this purpose.

In the rest of the paper, we review data table formalism and the decision logic in section 2. Basic components of our models, the information states and knowledge acquisition actions, are defined in section 3. In sections 4 and 5, the basic model and its extension are presented. Related works are surveyed and our results are summarized in section 6.

2. DATA REPRESENTATION AND QUERY LANGUAGE

The most popular data representation is data table([12]). The data in many application domains, such as medical records, financial transactions, employee data, etc., can be represented as tables. A data table is a simplification of a relational database, since the latter consists of a number of data tables. A formal definition of data table is given in [12].

DEFINITION 1 *A data table*[1] *is a pair* $T = (U, A)$ *such that*

- *U is a nonempty finite set of individuals, called the population or the universe,*
- *A is a nonempty finite set of primitive attributes, and*
- *every primitive attribute $a \in A$ is a total function $a : U \rightarrow V_a$, where V_a is the set of values of a, called the domain of a.*

[1]Also called knowledge representation system, information system, or attribute-value system

The attributes of a data table can be divided into three sets. The first contains the *key attributes* that can be used to identify to whom a data record belongs. These attributes are always masked in a query response. Since key attributes uniquely identify individuals, we can assume that they are associated with elements in the universe U and omit them. We also have a set of *easy-to-know attributes,* the values of which are easily discovered by the public. For example, [14] points that some attributes, like birth-date, gender, and ethnicity, are available in some public databases, such as census or voter registration lists. The last attribute set is *confidential* attributes, the values of which are mainly the target we have to protect. At times, there is an asymmetry in possible values of a confidential attribute. For example, if the attribute is a HIV test result, the revelation of a positive result may cause a serious privacy invasion, whereas the revelation of a negative result is benign. For simplicity, we assume there is exactly one confidential attribute in a data table. This assumption is not essential, since we can encode multiple attributes into a single attribute by their Cartesian product. Thus, a data table is usually written as $T = (U, A \cup \{c\})$ where A is the set of easy-to-know attributes and c is the confidential one.

Let $V_c = \{s_0, s_1, \ldots, s_{t-1}\}$ be the set of possible values for the confidential attribute c. It is assumed that the *a priori* information of the user is the probability distribution of the population on V_c. In other words, we assume that the user knows the value

$$\frac{|\{u \in U \mid c(u) = s_i\}|}{|U|}$$

for all $0 \leq i \leq t-1$. The user can improve his knowledge by investigating some sampled individuals of the population, or querying the data center that stores the data table. On one hand, the user can discover the exact value of the confidential attribute of the chosen individuals by using investigation, however, it is difficult to conduct this kind of investigation. On the other hand, a query may ask for the probability distribution of confidential values in a specific subset of the population. Once the query is correctly answered, the user not only knows the probability distribution of the specific sub-population, but also that of its complement on V_c. Thus we need a language to specify a subset of individuals. To achieve this purpose, we use the decision logic (DL) proposed in [12]. DL is originally designed for the representation of rules induced from a data table by data mining techniques. It is also perfectly suitable for the query of a data table since each formula of the logic is satisfied by some individuals in the data table.

The atomic formula of decision logic with respect to a data table $T = (U, A \cup \{c\})$ is of the form (a, v), where $a \in A$ is an easy-to-know attribute and $v \in V_a$ is a possible value of the attribute a. The well-formed formulas (wff) of the logic are then formed by the Boolean connectives negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), and implication ($\rightarrow$):

- Each atomic formula is a wff.
- If φ is a wff, so is $\neg\varphi$.
- If φ and ψ are wffs, so are $\varphi \wedge \psi$, $\varphi \vee \psi$, and $\varphi \rightarrow \psi$.

The satisfaction relation $\models_T$ between U and the wffs is defined recursively by the following clauses:

1 $u \models_T (a, v)$ iff $a(u) = v$

2 $u \models_T \neg\varphi$ iff $u \not\models_T \varphi$

3 $u \models_T \varphi \wedge \psi$ iff $u \models_T \varphi$ and $u \models_T \psi$

4 $u \models_T \varphi \vee \psi$ iff $u \models_T \varphi$ or $u \models_T \psi$

5 $u \models_T \varphi \rightarrow \psi$ iff $u \not\models_T \varphi$ or $u \models_T \psi$

Any individual satisfying (a, v) has v as the value of his attribute a.

