Data Publishing against Realistic Adversaries

Ashwin Machanavajjhala
Yahoo! Research
Santa Clara, CA
mvnak@yahoo-inc.com

Johannes Gehrke
Cornell University
Ithaca, NY
johannes@cs.cornell.edu

Michaela Götz
Cornell University
Ithaca, NY
goetz@cs.cornell.edu

ABSTRACT

Privacy in data publishing has received much attention recently. The key to defining privacy is to model knowledge of the attacker – if the attacker is assumed to know too little, the published data can be easily attacked, if the attacker is assumed to know too much, the published data has little utility. Previous work considered either quite ignorant adversaries or nearly omniscient adversaries.

In this paper, we introduce a new class of adversaries that we call *realistic adversaries* who live in the unexplored space in between. Realistic adversaries have knowledge from external sources with an associated stubbornness indicating the strength of their knowledge. We then introduce a novel privacy framework called epsilon-privacy that allows us to guard against realistic adversaries. We also show that prior privacy definitions are instantiations of our framework. In a thorough experimental study with real census data we show that e-privacy allows us to publish data with high utility while defending against strong adversaries.

1. INTRODUCTION

Data collection agencies, like the U.S. Census Bureau, the World Bank, and hospitals, want to publish structured data about individuals (also called *microdata*) to support research on this data. However microdata contains much information that is sensitive (e.g., information about salaries or diseases). Privacy-preserving data publishing (PPDP) aims to publish microdata such that (*i*) aggregate information about the population is preserved, while (*ii*) guaranteeing privacy of individuals by ensuring that their sensitive information is not disclosed. There has been research on the problem of formally defining privacy in data publishing for more than half a century. The key to formally defining privacy is to correctly model how much sensitive information an *adversary* can deduce about an individual in the published data. This heavily depends both on the published data as well as on any knowledge the adversary possesses about the world. Let us illustrate this through a simple example.

VLDB '09, August 24-28, 2009, Lyon, France

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Disease
1	130**	< 30	*	Heart Disease
2	130**	< 30	*	Heart Disease
3	130**	< 30	*	Viral Infection
4	130**	< 30	*	Viral Infection
5	1485*	≥ 40	*	Cancer
6	1485*	≥ 40	*	Heart Disease
7	1485*	≥ 40	*	Viral Infection
8	1485*	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Table 1: Inpatient Microdata

EXAMPLE 1. *Table 1 shows medical records published by Gotham City Hospital. Each tuple in this table corresponds to a unique patient in the hospital. Each patient considers her disease to be sensitive. The table also contains non-sensitive attributes, that have been either been suppressed or coarsened to ensure that no patient can be uniquely identified. Supression and coarsening are popular methods for PPDP that are both instances of* generalization*; Table 1 is called a* generalized table*. The hospital should ensure that an adversary cannot link any patient to his disease using Table 1.*

Suppose Rachel is an individual in the population. Given access to only this table, the adversary Alice, may not be able the deduce Rachel's disease. But if Alice knows that (*a*) *Rachel is one of the individuals whose medical record is published in Table 1, and that* (*b*) *Rachel is a 35 year old American living in zip code 13068, Alice can deduce that Rachel has cancer.*

Alice's knowledge can take many different forms: Facts like Rachel is in Table 1, and men don't get ovarian cancer, or more uncertain knowledge that the likelihood of an arbitrary individual to have AIDS is less than 10%. Alice may have gained her knowledge from phone books, publicly available Census statistics, the web, or even by eavesdropping on an individual, just to name a few.

Recently, there has been much work on adversary models and associated formal definitions of privacy for data publishing [16, 14, 9]. Each of these definitions make different assumptions about the adversary's knowledge. For example, ℓ-diversity assumes that Alice knows at most $\ell - 2$ facts of

the form "men do not have ovarian cancer," but that Alice otherwise believes that all diseases for Rachel are equally likely. t-closeness assumes that Alice knows the distribution of diseases in the table, and she assumes that Rachel's chances of having a disease follow the same odds. Differential privacy assumes that Alice knows the exact diseases of all patients except Rachel, and that Alice may have other arbitrary statistical information about Rachel's disease.

However, as our short discussion already shows, a close look at this prior work reveals an unfortunate dichotomy: Existing privacy definitions either make very *specific* assumptions about Alice's knowledge that may be easily violated in practice (leading to *weak* privacy guarantees), or they make very *minimal* assumptions that allow Alice to know unrealistic amounts of information (but leading to very *strong* privacy guarantees). For example, it is quite reasonable to assume that Alice's knowledge about Rachel's disease in Table 1 is neither a uniform distribution, nor matches its distribution in the published table, but rather comes from some other background information. ("It is flu season, and there will be a lot of elderly patients with flu symptoms in the hospital.") Neither ℓ-diversity nor t-closeness can capture such background knowledge. Differential privacy, on the other hand, does not exclude any reasonable knowledge that Rachel could have, but the assumption that Alice knows everything about all patients except Rachel is equally unrealistic. Because of this strong guarantee, we cannot publish much useful information [17]; for example, to achieve differential privacy (or one of its variants [17]), we cannot even use the powerful technique of generalization, since no generalized table satisfies differential privacy.

Thus while privacy has been defined both for weak adversaries and for very strong adversaries, defining privacy for adversaries that inhabit the "realistic" space in between is an important open problem. This is especially important from the point of view of a practioner who would like to publish useful data, but also would like to provide provable privacy guarantees; for example, for useful PPDP algorithms like generalization, either restrictive assumptions must be made about the adversary's knowledge, or no data whatsoever can be published if the data publisher is unsure about the adversary's knowledge and want to be on the safe side.

Contributions of This Paper. In this paper we make a first step towards this middle ground between weak and strong adversaries. We introduce *realistic* adversaries and an associated novel privacy framework called **ϵ-privacy**, that allow us to reason about privacy across the spectrum of weak to strong adversaries. Realistic adversaries form their knowledge using external datasets, and they revise their knowledge from the published data if they see enough contradicting information in the published data. The adversary's willingness to revise her knowledge is characterized by a parameter σ called the adversary's *stubbornness* that depends on the size of the external dataset used to form her knowledge.

We claim not only that our novel privacy definition ϵ-privacy spans the space in the middle, but we can actually prove that existing privacy definitions, including ℓ-diversity, t-closeness, differential privacy and perfect privacy [19] are instances of ϵ-privacy.[1] Because ϵ-privacy can be instantiated with differential privacy, it may be surprising that ϵ-privacy still permits generalization; we show how to publish generalized tables guaranteeing ϵ-privay against adversaries with finite stubbornness, even when we are unsure about the adversary's knowledge. In a thorough experimental evaluation on real census datasets, we show that generalized tables with significantly more utility can be published by considering our new realistic adversaries with finite stubbornness.

The remainder of this paper is organized as follows. Section 2 describes our new ϵ-privacy framework. In Section 3 we identify four important classes of adversary models and derive conditions under which a generalized table guarantees ϵ-privacy against each of these adversary classes. We also propose efficient algorithms for PPDP. We experimentally show the utility of ϵ-private generalizations in Section 4. Section 5 illustrates that ϵ-privacy spans both the space of adversaries that are neither too strong or too weak, but also includes existing work as extreme instances. We discuss related work and conclude in Sections 6 and 7, respectively.

