The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing

Justin Brickell and Vitaly Shmatikov
Department of Computer Sciences
The University of Texas at Austin
jlbrick@cs.utexas.edu, shmat@cs.utexas.edu

ABSTRACT

Re-identification is a major privacy threat to public datasets containing individual records. Many privacy protection algorithms rely on generalization and suppression of "quasi-identifier" attributes such as ZIP code and birthdate. Their objective is usually *syntactic* sanitization: for example, k-anonymity requires that each "quasi-identifier" tuple appear in at least k records, while ℓ-diversity requires that the distribution of sensitive attributes for each quasi-identifier have high entropy. The utility of sanitized data is also measured syntactically, by the number of generalization steps applied or the number of records with the same quasi-identifier.

In this paper, we ask whether generalization and suppression of quasi-identifiers offer any benefits over trivial sanitization which simply separates quasi-identifiers from sensitive attributes. Previous work showed that k-anonymous databases can be useful for data mining, but k-anonymization does not guarantee any privacy. By contrast, we measure the tradeoff between privacy (how much can the adversary learn from the sanitized records?) and utility, measured as accuracy of data-mining algorithms executed on the same sanitized records.

For our experimental evaluation, we use the same datasets from the UCI machine learning repository as were used in previous research on generalization and suppression. Our results demonstrate that even modest privacy gains require almost complete destruction of the data-mining utility. In most cases, trivial sanitization provides equivalent utility and better privacy than k-anonymity, ℓ-diversity, and similar methods based on generalization and suppression.

Categories and Subject Descriptors

H.2.7 [**Database Management**]: Database Administration—*Security, integrity, and protection*; H.2.8 [**Database Management**]: Database Applications—*Data mining*

General Terms

Algorithms, Security

1. INTRODUCTION

Microdata records contain information about specific individuals. Examples include medical records used in public-health research, individual transactions or preferences released to support the development of new data-mining algorithms, and records published to satisfy legal requirements.

In contrast to statistical databases and randomized response methods, the records in question contain actual, unperturbed data associated with individuals. Some of the attributes may be sensitive, *e.g.*, health-related attributes in medical records. Therefore, identifying attributes such as names and Social Security numbers are typically removed from microdata records prior to release. The published records may still contain "quasi-identifiers," *e.g.*, demographic attributes such as ZIP code, age, or sex. Even though the quasi-identifier attributes do not directly reveal a person's identity, they may appear together with the identity in another public database, or it may be easy to reconstruct their values for any given individual. Microdata records may also contain "neutral" attributes which are neither quasi-identifying, nor sensitive.

The association of quasi-identifiers with sensitive attributes in public records has long been recognized as a privacy risk [17,32]. This type of privacy breach is known as *sensitive attribute disclosure*, and is different from *membership disclosure*, *i.e.*, learning whether a certain individual is included in the database [12, 26, 30].

It is very easy to prevent sensitive attribute disclosure by simply not publishing quasi-identifiers and sensitive attributes together. *Trivial sanitization* that removes either all quasi-identifiers, or all sensitive attributes in each data release provides the maximum privacy possible against an adversary whose knowledge about specific individuals is limited to their quasi-identifiers (this adversary is very weak, yet standard in the microdata sanitization literature [10,22,34]).

There is large body of research on techniques such as k-anonymity and ℓ-diversity that apply domain-specific generalization and suppression to quasi-identifier attributes and then publish them together with unmodified sensitive attributes. In this paper, we ask a basic question: **what benefit do these algorithms provide over trivial sanitization?** The only reason to publish generalized quasi-identifiers and sensitive attributes together is to support data-mining tasks that consider both types of attributes in the sanitized database. Our goal in this paper is to evaluate the tradeoff between this incremental gain in data-mining utility and the degradation in privacy caused by publishing quasi-identifiers together with sensitive attributes.

KDD'08, August 24–27, 2008, Las Vegas, Nevada, USA.
Copyright 2008 ACM 978-1-60558-193-4/08/08 ...$5.00.

Our contributions. First, we give a *semantic* definition of sensitive attribute disclosure. It captures the gain in the adversary's knowledge due to his observations of the sanitized dataset. This definition is somewhat similar to privacy definitions used in random-perturbation databases [13], but is adapted to the generalization and suppression framework.

Second, we give a methodology for measuring the tradeoff between the loss of privacy and the gain of utility. Privacy loss is the increase in the adversary's ability to learn sensitive attributes corresponding to a given identity. Utility gain is the increase in the accuracy of machine-learning tasks evaluated on the sanitized dataset. The baseline for both is the *trivially sanitized* dataset, which simply omits either all quasi-identifiers, or all sensitive attributes, thus providing maximum privacy and minimum utility.

Third, we evaluate our methodology on the same datasets from the UCI machine learning repository as used in previous research on sanitized microdata utility [20–22]. We show that non-trivial generalization and suppression either results in large privacy breaches, or provides little incremental utility vs. a trivially sanitized dataset. Therefore, even if the adversary's knowledge is limited to quasi-identifiers, the data-mining utility must be destroyed to achieve only marginal privacy. To protect against an adversary with auxiliary knowledge, the loss of utility must be even greater.

2. RELATED WORK

Privacy in *statistical* databases has been a topic of much research [2,35]. Techniques include adding random noise to the data while preserving certain statistical aggregates [4,8,13] and interactive output perturbation [6,11].

By contrast, *microdata* publishing involves releasing unperturbed records containing information about individuals. k-anonymity is a popular interpretation of privacy [10,31,34]. Many methods have been proposed for achieving it [5,14,15,18,19,27,29,33,41]; most apply generalization and suppression to quasi-identifiers only. In Section 6, we compare our experimental methodology to previous work.

Limitations of k-anonymity are: (1) it does not hide whether a given individual is in the database [26,30], (2) it reveals individuals' sensitive attributes [21,22], (3) it does not protect against attacks based on background knowledge [22,23], (4) mere knowledge of the k-anonymization algorithm can violate privacy [43], (5) it cannot be applied to high-dimensional data without complete loss of utility [3], and (6) special methods are required if a dataset is anonymized and published more than once [7,37,40].

In [28], Øhrn and Ohno-Machado proposed that the sensitive attributes associated with each quasi-identifier be "diverse." This is similar to p-sensitivity [36], ℓ-diversity [22], and others [39,42]. Diversity of sensitive attributes, however, is neither necessary, nor sufficient to prevent sensitive attribute disclosure (see [21] and Section 4). A stronger definition appears in [24], but it is unclear whether it can be achieved in the data access model considered in the generalization and suppression framework.

In the k-anonymity literature, the adversary's knowledge is limited to quasi-identifiers such as age and ZIP code. Stronger adversaries with background knowledge are considered in [9,23]. Our results show that generalization and suppression do not protect privacy even against very weak adversaries who only know the quasi-identifiers; privacy obviously fails against stronger adversaries as well.

This paper is about sensitive attribute disclosure. Membership disclosure, *i.e.*, learning whether a given individual is present in the sanitized database, is a different, incomparable privacy property. Methods for preventing membership disclosure such as [12,26,30] are complementary to our work.