Using semantics of decision logic, we define the truth set of a wff φ with respect to the data table T as $\{u \in U \mid u \models_T \varphi\}$. The truth set is denoted by $|\varphi|_T$. Each wff φ specifies a subset of individuals $|\varphi|_T$ in the data table. A query φ submitted to the data center means a user wants to know the distribution of the sub-population $|\varphi|_T$ on V_c. If the query is correctly answered, the user would also know the distribution of the sub-population $U - |\varphi|_T$ by the axioms of probability. In other words, a correctly answered query would partition the population into two sub-populations and the distributions of confidential attribute values in these two sub-populations are both known. In this way, the user can subsequently query the data center to refine his knowledge regarding the distributions of confidential attribute values within different sub-populations. To model the evolution of user information after different queries, we need a formal representation of user information states. The next section will be devoted to these definitions.

3. THE INFORMATION STATES

Let us set a data table $T = (U, A \cup \{c\})$. Let $V_c = \{s_0, s_1, \ldots, s_{t-1}\}$ be the set of possible values for the confidential attribute and let $U = \{u_1, u_2, \ldots, u_n\}$ be the set of individuals. A *logical partition* of U is a subset of DL wffs $\Pi = \{\varphi_1, \varphi_2, \ldots, \varphi_m\}$ such that $|\varphi_i|_T \neq \emptyset, \forall 1 \leq i \leq m$, $|\varphi_1|_T \cup \cdots \cup |\varphi_m|_T = U$, and $|\varphi_i|_T \cap |\varphi_j|_T = \emptyset$ if $i \neq j$. Each $|\varphi_i|_T$ is called an equivalence class of Π. A piece of information known to the user is represented by a logical partition of U, a set of probability distributions indexed by the wffs of the partition, and the number of investigated individuals. Hereafter, we use $|\varphi|$ to denote the cardinality of $|\varphi|_T$.

DEFINITION 2 *An information state (or a knowledge state)* $\mathcal{I}$ *for the set of possible confidential attribute values* V_c *and the set of individuals U is a triple* $(\Pi, (\mu_i)_{0\le i\le t-1}, (\kappa_i)_{0\le i\le t-1})$, *where* Π *is a logical partition on U, and for all* $0 \le i \le t-1$, $\mu_i : \Pi \to [0,1]$ *and* $\kappa_i : \Pi \to \mathcal{N}$ *(*$\mathcal{N}$ *being the set of natural numbers) are functions satisfying, for any* $\varphi \in \Pi$, *the constraints*

(i) $\sum_{i=0}^{t-1} \mu_i(\varphi) = 1$,

(ii) $|\varphi| \cdot \mu_i(\varphi)$ *is a natural number, and*

(iii) $\kappa_i(\varphi) \le |\varphi| \cdot \mu_i(\varphi)$.

For convenience, we use vector notations to denote μ_i's and κ_i's. Thus $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_{t-1})$ and $\boldsymbol{\kappa} = (\kappa_0, \ldots, \kappa_{t-1})$ denote vector mappings which can be applied to elements of Π. The result is a vector consisting of the results of applying its component functions on the element. The dimension of each vector is evident from the context and not explicitly specified. Using vector notation, the information state defined above is $(\Pi, \boldsymbol{\mu}, \boldsymbol{\kappa})$. Let $\mathcal{I} = (\Pi, \boldsymbol{\mu}, \boldsymbol{\kappa})$ be an information state, then $(\Pi, \boldsymbol{\mu})$ is called a *partial knowledge state* compatible with $\mathcal{I}$. Note that a partial knowledge state may be compatible with various information states.

Within an information state, the user partitions the population into a number of sub-populations. He knows the probability distribution of confidential attribute values in a subpopulation. $\mu_i(\varphi)$ is the proportion of the individuals in sub-population $|\varphi|_T$ which have confidential attribute value s_i, whereas $\kappa_i(\varphi)$ is the number of investigated individuals in sub-population $|\varphi|_T$ which have confidential attribute value s_i. Since each DL wff φ is composed from atomic formulas with easy-to-know attributes, it can be assumed that it is easy for a user to verify whether a given individual satisfies φ. It can also be assumed that the cardinality of the truth set of each φ is known to the public. Note that it may sometimes be very difficult for the user to locate an individual satisfying a specific φ from the entire population *U*.