2. ϵ-PRIVACY

In this section, we describe our novel privacy framework. Our framework has the following attractive properties:

- Our framework allows sensitive information about each individual to be specified using a flexible language.
- Privacy is measured as the difference in an adversary's belief about an individual when the individual is in the published data compared to when the individual's information is left out of the published data.
- Our framework formally models realistic adversaries who form their priors using existing data and thus allows both strong and weak adversaries and adversaries that inhabit the middle ground.

Consider a data publisher who wants to publish data collected from a set of individuals. We assume that this data T has a relational structure with categorical attributes (A_1, ...,A_a) with a finite domain. The attributes are partitioned into a set of *non-sensitive* attributes whose disclosure is not our concern, and a set of *sensitive* attributes S whose privacy we should guarantee. We denote the non-sensitive and the sensitive attributes using two multi-dimensional attributes N and S, respectively. We assume that every tuple $t_u \in T$ contains information about one unique individual u. The data publisher uses a procedure R that takes T as input and outputs another table T_{pub} which is published. For historical reasons, we will (incorrectly) refer to R also as the *anonymization* procedure and T_{pub} as the *anonymized* table.

2.1 Sensitive Information

Let us formally describe an individual's sensitive information. An individual, we will call her Rachel, may want to protect many kinds of information. First Rachel may not want the adversary Alice to learn that she has some specific disease like cancer or the flu; such disclosures are called *positive disclosures*. Next, positive disclosures can occur on groups of diseases; for instance, Rachel may not want to disclose the fact that she has some stomach disease (such

[1]These privacy definitions guard against *infinitely stubborn* adversaries who (unrealistically) would need infinite amounts of external data to form their priors. For this reason, we call our new adversaries (maybe somewhat presumptuously) *realistic* adversaries.

as ulcer or dyspepsia). Finally, Rachel may not even want Alice to learn that she does *not* have cancer; this is called a *negative disclosure*. Note that a negative disclosure for cancer is the same as a positive disclosure that Rachel has one of the diseases in the set $Disease - \{cancer\}$.

We model sensitive information using positive disclosures on a set of *sensitive predicates*. The data publisher or an individual can specify a set of sensitive predicates Φ of the form *"Rachel[Disease] = cancer"*, or *"Rachel[Disease] ∈ {ulcer,dyspepsia}"*. Let $dom(S)$ denote the domain of the sensitive attribute. Each individual is associated with a set of sensitive predicates $\Phi(u)$. Each predicate $\phi(u) \in \Phi(u)$ takes the form $t_u.S \in S_\phi$, $S_\phi \subseteq dom(S)$. Informally, inferring that $\phi(u) = true$ from the published data for some $\phi(u) \in \Phi(u)$ breaches u's privacy.

EXAMPLE 2. *Let us continue our example from the introduction. Assume that the domain of the Disease attribute is {Flu, Cancer, Ulcer, Dyspepsia}. If Rachel wants to protect against positive disclosures for flu, cancer and any stomach related disease, she has three sensitive predicates, namely:*

$$\begin{aligned} \Phi &= \{\phi^1, \phi^2, \phi^3\} \\ \phi^1(Rachel) &: t_{Rachel}[S] \in \{Flu\} \\ \phi^2(Rachel) &: t_{Rachel}[S] \in \{Cancer\} \\ \phi^3(Rachel) &: t_{Rachel}[S] \in \{Ulcer, Dyspepsia\} \end{aligned}$$

Rachel can protect against any kind of disclosures related to her sensitive attribute if Φ contains one sensitive predicate ϕ for every subset $S_\phi \subset dom(S)$.

2.2 Disclosure

Our measure of disclosure about a sensitive predicate attempts to capture the privacy risk faced by an individual when allowing the data publisher to publish her information. Suppose a patient Rachel is deciding whether or not to permit Gotham City Hospital to include her tuple in Table 3. Rachel can achieve the most privacy by not permitting the hospital to include her tuple in the published data. However, since she cannot influence the privacy preferences of other individuals, despite disallowing the release of her information, the adversary Alice may infer some properties of Rachel's sensitive attribute based on her prior knowledge and based on the other individuals in the table who look like Rachel. Therefore, a data publisher should ensure that the individual does not regret having given permission to release her information; i.e., Alice's belief about the true values of Rachel's sensitive predicates when her tuple is included in T_{pub} should not be much higher when Rachel's tuple is not in T_{pub}.[2]

We can now describe our adversary model and explain how adversaries form their beliefs.

2.3 Adversaries And Their Knowledge

Recent research has established that an adversary's belief about an individual's sensitive information is determined by the published information and her knowledge [6, 9, 16, 18], also called her *prior* in statistical terms. Broadly, there are two kinds of prior knowledge an adversary possesses – (a) knowledge about the population from sources other than T, and (b) knowledge about the individuals in T. We describe these in turn.

(a) Knowledge about the population from sources other than the table being published.

Such knowledge is usually modeled by assuming that the individuals in the table are drawn from a joint distribution P over N and S, and that the adversary knows this distribution. More precisely, when both N and S are categorical, P can be described as a vector of probabilities $\overrightarrow{p} = (\ldots, p_i, \ldots), i \in N \times S$ such that $\sum_{i \in N \times S} p_i = 1$. For instance, many papers in the privacy literature use the *random worlds* prior [2, 6, 16, 18, 23, 24], where an adversary is assumed to have no preference for any value of i. This can be modeled by a uniform distribution where for all $i \in N \times S$, $p_i = 1/|N \times S|$. However, such a model has two main problems:

1. *Where does the adversary learn her prior P?* Typically, adversaries form their priors based on statistics T_S that have been made public before the publication of T. Such adversaries do not know which $\overrightarrow{p}$ is the right distribution, instead they use statistics they have collected about the population to determine $\overrightarrow{p}$. Thus such adversaries may not have complete confidence in their prior. For instance, suppose a hospital conducted a study on a sample of s individuals of the U.S. population, and found out that 10% of the individuals have cancer ($s_i = s/10$). How does Alice generalize these statistics to the population? One way would be to set $\overrightarrow{p}$ such that $p_i = s_i/s = 0.1$, the fraction of individuals in the sample having cancer, and to assume that all the individuals are drawn independently from $\overrightarrow{p}$. But in reality, Alice may not be willing to believe that 10% of the U.S. population have cancer when s is small (e.g., only 10 persons). However, when s is large (e.g., 1 million persons), Alice would be quite confident, or *stubborn*, that 10% of the population have cancer.