3. DEFINITIONS AND NOTATION

Let $T = \{t_1, \ldots t_n\}$ be a data table. Each t_i is a tuple of attribute values representing some individual's record. Let $\mathcal{A} = \{a_1, \ldots a_m\}$ be the set of attributes; $t[a_i]$ denotes the value of attribute a_i for tuple t. We use the following notation for subsets of attributes and tuples. If $\mathcal{C} = \{c_1, c_2, \ldots c_p\} \subseteq A$, then $t[C]$ denotes $(t[c_1], \ldots t[c_p])$. If $U = \{u_1, u_2, \ldots u_p\} \subseteq T$, then $U[a]$ denotes $(u_1[a], \ldots u_p[a])$.

Let $\mathcal{S} \in \mathcal{A}$ be the *sensitive attribute*. This is an attribute whose value the adversary should not be able to associate with an individual (*e.g.*, medical information). Let $S = \{s_1, \ldots s_l\}$ be the set of possible attribute values for the sensitive attribute $\mathcal{S}$. All of the concepts in this paper are easily explained in the single sensitive attribute setting, but can also be generalized to multiple sensitive attributes.

Let $\mathcal{Q} \in \mathcal{A} \setminus \mathcal{S}$ be the *quasi-identifier*, *i.e.*, the set of non-sensitive (*e.g.*, demographic) attributes whose values may be known to the adversary for a given individual.

Two tuples t_i and t_j are $\mathcal{Q}$-equivalent (denoted $t_i \overset{\mathcal{Q}}{\equiv} t_j$) if $t_i[\mathcal{Q}] = t_j[\mathcal{Q}]$. This equivalence relation partitions T into quasi-identifier equivalence classes, denoted as $\langle t_j \rangle$, where $t_i \in \langle t_j \rangle$ iff $t_i \overset{\mathcal{Q}}{\equiv} t_j$. Let $\mathcal{E}_\mathcal{Q} \subseteq T$ be a set of representative records for each equivalence class imposed by $\overset{\mathcal{Q}}{\equiv}$.

We make the standard assumption that the adversary knows only the quasi-identifiers [10,18,34,41]. This weak adversary model makes our results *stronger* because if privacy fails against the weak adversary, it will also fail against adversaries who have additional knowledge [22,23].

T may also contain attributes in $\mathcal{A} \setminus (\mathcal{Q} \cup \mathcal{S})$, which are neither sensitive, nor quasi-identifying. For example, a user may wish to construct a classifier that predicts the values of these "neutral" attributes (*e.g.*, length of hospital stay) on the basis of both quasi-identifiers (*e.g.*, age) and sensitive attributes (*e.g.*, diagnosis).

Consider a subset of tuples $U = \{u_1, u_2, \ldots u_p\} \subseteq T$, and the distribution of sensitive attribute values within U. For any sensitive attribute value s, denote by U_s the set $\{u \in U \mid u[\mathcal{S}] = s\}$ of tuples in U whose sensitive attribute value is equal to s, and denote by $p(U, s)$ the corresponding fraction of tuples in U, computed as $\frac{|U_s|}{|U|}$. The notation $p(U, s)$ can be understood as "the probability that a randomly chosen member of U has sensitive attribute value s."

We assume that whenever an adversary is provided with a sanitized table T', the record rows appear in a random order to prevent "unsorted matching attacks" [34].

4. SENSITIVE ATTRIBUTE DISCLOSURE

Sensitive attribute disclosure occurs when the adversary learns information about an individual's sensitive attribute(s). This form of privacy breach is different and incomparable to learning whether an individual is included in the database, which is the focus of differential privacy [12].

To obtain a meaningful definition of data privacy, it is necessary to quantify the knowledge about sensitive attributes that the adversary gains from observing the sanitized database. We call our definitions *semantic* because they capture this shift in the adversary's knowledge. The need for semantic definitions of privacy is well-understood for random-perturbation databases (*e.g.*, [13]). By contrast, research on microdata privacy has focused on purely *syntactic* privacy definitions such as k-anonymity and ℓ-diversity (surveyed below), which only consider the distribution of attribute values in the sanitized database, without directly measuring what the adversary may learn.

4.1 Attack model

We use the standard model from the literature [10, 22]. The adversary is given a sanitized table T' generated from an original table T, and the quasi-identifier $t[\mathcal{Q}]$ for some target individual t known to be in the table T (*i.e.*, we are not considering membership disclosure). We re-emphasize that giving the adversary *more* background knowledge will result in even worse disclosure than we demonstrate.

To keep the sanitized database "truthful" [31, 34], generalization and suppression are applied only to quasi-identifiers, with sensitive attributes left intact. Therefore, the most "private" sanitized table possible with this approach is the trivial sanitization in which all $\mathcal{Q}$ are suppressed. Equally effective is the trivial sanitization in which all $\mathcal{S}$ are suppressed (and released in a separate, unlinked table).

The adversary's *baseline knowledge* $\mathcal{A}_{\mathsf{base}}$ is the minimum information about sensitive attributes that he can learn after any sanitization, including trivial sanitization which releases quasi-identifiers and sensitive attributes separately. $\mathcal{A}_{\mathsf{base}}$ is the distribution of sensitive attributes in the original table, which is revealed by *any* generalization and suppression algorithm because sensitive attributes are left untouched to keep them "truthful." We are concerned about privacy leaks *in excess of* this baseline knowledge; for example, if 90% of the individuals in T have cancer, then it should *not* be considered an attribute disclosure if the adversary concludes that t has cancer with probability 90%, since this baseline distribution is always revealed to the adversary. We formally define $\mathcal{A}_{\mathsf{base}}$ as the vector of probabilities representing the distribution of sensitive attribute values in the entire table T: $\mathcal{A}_{\mathsf{base}} = \langle p(T, s_1), p(T, s_2), \ldots, p(T, s_l)\rangle$.

The adversary's *posterior knowledge* $\mathcal{A}_{\mathsf{san}}$ is what he learns from the sanitized table T' about the sensitive attributes of his target individual $t \in T$. Unlike $\mathcal{A}_{\mathsf{base}}$, $\mathcal{A}_{\mathsf{san}}$ takes quasi-identifiers into account, because the records in T' contain a mixture of generalized and suppressed quasi-identifiers. Because the generalization hierarchy on quasi-identifiers is required to be totally ordered [10], the adversary can uniquely identify the quasi-identifier equivalence class $\langle t\rangle$ containing the sanitized record of t in T'. $\mathcal{A}_{\mathsf{san}}$ is the distribution of sensitive attribute values within this class $\langle t\rangle$: $\mathcal{A}_{\mathsf{san}}(\langle t\rangle) = \langle p(\langle t\rangle, s_1), p(\langle t\rangle, s_2), \ldots, p(\langle t\rangle, s_l)\rangle$.

Sensitive attribute disclosure is the difference between the adversary's posterior knowledge $\mathcal{A}_{\mathsf{san}}$ and his baseline knowledge $\mathcal{A}_{\mathsf{base}}$. It can measured additively or multiplicatively.

$$\mathcal{A}_{\mathsf{diff}}(\langle t\rangle) = \tfrac{1}{2}\textstyle\sum_{i=1}^{l} |p(T, s_i) - p(\langle t\rangle, s_i)|$$
$$\mathcal{A}_{\mathsf{quot}}(\langle t\rangle) = \left|\log \tfrac{p(\langle t\rangle, s)}{p(T,s)}\right|$$

Informally, it captures how much *more* the adversary learns by observing sanitized quasi-identifiers than he would have learned from a "maximally private" database where sensitive attributes are separated from the quasi-identifiers.