Information states can change through investigation of individuals in a specific sup-population and through queries posed to and answered by the data center. This process of knowledge refinement can be modeled by knowledge acquisition actions. A logical partition Π_2 is a refinement of another logical partition Π_1, denoted by $\Pi_2 \sqsubseteq \Pi_1$, if for all $\varphi_2 \in \Pi_2$, there exists $\varphi_1 \in \Pi_1$ such that $|\varphi_2|_T \subseteq |\varphi_1|_T$. If $\Pi_2 \sqsubseteq \Pi_1$, then each $|\varphi_1|_T$ such that $\varphi_1 \in \Pi_1$ can be written as a union of the truth sets of some wffs in Π_2.

DEFINITION 3 *Let* $\mathcal{I}_1 = (\Pi_1, \boldsymbol{\mu}_1, \boldsymbol{\kappa}_1)$ *and* $\mathcal{I}_2 = (\Pi_2, \boldsymbol{\mu}_2, \boldsymbol{\kappa}_2)$ *be two information states.* $\mathcal{I}_2$ *is a refinement of* $\mathcal{I}_1$, *denoted as* $\mathcal{I}_2 \sqsubseteq \mathcal{I}_1$, *if both of the following conditions are satisfied.*

1 $\Pi_2 \sqsubseteq \Pi_1$.

2 *For each* $\varphi \in \Pi_1$, *if* $|\varphi|_T = \bigcup_{1 \le i \le l} |\varphi_i|_T$ *for some set* $\{\varphi_1, \ldots, \varphi_l\} \subseteq \Pi_2$, *then*

$$|\varphi| \cdot \boldsymbol{\mu_1}(\varphi) = \sum_{i=1}^{l} |\varphi_i| \cdot \boldsymbol{\mu_2}(\varphi_i),$$

and

$$\boldsymbol{\kappa_1}(\varphi) \le \sum_{i=1}^{l} \boldsymbol{\kappa_2}(\varphi_i).$$

Note that the arithmetics (addition and multiplication) and comparison of vectors are defined as usual. For example, the addition of two vectors is carried out point-wise and results in a vector of the same dimension.

In our framework, there are two kinds of knowledge acquisition actions which can refine the user information states. One is query, and each query is represented by a wff φ in DL. The answer of the query is the distribution of the confidential values within the selected population $|\varphi|_T$ in the database. The other is investigation, which is specified by a wff φ and a positive integer number k. This means that the user has investigated k individuals from the set $|\varphi|_T$. For uniformity, each knowledge acquisition action is written as $\alpha = (\varphi, k)$ for a DL wff φ and $k \ge 0$. $k > 0$ means an investigation, whereas $k = 0$ means a query.

DEFINITION 4

1 *A knowledge acquisition action* $(\varphi, 0)$ *is applicable under the information state* $\mathcal{I}_1 = (\Pi_1, \boldsymbol{\mu_1}, \boldsymbol{\kappa_1})$ *and results in a state* $\mathcal{I}_2 = (\Pi_2, \boldsymbol{\mu_2}, \boldsymbol{\kappa_2})$ *if*

(a) *there exists* $\varphi' \in \Pi_1$ *such that* $|\varphi|_T \subset |\varphi'|_T$,

(b) $\Pi_2 = \Pi_1 - \{\varphi'\} \cup \{\varphi, \varphi' \wedge \neg\varphi\}$,

(c) $\mathcal{I}_2$ *is a refinement of* $\mathcal{I}_1$,

(d) $\boldsymbol{\kappa_2}(\psi) = \boldsymbol{\kappa_1}(\psi)$ *for any* $\psi \in \Pi_1 - \{\varphi'\}$, *and*

(e) $\boldsymbol{\kappa_2}(\varphi) + \boldsymbol{\kappa_2}(\varphi' \wedge \neg\varphi) = \boldsymbol{\kappa_1}(\varphi')$.