2. *An adversary may change her prior.* Modeling the adversary Alice's prior correctly becomes even more crucial since we are interested in computing Alice's beliefs about the true values of Rachel's sensitive predicates when Rachel's tuple is not included in the published data. For instance, suppose Alice forms her prior based on a survey of only $s = 100$ women, where $s_i = 50$ women have cancer; Alice creates a prior $\overrightarrow{p}$ with $p_i = 0.5$. Let Alice assume that all individuals, including Rachel, are drawn *independently* from this prior distribution $\overrightarrow{p}$. Suppose she now sees Table 3, which does not contain Rachel and where only 2000 out of 20,000 women have cancer. If Alice strongly believes in her prior $\overrightarrow{p}$ and assumes that individuals are drawn independently from $\overrightarrow{p}$, then she will continue to believe that the probability of Rachel having cancer is 0.5. However, if Alice takes the published Table 3 into account, which has overwhelming evidence that p_i is close to 0.1 rather than to 0.5, then Alice is likely to change her prior accordingly.

The key to correctly modeling adversarial reasoning is to relax the assumptions that (a) the adversary knows of a single prior $\overrightarrow{p}$, and (b) all individuals are drawn independently from $\overrightarrow{p}$. In fact, by not committing to a single prior, individuals in the table are no longer independent of each other. To understand this better, suppose there are two populations of equal size – Ω_1 having only healthy individuals and Ω_2 having only sick individuals. Suppose a table T is either created only from Ω_1 or created only from Ω_2. If we do not know which population T is picked from, using the principle

[2]The same intuition powers differential privacy [9]. ϵ-Privacy differs from differential privacy in the adversary model and in the modeling of the sensitive information.

of indifference, our best guess for the probability that any individual in the table T is healthy is 0.5. However, if we know that one individual in the table T is healthy, then we can be sure that the rest of the individuals in the table are also healthy (if we assume individuals are independent of each other, the probability would be 0.5).

To formally be able to model such reasoning, we first introduce the notion of exchangeability.

DEFINITION 1 (EXCHANGEABILITY). *A sequence of random variables $X_1, X_2, \ldots$ is exchangeable if every finite permutation of these random variables has the same joint probability distribution.*

The set of individuals in the table T are exchangeable: if H means *healthy* and S means *sick*, the probability of seeing $HHSSH$ is the same as the probability of $SSHHH$. It is easy to see that independent random variables are indeed exchangeable. The real power of exchangeable random variables arises from deFinetti's Representation Theorem [8, 20]. Informally, deFinetti's theorem states that an exchangeable sequence of random variables is mathematically equivalent to (i) choosing a data-generating distribution θ at random, and (ii) creating the data by *independently* sampling from this chosen distribution θ. In the above example, each population represents one data-generating distribution $\theta \in (\Omega_1, \Omega_2)$. The table is then created by choosing individuals independently from θ. Note that the prior probability that an arbitrary individual t in T is healthy is

$$\begin{aligned} Pr[t = H] &= \sum_{i=1}^{2} Pr[t = H | T \subseteq \Omega_i] P[T \subseteq \Omega_i] \\ &= 1 \cdot Pr[T \subseteq \Omega_1] + 0 \cdot Pr[T \subseteq \Omega_2] \;=\; 0.5 \end{aligned}$$

On the other hand if we know that there is one individual in T who is healthy. Then $Pr[t = H]$ changes to,

$$\begin{aligned} &Pr[t = H | t_1 = H \in T] \\ &= \sum_{i=1}^{2} Pr[t = H | H \in T \wedge T \subseteq \Omega_i] Pr[T \subseteq \Omega_i | H \in T] \\ &= 1 \cdot Pr[T \subseteq \Omega_1 | t_1 = H \in T] + 0 \cdot Pr[T \subseteq \Omega_2 | H \in T] \\ &= 1 \text{ since a } T \text{ drawn from } \Omega_2 \text{ would not contain a } H. \end{aligned}$$

More generally, under the notion of exchangeability, the original table T (of size n) can be assumed to be generated in the following two steps. First one out of an infinite set of probability vectors $\overrightarrow{p}$ is drawn from some distribution D. Then n individuals are drawn i.i.d. from the probability vector $\overrightarrow{p}$. D encodes the adversary's prior information. An agnostic adversary with no information can be modeled using a D that makes all $\overrightarrow{p}$ equally likely. An adversary who knows, e.g., that out of 10^6 individuals $999,999$ have cancer, should be modeled using a D that assigns $\overrightarrow{p}^\star$ with $p^\star(cancer) = 0.999$ a much higher probability than $\overrightarrow{p}^\dagger$ with $p^\dagger(cancer) = 0.001$. Since all our attributes are categorical, we adopt the *Dirichlet distribution*[3] (from the statistics literature [4]) to model such a prior over $\overrightarrow{p}$.

DEFINITION 2 (DIRICHLET DISTRIBUTION). *Let $\overrightarrow{p} = (p_1, \ldots, p_k)$ denote a vector of probabilities ($\sum_i p_i = 1$). The* Dirichlet distribution *with parameters $\overrightarrow{\sigma} = (\sigma_1, \ldots, \sigma_k)$, which is denoted by $D(\overrightarrow{\sigma})$, is a probability density function defined over the space of probability vectors such that*

$$D(\overrightarrow{p}; \overrightarrow{\sigma}) \;=\; \frac{\Gamma(\sigma)}{\prod_i \Gamma(\sigma_i)} \prod_i p_i^{\sigma_i - 1} \qquad (1)$$

where $\sigma = \sum_i \sigma_i$, and Γ is the gamma function.[4] *σ is called the* prior sample size *and $\overrightarrow{\sigma}/\sigma = (\sigma_i/\sigma, \ldots, \sigma_k/\sigma)$ is called the* shape *of the Dirichlet distribution.*

An adversary may form a Dirichlet prior $D(\sigma_1, \ldots, \sigma_k)$ as follows. An adversary without any knowledge about the population can be modeled by a prior of $D(1, \ldots, 1)$; this makes all the probability vectors equally likely and thus models the complete lack of information. Upon observing a dataset with counts $(\sigma_1 - 1, \ldots, \sigma_k - 1)$ the adversary can update his prior to $D(\sigma_1, \ldots, \sigma_k)$. With this updated prior not all probability vectors are equally likely. The probability vector with the maximum likelihood is $\overrightarrow{p^\star}$ such that $p_i^\star = \sigma_i/\sigma$.

As we increase σ without changing the shape of the Dirichlet, $\overrightarrow{p^\star}$ becomes more and more likely. That is, the adversary becomes more and more stubborn (or certain) that $\overrightarrow{p^\star}$ is the correct prior distribution. Hence, we call σ the *stubbornness* of an adversary. In particular, in the extreme case when $\sigma \to \infty$, $\overrightarrow{p^\star}$ is the only possible probability distribution.

DEFINITION 3. *An ∞–stubborn adversary is one whose prior sample size is such that $\sigma \to \infty$.*

Existing privacy definitions consider an extreme adversary who is ∞-stubborn. Here, the maximum likelihood vector $\overrightarrow{p^\star}$ is the prior belief of the adversary.

Before we move on, we note that there may be scenarios when a data publisher does not know exactly what knowledge the adversary has. Rephrased in our case, a data publisher might not know (i) the shape of the adversarial prior, (ii) the stubbornness, or (iii) both. Therefore, to define privacy in such scenarios, we consider classes of adversaries, all of whom have the same stubbornness (in (i)), or the same shape (ii) (and all possible adversaries in (iii)). We explain this in detail in Section 2.5.