4.2 Semantic privacy

To capture the incremental gain in the adversary's knowledge caused by the sanitized table T', we first consider his baseline knowledge $\mathcal{A}_{\mathsf{base}}$ as defined above. Recall that it consists of the distribution of sensitive attributes in the table T^*, where all quasi-identifiers have been suppressed (any sanitization that does not touch sensitive attributes necessarily reveals T^*). Furthermore, the adversary knows $t[\mathcal{Q}]$ for all $t \in T$, *i.e.*, the quasi-identifier attribute values for all individuals in the database. The adversary can easily learn these values from external databases and other resources.

Definition 1 (δ-disclosure privacy). *We say that an equivalence class $\langle t\rangle$ is δ-disclosure-private with regard to the sensitive attribute S if, for all $s \in S$*

$$\mathcal{A}_{\mathsf{quot}}(\langle t\rangle) = \left|\log \frac{p(\langle t\rangle, s)}{p(T, s)}\right| < \delta$$

A table T is δ-disclosure-private if for every $t \in \mathcal{E}_{\mathcal{Q}}$, $\langle t\rangle$ is δ-disclosure private.

Intuitively, a table is δ-disclosure private if the distribution of sensitive attribute values within each quasi-identifier class is roughly the same as their distribution in the entire table. In contrast to [21], we use a multiplicative definition. It correctly models disclosures when some value of the sensitive attribute occurs in certain quasi-identifier classes, but not in others. It also allows us to derive a bound on the gain in adversarial knowledge, by relating the δ parameter to information gain used by decision tree classifiers such as ID3 and C4.5. $\mathsf{Gain}(\mathcal{S}, \mathcal{Q})$ is defined as the difference between the entropy of $\mathcal{S}$ and the conditional entropy $H(\mathcal{S}|\mathcal{Q})$.

$$\mathsf{Gain}(\mathcal{S}, \mathcal{Q}) = H(\mathcal{S}) - H(\mathcal{S}|\mathcal{Q})$$

Lemma 1. *If T satisfies δ-disclosure privacy, then $Gain(\mathcal{S}, \mathcal{Q}) < \delta$. Let $\alpha_s = p(T, s)$ and let $\beta_{t,s} = p(\langle t\rangle, s)$. Note that $\alpha_s = \sum_{t\in\mathcal{E}_{\mathcal{Q}}} \frac{|\langle t\rangle|}{|T|}\beta_{t,s}$.*

Proof:

$$\begin{aligned}
&\mathrm{Gain}(\mathcal{S}, \mathcal{Q})\\
=\;& \sum_{s\in S} -\alpha_s \log \alpha_s - \sum_{t\in\mathcal{E}_{\mathcal{Q}}} \frac{|\langle t\rangle|}{|T|} \sum_{s\in S} -\beta_{t,s} \log \beta_{t,s}\\
=\;& \sum_{s\in S} \sum_{t\in\mathcal{E}_{\mathcal{Q}}} -\frac{|\langle t\rangle|}{|T|}\beta_{t,s} \log \alpha_s - \sum_{t\in\mathcal{E}_{\mathcal{Q}}} \frac{|\langle t\rangle|}{|T|} \sum_{s\in S} -\beta_{t,s} \log \beta_{t,s}\\
=\;& \sum_{t\in\mathcal{E}_{\mathcal{Q}}} \frac{|\langle t\rangle|}{|T|} \sum_{s\in S} (-\beta_{t,s} \log \alpha_s + \beta_{t,s} \log \beta_{t,s})\\
=\;& \sum_{t\in\mathcal{E}_{\mathcal{Q}}} \frac{|\langle t\rangle|}{|T|} \sum_{s\in S} \beta_{t,s} \log \frac{\beta_{t,s}}{\alpha_s}\\
<\;& \sum_{t\in\mathcal{E}_{\mathcal{Q}}} \frac{|\langle t\rangle|}{|T|} \sum_{s\in S} \beta_{t,s} \cdot \delta \;=\; \frac{\delta}{|T|} \sum_{t\in\mathcal{E}_{\mathcal{Q}}} |\langle t\rangle| \sum_{s\in S} \frac{|\langle t\rangle_s|}{|\langle t\rangle|}\\
=\;& \frac{\delta}{|T|} \sum_{t\in\mathcal{E}_{\mathcal{Q}}} \sum_{s\in S} |\langle t\rangle_s| \;=\; \delta
\end{aligned}$$

Lemma 1 shows that when a database satisfies δ-disclosure privacy, the ability to build a predictor for sensitive attributes $\mathcal{S}$ based on the quasi-identifier $\mathcal{Q}$ is bounded by δ. Note that definition 1 is *stronger* than the bound given by lemma 1, because it requires that the distributions $\mathcal{A}_{\mathsf{base}}$ and $\mathcal{A}_{\mathsf{san}}$ be similar, rather than just have similar entropies.

4.3 Syntactic privacy

k-anonymity. k-anonymity is based on the observation that *identity disclosure* can lead to *sensitive attribute disclosure*: if an adversary can determine which database record corresponds to the target individual, then he can determine this individual's sensitive attribute value.

Definition 2 (k-anonymity [31, 34]). *Table T is k-anonymous if and only if for each $t_j \in \mathcal{E}_\mathcal{Q}$, $|\langle t_j \rangle| \geq k$*

k-anonymity does not prevent sensitive attribute disclosure [22, 36], because an individual may belong to an equivalence class in which sensitive attributes have a low-entropy distribution. The attacker then learns the sensitive attribute *without* learning which record belongs to the individual.

ℓ-diversity. To prevent homogeneity attacks, sensitive attribute values within each equivalence class should be "diverse." This was observed in [28] and in [22].

Definition 3 (Recursive (c, ℓ)-diversity [22]). *Let r_i denote the number of times the ith most frequent sensitive value appears in $\langle t_i \rangle$. Given a constant c, $\langle t_i \rangle$ satisfies recursive (c, ℓ)-diversity if $r_1 < c(r_\ell + r_{\ell+1} + \ldots + r_m)$. A table T satisfies recursive (c, ℓ)-diversity if for every $t_i \in \mathcal{Q}$-groups(T), $\langle t_i \rangle$ satisfies recursive (c, ℓ)-diversity. We say that $(c, 1)$-diversity is always satisfied.*

Unfortunately, ℓ-diversity is neither necessary, nor sufficient to prevent sensitive attribute disclosure [21]. While it prevents an attacker from learning a sensitive attribute *exactly*, it can still reveal a lot of probabilistic information.

For example, consider a database in which 1% of individuals have a rare form of cancer and the quasi-identifier equivalence class $\langle t_i \rangle$ in which, say, 30% have this form of cancer (or any high percentage, as required by the diversity criterion). If the adversary's target individual $t \in \langle t_i \rangle$, then the adversary can immediately infer that his target is far more likely to have this form of cancer than a random individual in the database. In general, probabilistic sensitive attribute disclosure occurs whenever an attribute which is *not* diverse in the overall sensitive attribute distribution appears diverse in some quasi-identifier equivalence class. On the other hand, if only 1% of individuals in the equivalence class have this form of cancer, then there is no sensitive attribute disclosure, even though the class is *not* diverse.

t-closeness. In [21], Li *et al.* assume that the adversary knows the distribution $\mathcal{A}_{\mathsf{base}}$ of sensitive attribute values over the entire table, which is a reasonable assumption because this information would have been revealed even if the quasi-identifiers had been completely suppressed.