2 *A knowledge acquisition action* (φ, k) *where* $k > 0$ *is applicable under the information state* $\mathcal{I}_1 = (\Pi_1, \boldsymbol{\mu_1}, \boldsymbol{\kappa_1})$, *and* $\mathcal{I}_2 = (\Pi_2, \boldsymbol{\mu_2}, \boldsymbol{\kappa_2})$ *is a resultant state of the application if*

(a) $\varphi \in \Pi_1$ *and* $k \le |\varphi| - \sum_{i=0}^{t-1} \kappa_{1i}(\varphi)$

(b) $\Pi_1 = \Pi_2$,

(c) $\boldsymbol{\mu_1} = \boldsymbol{\mu_2}$,

(d) $\boldsymbol{\kappa_2}(\psi) = \boldsymbol{\kappa_1}(\psi)$ *for any* $\psi \ne \varphi$, *and*

(e) $\sum_{i=0}^{t-1} \kappa_{2i}(\varphi) = \sum_{i=0}^{t-1} \kappa_{1i}(\varphi) + k$.

Since the user's goal is to refine his knowledge through queries, he would pose queries so that the answers increase his knowledge. If the user's information state is (Π_1, μ_1, κ_1), then he should poses a query about a subset of an equivalence class in Π_1. This is the requirement of Condition 1a in Definition 4. After the query is answered, the corresponding equivalence class is partitioned into two parts — one satisfying φ and the other not, so we have Condition 1b in Definition 4. Condition 1c in Definition 4 further requires that the answer is correct so that the resultant information state is a refinement of the original one. Since the query does not investigate any new individuals, κ_2 agrees with κ_1 in the population that is not split by the query. For the split population, the number of investigated individuals does not changed in total. This is reflected respectively in Conditions 1d and 1e of Definition 4.

For investigation, we assume the user will only investigate individuals in a sub-population represented by a wff in Π_1. The assumption is not essential, because, if the investigated individuals are from different sub-populations, the investigation can be decomposed into a sequence of actions satisfying the applicability condition. Since it is assumed that the user knows the total number of individuals in $|\varphi|_T$, and that the number of investigated individuals is equal to $\sum_{i=0}^{t-1} \kappa_i(\varphi)$, he would not try to investigate more individuals than all un-investigated ones. This is required by the applicability condition of Definition 4.2a. Conditions 2b to 2d are obvious since these values are not affected by the investigation. What the investigation can affect is the total number of the investigated individuals in $|\varphi|_T$ and this is reflected in Condition 2e.

4. THE BASIC MODEL

To model the damage and compensation of privacy leakage, we create a simple game played between an agent, called the accuser, and an individual in U. The accuser tries to disseminate the private information of individuals. Assume that $(d_0, d_1, \ldots, d_{t-1}) \in \Re^t$ and $(c_0, c_1, \ldots, c_{t-1}) \in \Re^t$ are respectively the damage and compensation vectors of the game. If an individual is accused of s_i and he actually has the attribute value s_i, then his damage is d_i which is also the reward of the accuser. However, if he is accused of s_i and his private attribute value is not s_i, then he can receive compensation c_i from the accuser. Thus, if $\mathcal{I} = (\Pi, \mu, \kappa)$ is an information state, then the agent who wants to accuse an un-investigated individual satisfying $\varphi \in \Pi$ of s_i would have the risk of losing

$$L_i(\varphi) = (1 - p_i(\varphi))c_i - p_i(\varphi)d_i \qquad (1)$$

where $p_i(\varphi)$ is defined by

$$p_i(\varphi) = \frac{|\varphi| \cdot \mu_i(\varphi) - \kappa_i(\varphi)}{|\varphi| - \sum_{i=0}^{t-1} \kappa_i(\varphi)}. \tag{2}$$

The goal of privacy protection is then to make acquiring confidential information unprofitable for the accuser. This is done by raising his expected loss to a threshold level. The threshold level should be high enough to dissuade an agent (accuser). For convenience, we assume the threshold is zero. Thus, an information state $\mathcal{I} = (\Pi, \boldsymbol{\mu}, \boldsymbol{\kappa})$ is said to be *safe* if $L_i(\varphi) \geq 0$ for all $\varphi \in \Pi$ and $0 \leq i \leq t-1$.