(b) Knowledge about the individuals whose information resides in the table.

We consider adversaries who have full information about a subset of tuples $B \subset T$ in the table (like in [22]). Complex kinds of knowledge (like negation statements [16], and implications [18, 6]) have been considered in the literature; we plan to incorporate them into ϵ-privacy in future work.

2.4 Our Privacy Definition

Putting all the above discussions together, we can now finally introduce our novel privacy definition. After T_{pub} is published, the adversary's belief in a sensitive predicate $\phi(u)$ about individual u is given by the following *posterior probability*

$$p^{in}(T_{pub}, u, \phi, B, \overrightarrow{\sigma}) = Pr[\phi(u) \mid T_{pub} = R(T) \wedge B; D(\overrightarrow{\sigma})] \qquad (2)$$

[3]For numeric attributes Gaussian, Poisson or Pareto are good choices.

[4]$\Gamma(t)$ is the gamma function defined as $\int_0^\infty x^{t-1} e^{-x} dx$.

The adversary's belief in $\phi(u)$ conditioned on the published data when individual u's information is removed from the publication is:

$$p^{out}(T_{pub}, u, \phi, B, \overrightarrow{\sigma}) = Pr[\phi(u) \mid T_{pub} = R(T - \{t_u\}) \wedge B; D(\overrightarrow{\sigma})] \quad (3)$$

The distance between the two probabilities p^{in} and p^{out} measures how much more an adversaries learns about an individual's sensitive predicate $\phi(u)$ if her information was included in T. If p^{in} is much greater than p^{out} (or if $1-p^{in}$ is much smaller than $1-p^{out}$), then we say that the adversary has learnt a lot of information towards deducing that $\phi(u) = true$; we call this a privacy breach. But if the belief of an adversary is roughly the same no matter whether or not her information is included in the dataset, she has no reason to hold her data back.

DEFINITION 4 (ϵ-PRIVACY). *Let $D(\overrightarrow{\sigma})$ be the adversary's prior about the population and B be a subset of individuals in T whose exact information the adversary knows. Anonymization algorithm R is said to violate ϵ-privacy if for some output T_{pub} generated by R, for some individual u appearing in T and some $\phi(u) \in \Phi(u)$,*

$$\frac{p^{in}(T_{pub}, u, \phi, B, \overrightarrow{\sigma})}{p^{out}(T_{pub}, u, \phi, B, \overrightarrow{\sigma})} > \epsilon \text{ or} \quad (4)$$

$$\frac{1 - p^{out}(T_{pub}, u, \phi, B, \overrightarrow{\sigma})}{1 - p^{in}(T_{pub}, u, \phi, B, \overrightarrow{\sigma})} > \epsilon. \quad (5)$$

Note that we consider the change in the adversary's belief by removing an individual from the data only to measure the disclosure risk of an individual. We do not use it as an actual anonymization technique where we consider each individual u in turn and drop t_u if and only if u's privacy is breached. We either publish the output of R on the whole table, when no individual sees a privacy breach, or do not publishing anything.

2.5 Adversary Classes

In order to effectively demonstrate the power of the ϵ-privacy framework, we next identify four interesting classes of adversaries and then in the next section apply ϵ-privacy to generalizations in these adversarial settings. In each of the classes, we consider adversaries who form a single prior on the distribution of the sensitive attribute, that does not depend on the value of the non-sensitive attributes.That is, the adversary's prior is captured by a Dirichlet $D(\overrightarrow{\sigma})$, where $\overrightarrow{\sigma}$ has one parameter $\sigma(s)$ for every $s \in S$.

1. *Class I:* A set of adversaries with stubbornness at most σ and prior shape $\overrightarrow{\sigma}/\sigma$.

2. *Class II:* A set of adversaries with stubbornness at most σ, but with an arbitrary prior $D(\overrightarrow{\sigma})$ such that $\sum_{s \in S} \sigma(s) = \sigma$.

3. *Class III:* A set of adversaries having the same prior shape of $\overrightarrow{\sigma}/\sigma$, but with arbitrarily large stubbornness (i.e., an ∞-stubborn adversary).

4. *Class IV:* The set of all possible adversaries (including ∞-stubborn adversaries with an arbitrary prior shape).

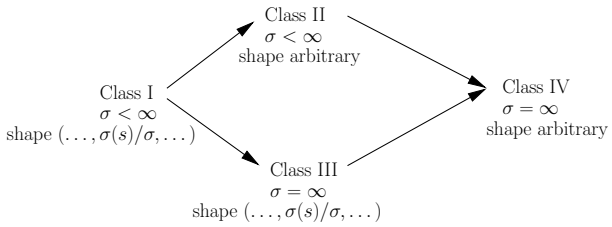


Figure 1: Adversary Classes: arrows point to the stronger adversary class.

We would like to note this is the first time class I and class II adversary are being considered. Moreover, since they assume adversaries who form their prior based on finite amounts of external data, we call them *realistic adversaries*. Figure 1 shows the relationships between the different adversary classes. Class I adversaries are weaker than both class II and class III adversaries, both of which are in turn weaker than the set of all adversaries (class IV). Before we go on to analyze the privacy of generalizations in these realistic settings, we compare and contrast how these adversaries reason about privacy in the following example.

EXAMPLE 3. *Continuing with our hospital example, suppose that Table 3 (T_{pub})is a generalized version of the complete inpatient medical records in Gotham City Hospital. At the time of publication, suppose the nearby Metropolis City Hospital has already published a dataset citing the distribution of flu and cancer; out of 29,998 individuals, 11,999 have flu and 17,999 have cancer.*

We consider the following example adversaries: a class I adversary who forms a prior of $D(12000, 18000)$ from the Metropolis dataset, a class II adversary with a stubbornness of 30,000 who formed his prior based on some other unknown dataset of size at most 29998, an ∞-stubborn class III adversary who believes that the distribution of flu versus cancer is $(.4, .6)$, and a class IV adversary whose prior is an arbitrary probability distribution. Assume $B = \emptyset$ for this example.

Let us now reason about the ϵ-privacy of an individual, say Rachel, whose tuple is in the table. We must calculate (i) $p^{in}(flu)$, the posterior probability that Rachel has flu when she is in T_{pub}, and (ii) $p^{out}(flu)$ when her tuple is excluded from the table for each of the four adversaries. We show in the appendix that in all four adversarial scenarios $p^{in}(flu) = 0.9$, which is the fraction of tuples in Rachel's anonymous group in T_{pub} who have cancer. That is, p^{in} only depends on the published data. However, adversaries from different classes reason about p^{out} differently.

A class I adversary uses both the published data and his prior knowledge and computes $p^{out} = \frac{18000+12000}{20000+30000} = .6$, where 12000 and 30000 are $\sigma(flu)$ and σ respectively. Note that even though the adversary's prior belief that Rachel has flu was .4, he revised it based on the evidence from T_{pub}.