If the adversary determines that the target individual is in an equivalence class with a sensitive attribute value distribution significantly different from $\mathcal{A}_{\mathsf{base}}$, then he learns a lot of information about the individual. The goal of t-closeness is to ensure that these distributions are never too different.

Definition 4 (t-closeness [21]). *An equivalence class $\langle t_i \rangle$ has t-closeness if the distance between the distribution of a sensitive attribute in this class $\mathcal{A}_{\mathsf{san}}(\langle t \rangle)$ and the distribution of the attribute in the whole table $\mathcal{A}_{\mathsf{base}}$ is no more than a threshold t. A table has t-closeness if all equivalence classes have t-closeness.*

The critical question is how to measure the distance between distributions. In [21], the *Earth Mover's distance* (EMD) is used, which for nominal attributes is equivalent to $\mathcal{A}_{\mathsf{diff}}$. This is an additive (as opposed to multiplicative) measure, and does not translate directly into a bound on the adversary's ability to learn sensitive attributes associated with a given quasi-identifier. By contrast, Lemma 1 gives such a bound for our semantic privacy definition.

Even though t-closeness does not directly bound the gain in adversary's knowledge, it is similar in its spirit to semantic privacy; it, too, attempts to capture the *difference* between the adversary's baseline knowledge and the knowledge he gains from the quasi-identifier equivalence classes in the sanitized table. As parameters (t and δ, respectively) approach 0, both t-closeness and our definition 1 converge to statistical independence of quasi-identifiers and sensitive attributes within the sanitized database.

4.4 Measuring privacy of sanitized databases

Semantic privacy definitions, such as our definition 1, bound sensitive attribute disclosure, but an actual database instance may have less sensitive attribute disclosure (and thus more privacy) than permitted by the definition.

Conventional privacy metrics rely on syntactic properties of the sanitized dataset: number of records with the same quasi-identifier (k-anonymity) or frequency of sensitive attributes within each quasi-identifier class (ℓ-diversity). Unfortunately, the two metrics are incomparable. In [22], k and ℓ are compared directly, even though the two have different domains: k can vary from 1 to the total number of records, while ℓ can vary from 1 to the number of different sensitive attribute values. For example, a 1000-record database with a binary sensitive attribute can never be more than 2-diverse, but it can be anywhere up to 1000-anonymous.

We propose two different metrics to quantify attribute disclosure allowed by a sanitized database T' as opposed to T^* where all quasi-identifiers have been trivially suppressed. The first is based on the attribute disclosure distance $\mathcal{A}_{\mathsf{diff}}$:

$$\mathcal{A}_{\mathsf{know}} = \frac{1}{|T|} \sum_{t \in \mathcal{E}_\mathcal{Q}} |\langle t \rangle| \cdot \mathcal{A}_{\mathsf{diff}}(\langle t \rangle)$$

$\mathcal{A}_{\mathsf{know}}$ stands for "adversarial knowledge gain." It is the average amount of information about the sensitive attributes of individual t that the adversary learns because he is able to identify the class $\langle t \rangle$ based on t's quasi-identifier.

One may also consider a metric based on $\mathcal{A}_{\mathsf{quot}}$, but only semantically private databases achieve a finite privacy score. Other privacy definitions allow sensitive attribute values to be absent from some quasi-identifier classes, enabling the adversary to learn with certainty that the corresponding individual does not have this value.

The second metric quantifies the adversary's ability to predict his target t's sensitive attribute using his best strategy, which is to guess the most common sensitive attribute in $\langle t \rangle$. For a quasi-identifier class $\langle t \rangle$, let $s_{\mathsf{max}}(\langle t \rangle)$ be the most common sensitive attribute value found in $\langle t \rangle$. Then,

$$\mathcal{A}_{\mathsf{acc}} = \left(\frac{1}{|T|} \sum_{t \in \mathcal{E}_\mathcal{Q}} |\langle t \rangle| \cdot p\left(\langle t \rangle, s_{\mathsf{max}}(\langle t \rangle)\right) \right) - p(T, s_{\mathsf{max}}(T))$$

$\mathcal{A}_{\mathsf{acc}}$ stands for "adversarial accuracy gain" and measures the increase in the adversary's accuracy after he observes the sanitized database T' compared to his baseline accuracy from observing T^*, which is the most private database that can be obtained by generalization and suppression.

$\mathcal{A}_{\mathsf{acc}}$ *underestimates* the amount of information leaked by the sanitized table T', because it does not consider shifts in the probabilities of non-majority sensitive attributes. It is still a useful metric because it can be directly compared to our metrics of data-mining utility, described in Section 5.

5. MEASURING UTILITY

Utility of any dataset, whether sanitized or not, is innately tied to the computations that one may perform on it. For example, a census dataset may support an extremely accurate classification of income based on education, but not enable clustering based on household size. Without a workload context, it is meaningless to say whether a dataset is "useful" or "not useful," let alone to quantify its utility.

Nevertheless, the stated goal of privacy-preserving microdata publishing is to produce sanitized datasets that have "good" utility for a large variety of workloads. The unknown workload is an essential premise—if the workloads were known in advance, the data publisher could simply execute them on the original data and publish just the results instead of releasing a sanitized version of the data.

The need for a workload-independent measure of utility has led to the use of syntactic properties as a proxy for utility. One approach is to minimize the amount of generalization and suppression applied to the quasi-identifier attributes to achieve a given level of privacy [10]. This "minimization" is done with respect to absolute difference, relative distance, maximum distribution, or minimum suppression. Other syntactic metrics include the number of generalization steps, average size of quasi-identifier equivalence classes, the sum of squares of class sizes [22], and preservation of marginals [16].

Workload-independent metrics quantify the "damage" caused by sanitization, but they do not measure how much utility remains. For example, small quasi-identifier equivalence classes do not imply anything about the accuracy of classifiers that one may compute on the sanitized data [25].

It has been recognized that utility of sanitized databases must be measured empirically, in terms of specific workloads such as classification algorithms [15, 20, 38]. This does not necessarily contradict the "unknown workload" premise of sanitization. It simply acknowledges that even when sanitization satisfies a syntactic damage minimization requirement, it may still destroy the utility of a dataset for certain tasks; it is thus essential to measure the latter when evaluating effectiveness of various sanitization methods.

We can assume that users of the sanitized database are interested in workloads that take advantage of attribute correlations within the database, *e.g.*, construction of classifiers. For workloads which consider attributes in isolation, the data publisher can achieve maximum privacy by simply publishing two tables, one with the permuted quasi-identifiers, the other with the remaining attributes since they cannot be linked to the quasi-identifiers. Intuitively, utility of a sanitized database should be measured by how well cross-attribute correlations are preserved after sanitization.

It is critically important to measure *both* privacy and utility using the same methodology. Otherwise, maximizing utility may lead to privacy violations. For example, if utility is measured as the ability to predict sensitive attributes from the quasi-identifiers, then it is exactly the same as adversarial sensitive attribute disclosure! Iyengar [15] concludes that classification accuracy is maximized when attributes are *homogeneous* within each quasi-identifier group: this directly contradicts the *diversity* requirement [22, 36]. Similarly, [41] says that the data publishing process should preserve correlation between quasi-identifiers and sensitive attributes. This contradicts both diversity and semantic privacy, and immediately leads to sensitive attribute disclosure.