EXAMPLE 1 Assume a person is being tested for a certain gene that increases the chance of suffering from some rare disease. If this person does have this gene, a potential employer may reject him if he acquires this information. Dissemination of this information could harm this person. Therefore, we should design some mechanism to make this information unprofitable for the potential employer. ■

A query can be answered only if doing so does not change an agent's information state to an unsafe one. An information state $\mathcal{I} = (\Pi, \boldsymbol{\mu}, \boldsymbol{\kappa})$ is safe if $p_i(\varphi) \leq \frac{c_i}{c_i+d_i}$ for any $\varphi \in \Pi$ and $0 \leq i \leq t-1$. However, since $p_i(\varphi)$ not only depends on $\mu_i(\varphi)$, but also on how many individuals have been investigated by the user, the data center cannot discern whether answering will maintain a safe state or not. To guarantee the safety of an information state, the data center can use worst-case analysis. Assume for each wff φ, the user can investigate at most K_φ individuals in $|\varphi|$ at an affordable cost. Then, given a partial knowledge state $\mathcal{P} = (\Pi, \boldsymbol{\mu})$ resulting from an answer to a query, the data center can guarantee safety, no matter which (affordable) investigation is made by the user, if the following condition holds for all $\varphi \in \Pi$ and $0 \leq i \leq t-1$:

$$\frac{|\varphi| \cdot \mu_i(\varphi)}{|\varphi| - K_\varphi} \leq \frac{c_i}{c_i + d_i}, \tag{3}$$

since by (2),

$$p_i(\varphi) \leq \frac{|\varphi| \cdot \mu_i(\varphi)}{|\varphi| - K_\varphi}.$$

Condition (3) can be rewritten as

$$\mu_i(\varphi) \leq \frac{c_i}{c_i + d_i} \cdot (1 - \frac{K_\varphi}{|\varphi|}). \tag{4}$$

Some cases in which Equation (4) is satisfied are given next.

1 If no investigative actions are possible ($K_\varphi = 0$), then (4) is satisfied if $\mu_i(\varphi) \leq \frac{c_i}{c_i+d_i}$. In this case, if $d_i = 0$ or $c_i \gg d_i$, then the information

state is safe even though $\mu_i(\varphi)$ is approximately equal to 1. This means that knowing that an individual is s_i will either not harm the individual $(d_i = 0)$, or compensation will be sufficient to cover the damage $(c_i \gg d_i)$. Hence, it does not matter if the accuser can be almost certain that a class of individuals has s_i value. On the other hand, if $c_i \approx d_i > 0$, then the information state is safe only when $\mu_i(\varphi)$ is less than 0.5. In other words, if compensation cannot sufficiently cover the damage, then the accuser should not make certain of the confidential value with a degree beyond 0.5.

2 If investigation is allowed for at most K_φ individuals, then the upper bound of $\mu_i(\varphi)$ is multiplied with the ratio $1 - \frac{K_\varphi}{|\varphi|}$ to maintain safety. The discount effect is alleviated when $|\varphi| \gg K_\varphi$. Thus, the larger the size of $|\varphi|$, the higher the possibility of achieving the safety requirement. This corresponds to the **k-anonymity** requirement for privacy protection in [15].

Based on the safety criterion, the data center can decide whether the query is answered or refused. Note that (4) is a sufficient condition for the safety of data release, so we may not have to test it for every i and φ. For example, if $d_i = 0$, then $L_i(\varphi) \geq 0$ holds, no matter how the investigation is carried out. We only have to test (4) for those i's such that $d_i > 0$.

In addition of refusing a query, the data center can use a pricing mechanism to discourage the user. To formulate the approach, we need two cost functions γ_{inv} and $\gamma_{acc} : \Phi \times \mathcal{Z}^+ \rightarrow \Re^+$ denoting respectively the cost of investigating k individuals satisfying φ and the cost of accusing k individuals satisfying φ. The minimum loss the user may incur under the partial knowledge state $\mathcal{P} = (\Pi, \mu)$ should then be

$$L^*(\varphi) = \min_{\mathbf{k},\mathbf{l}}[\sum_{i=0}^{t-1} L_i(\varphi, \mathbf{k}) \cdot l_i + \gamma_{inv}(\varphi, \sum_{i=0}^{t-1} k_i) + \gamma_{acc}(\varphi, \sum_{i=0}^{t-1} l_i)]$$