A class II adversary's stubbornness is constant, but his $\sigma(flu)$ could be arbitrary: if the data he consulted has no flu patients, then $\sigma(flu)$ could be as low as 1, and if it had no cancer patients, $\sigma(flu)$ could be as high as $29,999$. That is, a class II adversary might compute p^{out} to be as low as $\frac{18000+1}{20000+30000} = .36002$.

A class III adversary has infinite stubbornness; hence, he

Symbol	Description
T_{pub}	Generalized table
Q	Non-sensitive attribute
S	Sensitive attribute
n	Total number of tuples in T_{pub}
$n(q,s)$	Number of $t \in T_{pub}$ s.t. $t[Q]=q, t[S]=s$
$n(q,s)$	$\sum_{s\in S'} n(q,s)$, for $S' \subseteq S$
$\sigma(s)$	Dirichlet parameter for adversarial prior for $s \in S$
σ	$\sum_{s\in S'} \sigma(s)$, for $S' \subseteq S$
$\overrightarrow{\sigma}$	$(\sigma(s_1), \ldots, \sigma(s_{\lvert S\rvert}))$
B	Adversary knows exact information about all $b \in B$ ($B \subseteq T$).
$b(q,s)$	Number of $b \in B$ s.t. $b[Q]=q, b[S]=s$
b	size of B

Table 2: Notation

Non-Sensitive		Sensitive	
Age	Gender	Disease	Count
< 40	M	Flu	200
< 40	M	Cancer	300
≥ 40	M	Flu	1800
≥ 40	M	Cancer	2700
≥ 40	F	Flu	18000
≥ 40	F	Cancer	2000

Table 3: A generalized microdata table showing the distinct tuples and their multiplicities.

does not revise his beliefs based on the data in T_{pub}. He computes $p^{out} = .4$.

A class IV adversary has both infinite stubbornness and an arbitrary prior. His p^{out} could be any value in $[0,1]$, hence, there is no finite ϵ for which Rachel's data is private against this adversary.

It is easy to see that Rachel is guaranteed 4-, 6.4-, 6- and no privacy against class I, class II, class III and class IV adversaries, respectively.

3. PRIVACY OF GENERALIZATION

In this section we apply ϵ-privacy to generalizations. Table 2 summarizes the notation used in this section. Let Q denote the (multi-dimensional) attribute representing the *generalized* non-sensitive information in T_{pub}, and S the sensitive attribute. Let n denote the number of tuples, and hence individuals, in T_{pub}. We denote by $n(q,s)$ the number of tuples $t \in T_{pub}$ such that $t[Q]=q$ and $t[S]=s$. We call the group of tuples $t \in T$ that share the same non-sensitive information q as a q *anonymous group*; hence the size of an anonymous group is usually termed its *anonymity*. An anonymous group is called *diverse* if the distribution of the sensitive attribute in the group is roughly uniform; it is $(c,2)$-*diverse* if the most frequent sensitive value in the group appears with at most a $c/(c+1)$ fraction of the tuples in the group.

For ease of explanation we assume that the set of sensitive predicates for every individual u is $\Phi(u) = \{\{s\} \mid s \in S\}$. That is, every individual only cares about positive disclosures about specific values in the sensitive attribute domain. Extensions to other kinds of sensitive information is straightforward. We also assume that the data publisher does not know the composition of B, but only its size $b = |B|$.

In order to help us reason about the privacy of all individuals efficiently, we next state the conditions under which publishing a generalized table T_{pub} guarantees ϵ-privacy of all individuals in T_{pub} against each of the four adversary classes. A reader interested in the technical details is referred to Appendix A.

THEOREM 1 (PRIVACY CHECK FOR CLASS I). *Table T_{pub} is ϵ–private against a class I adversary with prior $D(\overrightarrow{\sigma})$ and $|B| = b$, if for all anonymous groups q, for all sensitive values s, the following conditions hold*

$$R1 \quad (a)\ n(q) - b \geq \frac{\sigma + b}{\epsilon - 1},\ or$$

$$(b)\ \frac{n(q,s)}{n(q)-b} \leq \frac{\epsilon}{1-\delta(q)} \cdot \frac{\sigma(s)-1}{\sigma+b}$$

$$R2 \quad \frac{n(q,s)}{n(q)-b} \leq 1 - \frac{1}{\epsilon' + \delta(q)} + \frac{1}{\epsilon' + \delta(q)} \cdot \frac{\sigma(s)-1}{\sigma+b}$$

where, $\delta(q) = (\epsilon - 1) \cdot \dfrac{n(q)-b}{\sigma+b}$ *and* $\epsilon' = \epsilon \cdot \left(1 - \dfrac{1}{\sigma+b}\right)$.

Theorem 1 has the following consequences. A combination of *anonymity* and *closeness* is sufficient to guarantee ϵ-privacy against a class I adversary; *R1(a)* requires each anonymous group to be sufficiently large, and *R1(b)*, *R2* require that the sensitive attribute distribution in each anonymous group to be close to the prior shape. For instance, consider Table 3 and an adversary with a prior of $D(12000, 18000)$ and $b = 0$. Table 3 satisfies 2.5-privacy because,

- In the two groups with males, the distribution of the sensitive attribute is identical to the shape of the prior (thus satisfying *R1(b)*, *R2*).
- The group with females satisfies the anonymity requirement $(20,000 \geq 30,000/1.5)$ (*R1(a)*) and the second closeness requirement *R2*.

Note that even though the distribution of sensitive attribute in the females anonymous group squarely contradicts the prior shape, privacy is not breached. This is because we are dealing with an adversary who is willing to change his prior if there is sufficient evidence to the contrary. Note that the right hand sides of both *R1(b)* and *R2* increase as $\delta(q)$ increases. Thus, as anonymity increases (keeping the sensitive attribute distribution the same in each anonymous group) a generalized table guarantees more privacy. Also for the same ϵ, anonymity can be traded-off with closeness. This is the first privacy definition that shows such a connection between anonymity and privacy.

THEOREM 2 (PRIVACY CHECK FOR CLASS II). *Table T_{pub} is ϵ-private against a class II adversary with stubbornness σ and $|B| = b$, if for all anonymous groups q, for all sensitive values s the following conditions hold*

$$R1 \quad n(q) - b \geq \frac{\sigma + b}{\epsilon - 1},$$

$$R2 \quad \frac{n(q,s)}{n(q)-b} \leq 1 - \frac{1}{\epsilon' + \delta(q)}$$

where, $\delta(q) = (\epsilon - 1) \cdot \dfrac{n(q)-b}{\sigma+b}$ *and* $\epsilon' = \epsilon \cdot \left(1 - \dfrac{1}{\sigma+b}\right)$.