We aim to measure the tradeoffs between privacy and utility in a single framework, using semantic definitions for both: privacy in terms of adversarial sensitive attribute disclosure, utility in terms of concrete machine-learning tasks.

First, for a given workload w, we measure workload-specific utility of trivially sanitized datasets, *i.e.*, datasets from which either all quasi-identifiers $\mathcal{Q}$, or all sensitive attributes $\mathcal{S}$ have been removed. Both provide the maximum privacy achievable using generalization and suppression. Let $\mathcal{U}^{(w)}_{\mathsf{base}}$ be the corresponding empirical utility (to compute $\mathcal{U}^{(w)}_{\mathsf{base}}$, we pick the trivial sanitization with the largest utility). We give specific workloads and utility metrics in Section 6; for example, when the workload w involves computing a classifier, $\mathcal{U}^{(w)}_{\mathsf{base}}$ is the accuracy of this classifier.

Then, we consider several non-trivially sanitized tables T', one for each value of the sanitization parameter. For each table, we compute its workload-specific utility $\mathcal{U}^{(w)}_{\mathsf{san}}$.

The critical metric is $\mathcal{U}^{(w)}_{\mathsf{san}} - \mathcal{U}^{(w)}_{\mathsf{base}}$. This is the incremental utility gain provided by the release of non-trivially sanitized data. Note that if $\mathcal{U}^{(w)}_{\mathsf{san}} - \mathcal{U}^{(w)}_{\mathsf{base}}$ is close to 0, then non-trivial sanitization is pointless for this specific workload. In this case, a trivial sanitization which suppresses all $\mathcal{Q}$ or removes all $\mathcal{S}$ provides as much utility as any sophisticated sanitization algorithm while providing as much privacy as possible.

Another metric we'll employ is $\mathcal{U}^{(w)}_{\mathsf{max}}$, the utility of workload w as measured on the original, pre-sanitization database. If $\mathcal{U}^{(w)}_{\mathsf{max}}$ is low (*e.g.*, the corresponding classifier has low accuracy), this means that the workload is inappropriate for the data regardless of sanitization. It does not make sense to measure utility in terms of this workload, because even if the users had been given the entire original database, the utility would have been low.

6. EXPERIMENTS

Our experiments demonstrate that a trivial sanitizer which simply suppresses all quasi-identifiers or all sensitive attributes produces datasets with equivalent utility and better privacy (or equivalent privacy and better utility) than non-trivial generalization and suppression.

This appears to contradict previous work. For example, it was shown that useful machine-learning workloads can be evaluated on k-anonymous datasets [15, 20]. Of course, k-anonymity is neither necessary, nor sufficient for privacy. The "useful" datasets in question simply don't prevent sensitive attribute disclosure.

At the other end of the spectrum, ℓ-diversity [22] and t-closeness [21] do limit sensitive attribute disclosure. Utility, however, is measured syntactically, by the number of generalization steps applied to quasi-identifiers, average size of quasi-identifier equivalence classes, sum of squares of class sizes, or preservation of marginals. In contrast to this paper, the actual *data-mining* utility is not measured.

Wang *et al.* [38] give a sanitization which ensures a strong privacy definition *and* better data-mining utility on the UCI

Attribute	Values	Generalization
Age	74	continuous
Workclass	7	hierarchy
Education	16	continuous
Marital Status	7	hierarchy
Occupation	14	hierarchy
Race	5	hierarchy
Sex	2	hierarchy
Native Country	41	hierarchy
Salary	2	hierarchy

Table 1: Summary of the UCI "Adult" dataset.

Adult dataset than simple removal of all sensitive attributes. They do not consider the other trivial sanitization, which is to remove all quasi-identifiers. We repeated their experiments and observed that their sanitization does not provide significantly better utility than the trivially sanitized dataset consisting of sensitive attributes only.

6.1 Achieving semantic privacy

Semantic privacy, as defined in Section 4.2, is easily incorporated into k-anonymity frameworks such as Incognito [18]. Like ℓ-diversity [22] and t-closeness [21], semantic privacy has the *monotonicity property*: a generalization of a semantically private table is itself semantically private.

We used the implementation of generalization and suppression from LeFevre *et al.* [20], and modified the constraint checking portion of the code to support recursive (c, ℓ)-diversity (c=3 in all of our tests), t-closeness for nominal sensitive attributes, and semantic privacy from this paper (in the figures, s stands for δ from definition 1). This implementation is "workload-aware," *i.e.*, when choosing quasi-identifiers in $\mathcal{Q}$ to generalize, it attempts to maximize information gain for some target attribute.

6.2 Experimental methodology

To enable direct comparison with previous microdata sanitization work [21, 22], we used the *same* data for our experiments: the 45,222-record Adult database from the UCI Machine Learning Repository [1], described in table 1. Our classifier learning used Weka with the default settings for C4.5 (J48), Random Forests, and Naive Bayes. For all classification experiments, we used 10-fold cross-validation.

Choosing the quasi-identifier. In a real database, the set of quasi-identifier attributes $\mathcal{Q}$ is domain-specific, and includes the attributes to which the adversary is most likely to have access via an external database (*e.g.*, demographic information). For our experiments, we examined several different sets of attributes for $\mathcal{Q}$. All were picked to maximize the likelihood that sanitization will produce a useful table.

It is common in the literature to choose large quasi-identifiers, sometimes consisting of all non-sensitive attributes. A larger quasi-identifier, however, gives more prior information to the adversary and requires heavier generalization and suppression during sanitization. Large quasi-identifiers thus underestimate utility of the dataset and increase the risk of a privacy breach. Our most important criterion for choosing $\mathcal{Q}$ was to keep it *small*, to make the adversary's task as hard as possible.

Furthermore, if a legitimate user (whom we will call "researcher") is to get more utility out of the sanitized database than the adversary, his task(s) must be different from the adversary's. If the sensitive attribute is also the researcher's target attribute and all other attributes are quasi-identifiers, then both the researcher and the attacker are trying to use $\mathcal{Q}$ to predict $\mathcal{S}$! This is why in our measurements of utility, we consider utility of classification on "neutral" attributes which are neither quasi-identifiers, nor sensitive.

Attribute	Intact	Suppressed
Workclass	74.8618%	74.6672%
Education	41.6899%	41.1658%
Marital Status	69.3623%	58.5777%
Occupation	32.2387%	30.0363%
Country	91.7960%	91.6147%
Salary	82.7916%	82.4311%

Table 2: The effect of including age, sex, and race on decision tree learning accuracy.

Choosing the workload and the sensitive attribute. We must also choose a workload for the legitimate researcher. As discussed in Section 5, classification is a good workload because quality of classification depends on the correlations between attributes in the database, and the entire purpose of "truthfully" publishing quasi-identifiers and sensitive attributes *together* is to preserve these cross-attribute correlations.

We will look at classification of both sensitive and neutral attributes. It is important to choose a workload (target) attribute v for which the presence of the quasi-identifier attributes $\mathcal{Q}$ in the sanitized table actually matters. If v can be learned equally well with or without $\mathcal{Q}$, then the data publisher can simply suppress all quasi-identifiers.