where

$$L_i(\varphi, \mathbf{k}) = c_i - (c_i + d_i) \cdot \frac{|\varphi| \cdot \mu_i(\varphi) - k_i}{|\varphi| - \sum_{i=0}^{t-1} k_i}$$

is the result of substituting (2) into (1) when $\kappa_i(\varphi) = k_i$ for $0 \leq i \leq t-1$. The minimization is taken over all $k_i \leq |\varphi| \cdot \mu_i(\varphi)$ for $0 \leq i \leq t-1$ and l_i such that $\sum_{i=0}^{t-1} l_i \leq |\varphi|$. If answers to a batch of queries result in the partial knowledge state (Π, μ), its price should be determined by $\sum_{\varphi \in \Pi} price(\varphi)$, where the price of each φ is the equation

$$price(\varphi) = \begin{cases} -L^*(\varphi), & \text{if } L^*(\varphi) < 0 \\ 0, & \text{otherwise.} \end{cases}$$

5. THE EXTENDED MODEL

In the basic model, we assumed the damage vector is associated with each specific value of the confidential attribute. This means that if an individual is known to have the attribute value s_i, then he will have damage d_i. Sometimes it is also harmful to an individual if his attribute value is known to be in some specific subset of V_c even if the subset is not a singleton.

EXAMPLE 2 Assume a fatal disease can be diagnosed and classified as a stage 0 – 5, where 0 is no disease, 1 through 3 are curable states, and 4 and 5 are deadly. Knowing that a person was diagnosed as stage 4 or 5 is harmful to that person. ■

Since it is reasonable that compensation is proportional to damage, we can simplify the model by assuming that there is a function $\alpha : \Re \rightarrow \Re$ that maps each damage value to its corresponding compensation. For example, it may be that $\alpha(x) = r \cdot x$ for some positive number r. We can concentrate on the estimate of damage in the extended model. We assume there is a damage function $\delta : (2^{\{0,\ldots,t-1\}} - \{\emptyset\}) \rightarrow \Re$. For any $S \subseteq \{0, \ldots, t-1\}$, $\delta(S)$ is the damage caused to an individual when it is known that his confidential attribute value belongs to $\{s_i \mid i \in S\}$. By using the game rule from the basic model, the expected loss of the agent accusing an individual in $\varphi \in \Pi$ of $\{s_i \mid i \in S\}$ would be

$$L_S(\varphi) = (1 - \sum_{i \in S} p_i(\varphi))\alpha(\delta(S)) - (\sum_{i \in S} p_i(\varphi))\delta(S)$$

where $p_i(\varphi)$ is defined in (2). The safety criterion for an information state $\mathcal{I} = (\Pi, \mu, \kappa)$ is extended to

$$L_S(\varphi) \geq 0$$

for all $\varphi \in \Pi$ and $S \subseteq \{0, \ldots, t-1\}$. This is equivalent to

$$\sum_{i \in S} p_i(\varphi) \leq \frac{\alpha(\delta(S))}{\alpha(\delta(S)) + \delta(S)}. \tag{5}$$

By using worst-case analysis of the basic model,

$$\sum_{i \in S} \frac{|\varphi| \cdot \mu_i(\varphi)}{|\varphi| - K_\varphi} \leq \frac{\alpha(\delta(S))}{\alpha(\delta(S)) + \delta(S)},$$

must be satisfied for all $\varphi \in \Pi$ and $S \subseteq \{0, \ldots, t-1\}$. Alternatively, this can be rewritten as

$$\sum_{i \in S} \mu_i(\varphi) \leq (\frac{\alpha(\delta(S))}{\alpha(\delta(S)) + \delta(S)}) \cdot (1 - \frac{K_\varphi}{|\varphi|}). \tag{6}$$

So far, the model does not address the issue of estimating the damage function δ. In fact, the damage vector in the basic model should be determined by an external mechanism, such as a legal system or a social convention, so we can assume that $\delta(\{i\}) = d_i$ for each $0 \leq i \leq t-1$. However, for a subset S other than singletons, it should be possible to impose some reasonable constraints so that $\delta(S)$ is (partially) determined by $\{\delta(\{i\}) \mid i \in S\}$.