Here, a combination of *anonymity* (*R1*) and *diversity* (*R2*) is sufficient to guarantee ϵ-privacy. For instance, Table 3 satisfies 3-privacy against a class II adversary with stubbornness $\sigma = 1,000$ and $b = 0$, because (i) $\delta(< 40, M) = 1$ and the most frequent sensitive value in this group may appear in at most $\frac{3}{4}$ of the tuples; (ii) $\delta(\geq 40, M) = 9$ and the most frequent sensitive value in this group may appear in at most $\frac{11}{12}$ of the tuples; and (iii) $\delta(\geq 40, F) = 40$ and the most frequent sensitive value in this group may appear in at most $\frac{42}{43}$ of the tuples. When $\sigma = 30,000$, however, Table 3 only satisfies ϵ-privacy for $\epsilon \geq 61$ for males of age less than 40. The anonymous group is not large enough (500) to force a 30,000-stubborn adversary revise his prior, which may be very different from $(.4, .6)$.

We would like to note that this is the first privacy definition that allows a generalized table to be published with formal privacy guarantees even though the data publisher is unsure about the adversary's prior shape. Also note that for the same ϵ, as the anonymity of a group increases, the distribution of the sensitive attribute in that group is allowed to be more and more skewed. Conversely, in order for smaller groups to satisfy privacy, the distribution of the sensitive attribute must be close to uniform.

THEOREM 3 (PRIVACY CHECK FOR CLASS III). *Table T_{pub} is ϵ–private against a class III adversary with a prior shape of $(\ldots, \sigma(s)/\sigma, \ldots)$ and $|B| = b$, if for all anonymous groups q, for all sensitive values s the following conditions hold*

$$\textit{(R1)}\ \frac{n(q,s)}{n(q)-b} \leq \epsilon \cdot \frac{\sigma(s)}{\sigma}$$

$$\textit{(R2)}\ \frac{n(q,s)}{n(q)-b} \leq 1 - \frac{1}{\epsilon} \cdot \left(1 - \frac{\sigma(s)}{\sigma}\right)$$

Theorem 3 requires that the distribution in each anonymous group be close to the prior. Recall that a class III adversary (due to ∞-stubbornness) assumes that the individuals in the data are drawn independently from a single distribution $\overrightarrow{p^\star}$ such that $p^\star(s) = \sigma(s)/\sigma$. That is, no matter how much evidence is seen contradicting the prior $\overrightarrow{p^\star}$ the adversary will not update his beliefs. Hence, increasing anonymity has no effect on the privacy guaranteed against a class III adversary.

Finally, we consider the set of all possible adversaries. Since the stubbornness and prior shape are arbitrary, no privacy can be guaranteed by generalization.

THEOREM 4 (PRIVACY CHECK FOR CLASS IV). *A generalized table T_{pub} does not guarantee ϵ–private against class IV adversaries for any value of ϵ even when $b = 0$.*

3.1 Finding a Generalized Table

Algorithms for finding a private generalization (like Incognito [12], Mondrian [13], etc.) usually involve two parts - algorithm P that checks whether a generalization satisfies privacy, and, algorithm A that searches for a generalization that satisfies P and has the best utility. For all three classes of adversaries in this paper, we can show that algorithm P terminates in $O(N)$ time. Note that we only need to check privacy conditions for those sensitive values s that appear in the table (else privacy is automatically guaranteed).

Most algorithms A find a minimal generalization. A table T can be generalized in many different ways. For instance, Table 3 (T) can be generalized by suppressing either age (giving T_a) or gender (T_g), or both ($T_\top$). We can impose a partial ordering $\preceq$ on generalizations of T; i.e., if T_1 and T_2 are generalizations of T, then $T_1 \preceq T_2$ if and only every anonymous group in T_2 is a union of one or more anonymous groups in T_1. Again in our example, $T \preceq T_a, T_g \preceq T_\top$, but $T_a \not\preceq T_g$ or vice versa. Note that we lose information with every generalization ($T_\top$ has the least information). Hence, we would like to efficiently find a generalization that is as far away from $T_\top$ as possible that satisfies ϵ-privacy; this is called a minimal generalization.

Starting from an original table, we can find the set of all minimal generalizations efficiently using existing algorithms [12, 3, 13] if the check for ϵ-privacy satisfies the following *monotonicity* property.

DEFINITION 5 (MONOTONICITY). *Let f be a function that takes a table $T \in dom(T)$ and outputs either true or false. f is said to be monotonic on a partial ordering $\preceq$ on $dom(T)$, if*

$$T_1 \preceq T_2 \wedge f(T_1) = true \quad \Longrightarrow \quad f(T_2) = true$$

It can be easily shown that the check for ϵ-privacy against class I, II and III adversaries (Theorems 1, 2 and 3) satisfies the monotonicity property, thus allowing a data publisher to efficiently compute a minimal generalization.

3.2 Choosing Parameters

We conclude this section with a brief discussion on how a data publisher can choose parameters, namely ϵ, stubbornness σ, and the prior shape, to instantiate ϵ-privacy. Since the choice of parameters is application dependent, our discussion will be based on a real Census application called OnTheMap[5], which publishes anonymized commute patterns of workers in the US. The current OnTheMap (v3) algorithm [17] is based on differential privacy (and thus a class IV adversary). The parameter ϵ is set between 10 an 100. OnTheMap could greatly benefit from using realistic adversaries; here, the prior can be set based on the distribution of commute patterns from a previous version (or versions) of the Census data (for example, the Census Transportation Planning Package, CTPP[6]). The stubbornness can be set based on the number of individuals contributing to each demographic in CTPP. The stubbornness may also be set higher or lower taking into account recency of the previous versions. We would like to point out that the parameters do not have to be exact, but rather, only need to upper bound the strength of the adversaries considered. Also, if the Census is worried that either the prior or the stubbornness has changed since previous releases, a stronger adversary model (II, III or IV) can be used to account for uncertainty in one or both of these parameters.

4. EXPERIMENTS

In this section we experimentally evaluate the utility of generalized tables that satisfy ϵ-privacy in the new space of realistic adversaries, and compare it to the utility of generalized tables satisfying ℓ-diversity and t-closeness, using real census data. We do not consider differential privacy since no generalization guarantees such strong privacy.

[5] http://lehdmap3.did.census.gov/.

[6] www.fhwa.dot.gov/ctpp/.

	Attribute	Domain size	Generalizations type	Ht.
1	Age	73	ranges-5,10,20,40,⋆	6
2	Marital Status	6	Taxonomy tree	3
3	Race	9	Suppression	2
4	Gender	2	Suppression	2
8	Salary class	2	*Sensitive att.*	

Table 4: Selected Attributes of the ACS Database

In Section 4.1 we compare the utility of generalizations against class I, II and III adversaries. Within class I and class II, we show that we can generate tables with strictly more information as the stubbornness of the adversary decreases. For the first time, we show that generalization can be used to publish useful tables against strong adversaries with arbitrary priors.

In Section 4.2 we compare the utility of the generalizations satisfying $(c, 2)$-diversity and t-closeness to that of generalizations guarding against realistic class I and class II adversaries. We show that more useful information can be published when considering realistic adversaries. We also uncover an interesting fact that a class III adversary with a uniform prior provides equivalent utility as $(c, 2)$-diversity, suggesting that $(c, 2)$-diversity is an instantiation of ϵ-privacy. We explore this in more detail in Section 5.