Table 2 shows the difference in decision tree learning accuracy depending on whether or not the quasi-identifier (age, sex, race) is included. Only marital status shows a significant drop when the quasi-identifier is entirely suppressed, thus we choose it as the workload attribute for the "Occupation" dataset. Even though salary is intuitively the sensitive attribute in the "Adult" dataset, when the workload w is "learning salary," then $\mathcal{U}^{(w)}_{\mathsf{max}} \approx \mathcal{U}^{(w)}_{\mathsf{base}}$. Since we are interested in measuring utility of non-trivial sanitization (*i.e.*, how much utility it provides over the table in which all quasi-identifiers have been suppressed), we are only interested in scenarios where $\mathcal{U}^{(w)}_{\mathsf{san}} > \mathcal{U}^{(w)}_{\mathsf{base}}$; otherwise, complete suppression of $\mathcal{Q}$ provides better privacy and same utility as any non-trivial sanitization. When $\mathcal{U}^{(w)}_{\mathsf{max}} \approx \mathcal{U}^{(w)}_{\mathsf{base}}$, it cannot be the case that $\mathcal{U}^{(w)}_{\mathsf{san}} > \mathcal{U}^{(w)}_{\mathsf{base}}$, thus we do *not* choose salary as the sensitive attribute, and use marital status instead.

Datasets used. In the "Marital" dataset, $\mathcal{Q}$=(age, occupation, education), $\mathcal{S}$=marital status, the workload attribute is salary. In the "Occupation" dataset, $\mathcal{Q}$=(age, sex, race), $\mathcal{S}$=occupation, the workload attribute is marital status.

6.3 Experimental results

Learning the sensitive attribute $\mathcal{S}$. The researcher may wish to build a classifier for the sensitive attribute $\mathcal{S}$ using both the quasi-identifiers and the neutral attributes as predictors. Of course, if sanitization has been correctly performed, it is impossible to build a good classifier for $\mathcal{S}$ based *only* on $\mathcal{Q}$, because good sanitization must destroy any correlation between $\mathcal{S}$ and $\mathcal{Q}$ (otherwise, the adversary will eas-

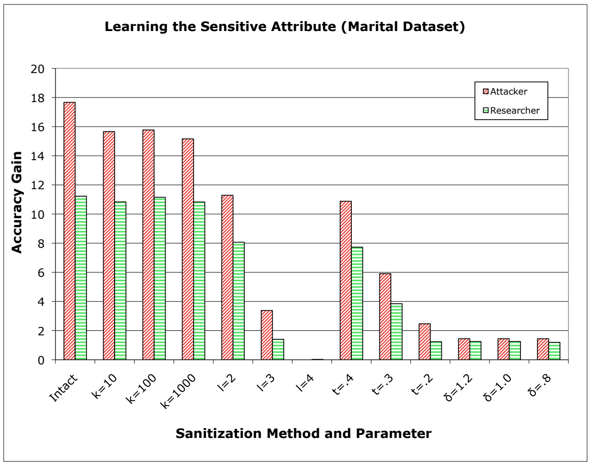

Figure 1: Gain in classification accuracy for the sensitive attribute (marital) in the "Marital" dataset. With trivial sanitization, accuracy is 46.56% for the adversary and 58.30% for the researcher.

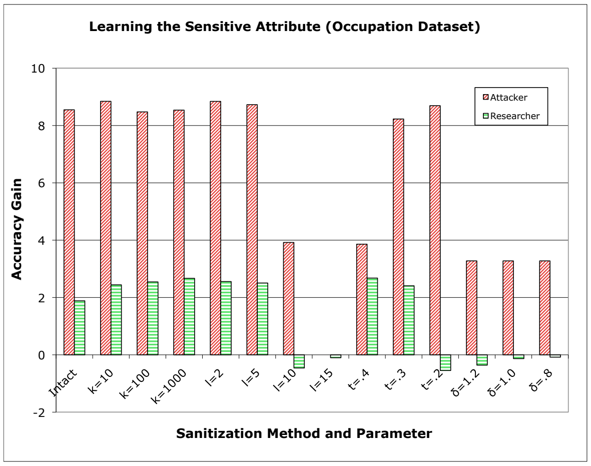

Figure 2: Gain in classification accuracy for the sensitive attribute (occupation) in the "Occupation" dataset. With trivial sanitization, accuracy is 13.31% for the adversary and 30.18% for the researcher.

ily learn the sensitive attributes associated with any quasi-identifier). Our results demonstrate that the researcher can build a classifier for $\mathcal{S}$ without using the attributes in $\mathcal{Q}$ just as well as when using sanitized versions of $\mathcal{Q}$.

Figure 1 shows the loss of privacy, measured as the gain in the accuracy of adversarial classification $\mathcal{A}_{\mathsf{acc}}$ for different sanitizations of the "Marital" dataset, and compares it with the gain in workload utility $\mathcal{U}^{(w)}_{\mathsf{san}} - \mathcal{U}^{(w)}_{\mathsf{base}}$ where the workload w is building a decision tree classifier for the "marital status" attribute. As explained in Section 4.4, accuracy of adversarial classification *underestimates* the actual amount of sensitive attribute disclosure. Figure 1 shows that releasing a sanitized table instead of simply suppressing all $\mathcal{Q}$ helps the adversary associate sensitive attributes with individuals much more than it helps the researcher to build legitimate classifiers. Figure 2 shows the same result for the "Occupation" dataset, where the workload w is building a decision tree classifier for the "occupation" attribute.

Learning a non-sensitive workload attribute. Perhaps it is not surprising that sanitization makes it difficult to build an accurate classifier for the sensitive attribute. We now consider the case when the researcher wishes to build a classifier for a *non-sensitive* attribute v.

If both $\mathcal{Q}$ and $\mathcal{S}$ are correlated with v, then a classifier based on both $\mathcal{Q}$ and $\mathcal{S}$ may have higher accuracy than one that considers only $\mathcal{Q}$, only $\mathcal{S}$, or neither. Our results show that sanitization which removes the correlation between $\mathcal{S}$ and $\mathcal{Q}$ also destroys the correlation between $\mathcal{S}$ and v.

In these experiments, we compute $\mathcal{U}^{(w)}_{\mathsf{base}}$ by running different machine learning algorithms on both trivially sanitized versions of the database; $\mathcal{U}^{(w)}_{\mathsf{base}}$ is the accuracy of the best classifier. We then compute $\mathcal{U}^{(w)}_{\mathsf{base}} - \mathcal{U}^{(w)}_{\mathsf{san}}$ for different sanitizations and different machine learning algorithms, and compare this to the increase in adversarial accuracy $\mathcal{A}_{\mathsf{acc}}$.

Figure 3 compares gains in adversary's and researcher's respective accuracies for the "Marital" dataset (workload w is learning the "salary" attribute). The classification accuracies with the sensitive attribute removed were 80.73% for J48, 77.12% for Random Forests, and 79.45% for Naive Bayes. Thus, the baseline for utility was set to 80.73%.