EXAMPLE 3 Below are some possible conditions that the damage function $\delta : (2^{\{0,\ldots,t-1\}} - \{\emptyset\}) \rightarrow R$ should satisfy.

1 $\delta(\{0,\ldots,t-1\}) = 0$.

2 $\delta(S_1) \leq \delta(S_2)$ if $S_2 \subseteq S_1$.

3 $\delta(S) = 0$ if $|S| > 1$.

4 $\delta(S) = \min_{i \in S} \delta(\{i\})$.

Condition 1 ensures that if there is no privacy leakage, there is no damage. Since it is known that all possible values of the confidential attribute are in V_c, the index set $\{0,\ldots, t-1\}$ corresponds to the situation of no privacy leakage. Condition 2 means the more specific information is known, the more damage is caused. Condition 3 corresponds to the basic model in which only the damage value of the singleton is considered. Condition 4 is due to the principle of least commitment. The principle implies that if an individual is accused of a set of possible faults disjunctively, it can only be sure that he has the least harmful fault, so that the damage to him caused by such accusation would be equivalent to the minimal one of accusing him of a specific fault in the set. Note that Conditions 3 and 4 are not compatible if there are at least two indices i and j such that $\delta(\{i\}) > 0$ and $\delta(\{j\}) > 0$. However, both Conditions 1 and 2 are implied by Conditions 3, and Condition 4 implies Condition 2. Furthermore, Condition 4 also implies Condition 1 provided that i exists such that $\delta(\{i\}) = 0$. ■

An alternative way to estimate the damage value of a subset is by the information theoretic approach. If the *a priori* probability function on the possible values of the confidential attribute is given by $\mu(\top)$, then we can compute the *a posteriori* probability for any $S \subseteq \{0, 1, \ldots, t-1\}$ as

$$Pr(s_i|S) = \begin{cases} \frac{\mu_i(\top)}{\sum_{j \in S} \mu_j(\top)}, & \text{if } i \in S; \\ 0, & \text{otherwise.} \end{cases}$$

Then a possible constraint on the damage function is

$$\delta(S) = \sum_{i \in S} m(\mu_i(\top), Pr(s_i|S)) \cdot \delta(\{i\}),$$

where $m : [0,1] \times [0,1] \rightarrow [0,1]$ is called an information distance function. The information distance function estimates how the user's information on some specific s_i increases by knowing the index i is in S. Typically, the information distance function can be defined as the relative difference between the entropy values of the two probabilities, i.e.,

$$m(p,q) = \frac{\log p - \log q}{\log p}.$$

6. CONCLUSION AND RELATED WORKS

In this paper, we present a quantitative model for privacy protection. The model is based on a formal representation of the user information states. We assume that the damage and compensation of revealing each specific confidential value is known. An information state is safe when a user can discover a specific confidential value only with a sufficiently small probability if the damage of revealing the value is large.

Quantifying the value of information is by no means a new problem. However, quantitative models for privacy protection provide a new angle to view the problem. A standard concept of information value has been discussed in decision theory [5, 10]. The decision-theoretic concept of information value is applied to privacy protection in [7]. This paper follows the framework of [7], but assess information value from a different viewpoint. It must be emphasized that the value of information is defined with respect to the particular user model. When other user models are considered, the value of information may be different. Some examples can be found in [9].

Some quantitative criteria for privacy protection have been proposed in [2–4, 11, 17]. In [2, 3], information value is estimated by the expected cost the user must pay to achieve a perfect knowledge state from the given information. In [4, 11, 17], the paradigm of granular computing is applied to the definition of safety criteria.

In contrast to the quantitative approach of this paper, some qualitative criteria for privacy protection have been proposed in [6, 8, 13–16]. These criteria are designed to protect sensitive information in the release of a microdata set, i.e. a set of records containing information about individuals. The main objective is to avoid the re-identification of individuals or in other words, to prevent the possibility of deducing which record corresponds to a particular individual even though the explicit identifier of the individual is not contained in the released information. Our models are concerned with the release of statistical information, which is generally less specific than microdata. However, microdata release can also be handled by our framework when the queries are specific enough. Let us define a complete specification formula (CSF) as a DL wff of the form $\wedge_{a \in A}(a, v_a)$, where A is the set of all easy-to-know attributes

and v_a is a value in the domain of A. The answer to the batch of queries Q consisting of all CSF's is equivalent to the microdata release of the whole data table T.