Data: We use the American Community Survey (ACS) Dataset from the Minnesota Population Center [5] for our studies. The ACS Dataset from 2006 contains nearly 3 million tuples. We adopted the same domain generalizations as [12]. Table 4 provides a brief description of the data including the attributes we used, the number of distinct values for each attribute, the type of generalization that was used (for non-sensitive attributes), and the height of the generalization hierarchy for each attribute. We call the original table T. We treat the *Salary class* as sensitive $|S| = 2$.

Generalization Lattice: Recall that there are many generalizations of a table, and that we can define a partial order $\preceq$ on these generalized tables. This gives us a *generalization lattice*. Each table in the generalization lattice is label $[x_1, x_2, x_3, x_4]$ where x_i denotes the number of levels attribute i has been generalized. For example $[4, 1, 1, 0]$ means that *Age* was partitioned into ranges of 40, Marital Status was partitioned into classes {married, never married}, Race was suppressed, and the Gender was not generalized. The sum of the generalization heights determines the level of a table. The most general table $T_\top$ is labeled $[5, 2, 1, 1]$ (level 9) and contains the least amount of information amongst all the tables in the lattice. The partial order on the tables is: $T[x_1, x_2, x_3, x_4] \preceq T'[y_1, y_2, y_3, y_4]$ if, $x_i \leq y_i$.

Utility Metric: Many metrics have been proposed to study the utility of generalizations – average size of an anonymous group, KL-divergence to the original table [16], discernibility [3], and average error on a pre-specified query workload. The utility of a privacy definition P is the maximum utility of a generalization that satisfies P. In practice, however, utility depends on the application that uses the data, and we do not know which properties are of interest for this application.

In addition to studying the above metrics, we look at all the 72 possible generalizations of the ACS dataset and compare which of these guarantee privacy according to each of the privacy definitions we are interested in. We say that privacy definition P_1 guarantees *strictly more utility* than privacy definition P_2 if every generalized table that satisfies P_2 also satisfies P_1. Consequently, P_1 has more utility than P_2 according to all of the above metrics.

Generalization		Smallest value of ϵ for privacy against					
		$\mathcal{A}_d^{10^6}$	$\mathcal{A}_c^{10^6}$	$\mathcal{A}_-^{10^3}$	$\mathcal{A}_-^{10^6}$	$\mathcal{A}_d^{\infty}$	$\mathcal{A}_c^{\infty}$
9	[5, 2, 1, 1]	1.13	1.0	1.1	1.6	1.5	1
8	[4, 2, 1, 1]	105.8	70.8	2.9	211.6	118.9	79.5
	[5, 1, 1, 1]	1.57	1.25	1.1	2.7	2.38	1.59
	[5, 2, 1, 0]	1.32	1.08	1.1	2.1	1.81	1.21
	[5, 2, 0, 1]	1.69	1.54	1.3	249.4	1.69	1.54
7	[3, 2, 1, 1]	105.8	70.7	2.9	211.6	119.0	79.6
	[4, 2, 1, 0]	377	251.9	11.1	753.9	407.0	272.0
	[4, 1, 1, 1]	213.4	142.6	6.9	426.8	229.8	153.6
	[5, 0, 1, 1]	15.3	10.3	1.3	30.5	17.7	11.8
	[5, 1, 1, 0]	3.01	2.14	1.1	5.7	4.38	2.92
	[5, 1, 0, 1]	4.71	1.25	3.6	1157.1	4.93	3.29
	[5, 2, 0, 0]	2.45	1.69	1.7	526.8	2.45	1.69
6	[3, 1, 1, 1]	213.4	142.6	6.9	426.8	229.8	153.5
	[3, 2, 1, 0]	377	251.9	11.1	753.9	406.9	271.9
	[5, 0, 1, 0]	19.8	13.3	1.6	55.9	22.3	14.9
	[5, 1, 0, 0]	9.5	6.34	7.2	2625.7	9.72	6.49
	[5, 0, 0, 1]	21.7	14.5	21.8	14.6	21.8	14.6

Table 5: Minimum parameter ϵ such that each generalization guarantees privacy against each of the adversaries.

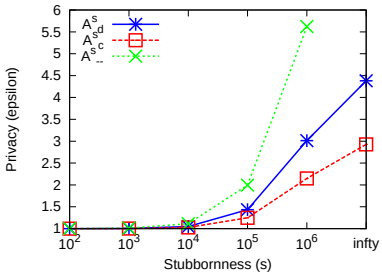
Privacy Definitions Compared: $(c, 2)$-Diversity and t-closeness are compared to the following adversaries:

- $\mathcal{A}_d^{\infty}(\epsilon)$: ϵ-privacy against a class III adversary with a uniform prior shape, i.e., p(salary = *high*) = p(salary = *low*) = 0.5.
- $\mathcal{A}_c^{\infty}(\epsilon)$: ϵ-privacy against a class III adversary whose prior shape is identical to the sensitive attribute distribution in the whole table ($\sigma(s)/\sigma = n(s)/n$); i.e., p(salary = *high*) = 0.334.
- $\mathcal{A}_d^{\sigma}(\epsilon)$: ϵ-privacy against a class I adversary with a uniform prior shape and stubbornness σ.
- $\mathcal{A}_c^{\sigma}(\epsilon)$: ϵ-privacy against a class I adversary whose prior shape is $\forall s, \sigma(s)/\sigma = n(s)/n$, and stubbornness σ.
- $\mathcal{A}_-^{\sigma}(\epsilon)$: ϵ-privacy against class II adversaries (arbitrary prior shape) with stubbornness σ.

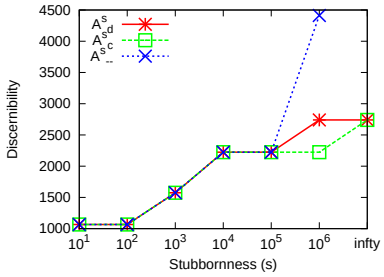
4.1 Generalization With Realistic Adversaries

We first compare the utility of generalization against realistic class I adversaries ($\mathcal{A}_d^{10^6}(\epsilon)$, $\mathcal{A}_c^{10^6}(\epsilon)$), realistic class II adversaries ($\mathcal{A}_-^{10^3}(\epsilon)$, $\mathcal{A}_-^{10^6}(\epsilon)$) and class III adversaries ($\mathcal{A}_d^{\infty}(\epsilon)$, $\mathcal{A}_c^{\infty}(\epsilon)$). For every generalization T_g, we calculate the smallest value of ϵ such that T_g satisfies ϵ-privacy against each of the 6 adversaries. Table 5 shows all the tables that satisfy one of these criterion with $\epsilon < 20$. For instance, the table [5,1,1,0] provides privacy against $\mathcal{A}_d^{10^6}(\epsilon)$ for all $\epsilon \geq 3.01$, against $\mathcal{A}_c^{10^6}(\epsilon)$ for all $\epsilon \geq 2.14$, against $\mathcal{A}_-^{10^3}(\epsilon)$ for all $\epsilon \geq 1.1$ and so on.