Figure 4 compares gains in adversary's and researcher's respective accuracies for the "Occupation" dataset (workload w is learning the "marital status" attribute). Here we see that $\mathcal{U}^{(w)}_{\mathsf{max}} = 69.52\%$ and $\mathcal{U}^{(w)}_{\mathsf{base}} = 69.30\%$ where the baseline comes from J48 learning with the sensitive attribute removed. With such a small gap between $\mathcal{U}^{(w)}_{\mathsf{max}}$ and $\mathcal{U}^{(w)}_{\mathsf{base}}$, it is not surprising that classification accuracies for sanitized datasets are below those of trivial sanitizations, where the sensitive attribute was simply removed.

Privacy of the sanitized database. Table 3 shows the $\mathcal{A}_{\mathsf{acc}}$ and $\mathcal{A}_{\mathsf{know}}$ scores for different sanitizations of the "Occupation" dataset (the accuracies are low even for the intact database because the quasi-identifiers do not identify a unique individual). Even for large k, k-anonymity barely changes the value of $\mathcal{A}_{\mathsf{know}}$ compared to the intact database. In other words, **k-anonymity provides no privacy improvement whatsoever** on this dataset. Furthermore, **ℓ-diversity is no better than trivial sanitization** because it requires complete suppression of the quasi-identifiers to substantially limit the gain in adversary's knowledge.

7. ACHIEVING PRIVACY AND UTILITY

Our experimental results in Section 6 indicate that, empirically, it is difficult to find a database table on which sanitization permits both privacy and utility. Any incremental utility gained by non-trivial sanitization (as opposed to simply removing quasi-identifiers or sensitive attributes) is more than offset by a decrease in privacy, measured as the adversarial sensitive attribute disclosure. It is possible, however, to construct an *artificial* database, for which sanitization provides both complete utility and complete privacy, even for the strongest definition of privacy (semantic privacy).

Consider table T, in which each tuple t has five attributes

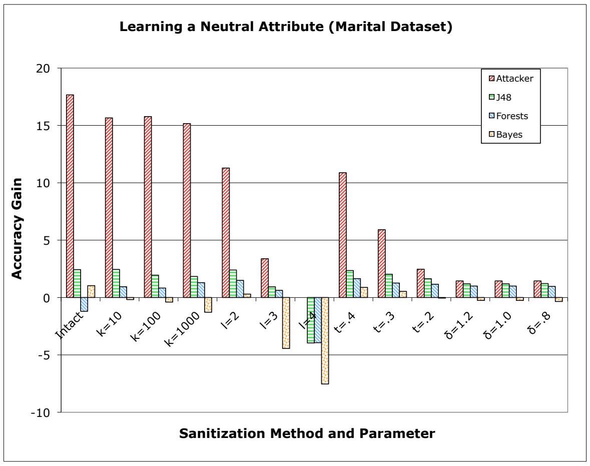

Figure 3: Gain in the adversary's ability to learn the sensitive attribute (marital) and the researcher's ability to learn the workload attribute (salary) for the "Marital" dataset. With the trivial sanitization, accuracy is 46.56% for the adversary, and 80.73% for the researcher.

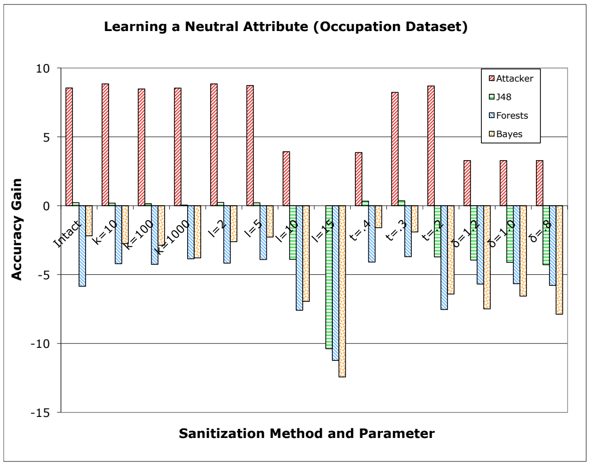

Figure 4: Gain in the adversary's ability to learn the sensitive attribute (occupation) and the researcher's ability to learn the workload attribute (marital) for the "Occupation" dataset. With the trivial sanitization, accuracy is 13.31% for the adversary, and 69.30% for the researcher.

a_1, a_2, a_3, a_4, a_5. Their values are defined by three coin flips r_1, r_2, r_3, which are generated independently at random for each tuple. The attributes are as follows:

$$a_1 = r_1 \quad a_2 = r_2 \quad a_3 = (r_2, r_3) \quad a_4 = r_1 \oplus r_3 \quad a_5 = r_1 \oplus r_2$$

Now consider the case where $\mathcal{Q} = \{a_1, a_2\}$, $\mathcal{S} = a_3$. This database is a candidate for sanitization, since $\mathcal{Q}$ provides a lot of information about $\mathcal{S}$ (half of the sensitive attribute can be predicted perfectly from the quasi-identifier). If we sanitize by suppressing a_2, then we are left with a database T' which is perfectly private, since a_1 reveals nothing about a_3. But this database also has perfect utility, since a researcher can learn a_4 exactly from a_1 and a_3, and he can learn a_5 exactly from a_1 and a_3, and he can learn a_3 exactly from a_1, a_4, and a_5. Furthermore, if $\mathcal{Q}$ were completely suppressed, the researcher could learn nothing about a_4 or a_5 and he could only learn half the information about a_3 ($r_2 \oplus r_3$). If $\mathcal{S}$ were omitted, the researcher could learn nothing about a_4 and only half of the information about a_5.

Sanitization	$\mathcal{A}_{\mathsf{acc}}$	$\mathcal{A}_{\mathsf{know}}$
Intact	0.1034	0.2492
k=10	0.0957	0.2331
k=100	0.0909	0.2236
k=1000	0.0885	0.2131
l=2	0.0966	0.2353
l=5	0.0940	0.2316
l=10	0.0400	0.1217
l=15	0	0
t=.4	0.0924	0.2264
t=.3	0.0861	0.2131
t=.2	0.0396	0.1213
δ=1.2	0.0328	0.0944
δ=1.0	0.0327	0.0937
δ=.8	0.0327	0.0915
Suppressed	0	0

Table 3: $\mathcal{A}_{\mathsf{acc}}$ and $\mathcal{A}_{\mathsf{know}}$ scores for different sanitizations of the "Occupation" dataset.

This artificial dataset is very unusual, and it is unclear whether any real datasets exhibit similar properties. For instance, sensitive attributes $\mathcal{S}$ can be split into two parts, one of which is 100% correlated with the quasi-identifiers $\mathcal{Q}$ and the other is completely independent of $\mathcal{Q}$. Sanitization can thus suppress the dependent part of $\mathcal{Q}$ entirely, while leaving the independent part intact. Furthermore, a_4 and a_5 are both completely determined by the joint distribution of $\mathcal{S}$ and $\mathcal{Q}$, but independent of either one taken alone. It is unclear how often attributes which are pairwise independent but jointly dependent arise in real data.

8. CONCLUSIONS

Microdata privacy can be understood as prevention of membership disclosure (the adversary should not learn whether a particular individual is included in the database) or sensitive attribute disclosure (the sanitized database should not reveal very much information about any individual's sensitive attributes). It is known that generalization and suppression cannot prevent membership disclosure [12, 26]. For sensitive attribute disclosure, perfect privacy can be achieved—against a very weak adversary who knows just the quasi-identifiers—by simply removing the sensitive attributes or the quasi-identifiers from the published data. Of course, these trivial sanitizations also destroy any utility that depended on the removed attributes.