The description of μ-ARGUS system [8] emphasized that re-identification of an individual can occur when the individual is rare in the population in respect to an easy-to-know attribute value. This is formulated as the *bin size* criterion in the Datafly system [14]. A bin is defined as an equivalence class of individuals who have exactly the same easy-to-know attribute values. The bin size criterion is that the size of each bin must be greater than some threshold level. To achieve the criterion, it may be necessary to generalize the data to a more imprecise level. These data modification techniques, mainly generalization and suppression, are formally investigated in [13, 15, 16]. In their framework, a formal requirement (called k-anonymity) is defined, and generalization and suppression techniques are employed to ensure that the requirement is satisfied. Both the bin size criterion and k-anonymity requirement can be easily enforced in our model if it is required that a query φ cannot be answered if size $|\varphi|$ is less than some threshold. However, instead of generalizing or suppressing the data, we try to assess the value or the damage of releasing such data, and discourage the misuse of the information by a pricing or penalty mechanism.

References

[1] L.J. Camp. *Trust and Risk in Internet Commerce.* The MIT Press, 2000.

[2] Y.C. Chiang, T.-s. Hsu, S. Kuo, C.J. Liau, and D.W. Wang. Preserving confidentiality when sharing medical database with the Cellsecu system. *International Journal of Medical Informatics,* 71:17–23, 2003.

[3] Y.C. Chiang, T.-s. Hsu, S. Kuo, and D.W. Wang. Preserving confidentially when sharing medical data. In *Proceedings of Asia Pacific Medical Informatics Conference,* 2000.

[4] Y.T. Chiang, Y.C. Chiang, T.-s. Hsu, C.J. Liau, and D.W. Wang. How much privacy? - a system to safe guard personal privacy while releasing database. In *Proceedings of the 3rd International Conference on Rough Sets and Current Trends in Computing,* LNCS 2475, pages 226–233. Springer-Verlag, 2002.

[5] G.D. Eppen and F.J. Gould. *Quantitative Concepts for Management.* Prentice Hall, 1985.

[6] T.-s. Hsu, C.J. Liau, and D.W. Wang. A logical model for privacy protection. In *Proceedings of the 4th International Conference on Information Security,* LNCS 2200, pages 110–124. Springer-Verlag, 2001.

[7] T.-s. Hsu, C.J. Liau, D.W. Wang, and Jeremy K.P. Chen. Quantifying privacy leakage through answering database queries. In *Proceedings of the 5th International Conference on Information Security,* LNCS 2433, pages 162–175. Springer-Verlag, 2002.

[8] A.J. Hundepool and L.C.R.J. Willenborg. "μ- and τ-ARGUS: Software for statistical disclosure control". In *Proceedings of the 3rd International Seminar on Statistical Confidentiality,* 1996.

[9] J. Kleinberg, C.H. Papadimitriou, and P. Raghavan. "On the value of private information". In *Proc. 8th Conf. on Theoretical Aspects of Rationality and Knowledge,* 2001.

[10] D.V. Lindley. *Making Decisions.* John Wiley & Sons, 1985.

[11] A. Ohrn and L. Ohno-Machado. "Using Boolean reasoning to anonymize databases". *Artificial Intelligence in Medicine,* 15:235–254, 1999.

[12] Z. Pawlak. *Rough Sets–Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers, 1991.

[13] P. Samarati. "Protecting respondents' identities in microdata release". *IEEE Transactions on Knowledge and Data Engineering,* 13(6):1010–1027, 2001.

[14] L. Sweeney. "Guaranteeing anonymity when sharing medical data, the Datafly system". In *Proceedings of American Medical Informatics Association,* 1997.

[15] L. Sweeney. "Achieving **k-Anonymity** privacy protection using generalization and suppression". *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems,* 10(5):571–588, 2002.

[16] L. Sweeney. **"k-Anonymity:** A model for protecting privacy". *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems,* 10(5):557–570, 2002.

[17] D.W. Wang, C.J. Liau, and T.-s. Hsu. "Medical privacy protection based on granular computing". *Artificial Intelligence in Medicine,* to appear, 2004.