Algorithms such as k-anonymity and ℓ-diversity leave all sensitive attributes intact and apply generalization and suppression to the quasi-identifiers. The goal is to keep the data "truthful" and thus provide good utility for data-mining applications, while achieving less than perfect privacy. We argue that utility is best measured by the success of data-mining algorithms such as decision tree learning which take advantage of relationships between attributes. Algorithms that need only aggregate statistical information can be executed on perturbed or randomized data, with much stronger privacy guarantees against stronger adversaries than achieved by k-anonymity, ℓ-diversity, and so on.

Our experiments, carried out on the same UCI data as was used to validate existing microdata sanitization algorithms, show that the privacy vs. utility tradeoff for these algorithms is very poor. Depending on the sanitization parameter, sanitized datasets either provide no additional utility vs. trivial sanitization, or the adversary's ability to compute the sensitive attributes of any individual increases much more than the accuracy of legitimate machine-learning workloads.

An important question for future research is whether there exists any real-world dataset on which quasi-identifier generalization supports meaningfully better data-mining accuracy than trivial sanitization without severely compromising privacy via sensitive attribute disclosure.

Another important question is how to design microdata sanitization algorithms that provide both privacy and utility. Sensitive attribute disclosure results, in part, from the fact that each individual t can only belong to a unique quasi-identifier equivalence class $\langle t \rangle$ in the sanitized table T'. This is a consequence of the requirement that the generalization hierarchy be totally ordered [10]. This requirement helps the adversary, but does not improve utility. If we consider $G(t)$, the set of records in T' whose quasi-identifier values are generalizations of $t[\mathcal{Q}]$, there is no privacy reason why each record of $G(t)$ must have the same quasi-identifier values. It is possible that a generalization strategy that uses, *e.g.*, DAGs instead of totally ordered hierarchies may provide better privacy than the existing algorithms.

9. REFERENCES

[1] D. N. A. Asuncion. UCI machine learning repository, 2007.

[2] N. Adam and J. Worthmann. Security-control methods for statistical databases: A comparative study. *ACM Computing Surveys*, 21(4), 1989.

[3] C. Aggarwal. On k-anonymity and the curse of dimensionality. In *VLDB*, 2005.

[4] R. Agrawal and R. Srikant. Privacy-preserving data mining. In *SIGMOD*, 2000.

[5] R. Bayardo and R. Agrawal. Data privacy through optimal k-anonymization. In *ICDE*, 2005.

[6] A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical privacy: the SuLQ framework. In *PODS*, 2005.

[7] J.-W. Byun, Y. Sohn, E. Bertino, and N. Li. Secure anonymization for incremental datasets. In *SDM*, 2006.

[8] S. Chawla, C. Dwork, F. McSherry, A. Smith, and H. Wee. Towards privacy in public databases. In *TCC*, 2005.

[9] B.-C. Chen, K. LeFevre, and R. Ramakrishnan. Privacy skyline: privacy with multidimensional adversarial knowledge. In *VLDB*, 2007.

[10] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati. k-anonymity. *Secure Data Management in Decentralized Systems*, 2007.

[11] I. Dinur and K. Nissim. Revealing information while preserving privacy. In *PODS*, 2003.

[12] C. Dwork. Differential privacy. In *ICALP*, 2006.

[13] A. Evfimievski, J. Gehrke, and R. Srikant. Limiting privacy breaches in privacy-preserving data mining. In *PODS*, 2003.

[14] B. Fung, K. Wang, and P. Yu. Top-down specialization for information and privacy preservation. In *ICDE*, 2005.

[15] V. Iyengar. Transforming data to satisfy privacy constraints. In *KDD*, 2002.

[16] D. Kifer and J. Gehrke. Injecting utility into anonymized datasets. In *SIGMOD*, 2006.

[17] D. Lambert. Measures of disclosure risk and harm. *J. Official Stat.*, 9, 1993.

[18] K. LeFevre, D. DeWitt, and R. Ramakrishnan. Incognito: Efficient full-domain k-anonymity. In *SIGMOD*, 2005.

[19] K. LeFevre, D. DeWitt, and R. Ramakrishnan. Mondrian multidimensional k-anonymity. In *ICDE*, 2006.

[20] K. LeFevre, D. DeWitt, and R. Ramakrishnan. Workload-aware anonymization. In *KDD*, 2006.

[21] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and ℓ-diversity. In *ICDE*, 2007.

[22] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. ℓ-diversity: Privacy beyond k-anonymity. In *ICDE*, 2006.

[23] D. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke, and J. Halpern. Worst-case background knowledge for privacy-preserving data publishing. In *ICDE*, 2007.

[24] G. Miklau and D. Suciu. A formal analysis of information disclosure in data exchange. In *SIGMOD*, 2004.

[25] M. Nergiz and C. Clifton. Thoughts on k-anonymization. In *PDM*, 2006.

[26] M. Nergiz, M. Atzori, and C. Clifton. Hiding the presence of individuals from shared database. In *SIGMOD*, 2007.

[27] M. Nergiz, C. Clifton, and A. Nergiz. Multirelational k-anonymity. In *ICDE*, 2007.

[28] A. Øhrn and L. Ohno-Machado. Using boolean reasoning to anonymize databases. *Artif. Intell. in Medicine*, 15, 1999.

[29] H. Park and K. Shim. Approximate algorithms for k-anonymity. In *SIGMOD*, 2007.

[30] V. Rastogi, D. Suciu, and S. Hong. The boundary between privacy and utility in data publishing. In *VLDB*, 2007.

[31] P. Samarati. Protecting respondents' identities in microdata release. *IEEE Trans. on Knowledge and Data Engineering*, 13(6), 2001.

[32] L. Sweeney. Weaving technology and policy together to maintain confidentiality. *J. of Law, Medicine and Ethics*, 25(2–3):98–110, 1997.

[33] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):571–588, 2002.

[34] L. Sweeney. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, 2002.

[35] J. Traub, Y. Yemini, and H. Wozniakowski. The statistical security of a statistical database. *ACM Transactions on Database Systems*, 9(4), 1984.

[36] T. Truta and B. Vinay. Privacy protection: p-sensitive k-anonymity property. In *PDM*, 2006.

[37] K. Wang and B. Fung. Anonymizing sequential releases. In *KDD*, 2006.

[38] K. Wang, B. Fung, and P. Yu. Template-based privacy preservation in classification problems. In *ICDM*, 2005.

[39] R. Wong, J. li, A. Fu, and K. Wang. (α,k)-anonymity: An enhanced k-anonymity model for privacy-preserving data publishing. In *KDD*, 2006.

[40] X. Xiao and T. Tao. m-invariance: Towards privacy preserving re-publication of dynamic datasets. In *SIGMOD*, 2007.

[41] X. Xiao and Y. Tao. Anatomy: Simple and effective privacy preservation. In *VLDB*, 2006.

[42] X. Xiao and Y. Tao. Personalized privacy protection. In *SIGMOD*, 2006.

[43] L. Zhang, S. Jajodia, and A. Brodsky. Information disclosure under realistic assumptions: Privacy versus optimality. In *CCS*, 2007.