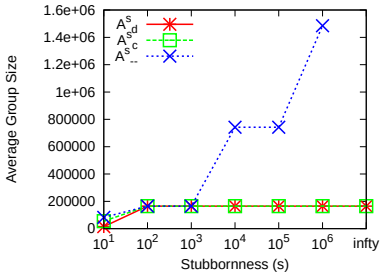
- *Class I strictly more useful than corresponding Class II or Class III:* The minimum ϵ required for any table is smaller for class I adversaries ($\mathcal{A}_d^{10^6}(\epsilon), \mathcal{A}_c^{10^6}(\epsilon)$) compared to either (i) their *corresponding* class III adversaries ($\mathcal{A}_d^{\infty}(\epsilon), \mathcal{A}_c^{\infty}(\epsilon)$), and (ii) class II adversaries ($\mathcal{A}_-^{10^6}(\epsilon)$) with greater stubbornness. Thus for any ϵ, class I adversaries produce *strictly*



(a) Minimum value of ϵ for $\mathcal{A}_d^\sigma(\epsilon)$, $\mathcal{A}_c^\sigma(\epsilon)$ and $\mathcal{A}_-^\sigma(\epsilon)$



(b) Minimum discernibility of tables satisfying $\mathcal{A}_d^\sigma(2.5)$, $\mathcal{A}_c^\sigma(2.5)$ and $\mathcal{A}_-^\sigma(2.5)$. Discernibility is in billions.



(c) Minimum average group size of tables satisfying $\mathcal{A}_d^\sigma(2.5)$, $\mathcal{A}_c^\sigma(2.5)$ and $\mathcal{A}_-^\sigma(2.5)$

Figure 2: Maximum utility vs. stubbornness σ for $\mathcal{A}_d^\sigma(\epsilon)$, $\mathcal{A}_c^\sigma(\epsilon)$, and $\mathcal{A}_-^\sigma(\epsilon)$.

more useful tables than class II or class III adversaries.

• *Utility despite adversaries with arbitrary prior shapes:* In the current state of the art, a data publisher unsure about the adversary's prior *cannot* publish a generalization with provable privacy guarantees. For the first time, our experiments show that we can publish useful generalization tables against strong realistic adversaries with an arbitrary prior shape; there are 4 tables that are private against $\mathcal{A}_-^{10^6}(20)$ adversaries, and 16 tables that can be published that are private against $\mathcal{A}_-^{10^3}(20)$ adversaries.

• *More utility by tolerating less stubborn adversaries:* The minimum values of ϵ increases consistently for all tables as stubbornness is increased from 10^6 to ∞ in the case of class I adversaries, and from 10^3 to 10^6 for class II adversaries. This is also seen in Figure 2(a), which plots the minimum value of ϵ such that the table [5,1,1,0] is private against realistic adversaries $\mathcal{A}_d^\sigma(\epsilon)$, $\mathcal{A}_c^\sigma(\epsilon)$, and $\mathcal{A}_-^\sigma(\epsilon)$ as stubbornness is increased (10^2 to ∞). As expected, we observe an increase in the minimum ϵ when we guard against increasingly stubborn adversaries. Hence, as the adversarial stubbornness decreases, for the same ϵ more tables guarantee privacy, thus yielding more utility.

Figures 2(a) and (b) show the same result using standard utility metrics discernibility and average group size. For $\epsilon = 2.5$, we computed the minimum value of discernibility and average group size achieved by some tables that satisfies $\mathcal{A}_d^\sigma(2.5)$, $\mathcal{A}_c^\sigma(2.5)$ and $\mathcal{A}_-^\sigma(2.5)$, respectively. Clearly, utility increases when less stubborn adversaries are considered.

• *Class II versus Class III adversaries:* Finally, we note that sometimes class II adversaries may allow more useful tables to be published than class III adversaries. For instance, the minimum value of ϵ for almost all tables is smaller in the case of class II adversary $\mathcal{A}_-^{10^3}(\epsilon)$ than in the case of class III adversaries $\mathcal{A}_d^\infty(\epsilon)$ and $\mathcal{A}_c^\infty(\epsilon)$. In Figure 2(b), tables satisfying $\mathcal{A}_-^\sigma(2.5)$ have the same utility as $\mathcal{A}_d^\sigma(2.5)$ and $\mathcal{A}_c^\sigma(2.5)$, for all $\sigma \leq 10^5$. Consequently, since $\mathcal{A}_d^\sigma(2.5)$ is strictly more useful than $\mathcal{A}_d^\infty(2.5)$, for $\sigma < 10^5$, this class II privacy definition provides strictly more utility than $\mathcal{A}_d^\infty(2.5)$.

We will see in the next section that $\mathcal{A}_d^\infty(2.5)$ guarantees privacy that is equivalent to (4,2)-diversity. This shows that realistic class II adversaries not only allows us to protect against adversaries with arbitrary prior shapes, but also can provide more utility than existing privacy definitions.

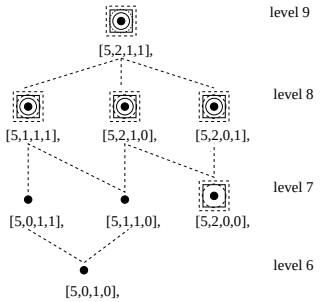


Figure 3: Graphical representation of part of the generalization lattice.

4.2 Comparison with Existing Work

To compare ϵ-privacy to existing work, we compare $(4,2)$-diversity to $\mathcal{A}_d^\infty(2.5)$, and 0.2-closeness to $\mathcal{A}_c^\infty(2.5)$.[7] We also include $\mathcal{A}_-^{10^6}(2.5)$ in our comparison. Figure 3 graphically represents a part of the generalization lattice. Generalizations which satisfy the above privacy definitions are denoted by a square (for $(4,2)$-diversity), a dashed square (for $\mathcal{A}_d^\infty(2.5)$), a circle (for 0.2-closeness), a dashed circle (for $\mathcal{A}_c^\infty(2.5)$), and a little filled circle (for $\mathcal{A}_-^{10^6}(2.5)$). Lines connect tables that are immediate generalizations.

• $\mathcal{A}_d^\infty(\frac{c+1}{2}) \equiv (c,l)$*-diversity:* We observe that all tables that satisfy $(4,2)$-diversity also satisfy 2.5-privacy against $\mathcal{A}_d^\infty(2.5)$, and vice versa. In fact, we found such an equivalence for every generalization in the lattice whenever $\epsilon = \frac{c+1}{2}$. This means that $(c,2)$-diversity and $\mathcal{A}_d^\infty(\frac{c+1}{2})$ allow tables with equivalent privacy and utility to be published. This in turn is not a co-incidence; we show in Section 5 that ϵ-privacy captures many existing privacy definitions.

In conjunction with results from Section 4.1, this equivalence also means that (i) strictly more useful tables can be published by considering class I $\mathcal{A}_d^\sigma(\frac{c+1}{2})$ adversaries than $(c,2)$-diversity; and (ii) in addition to guaranteeing privacy against adversaries with arbitrary prior shapes, class II adversaries (like $\mathcal{A}_-^\sigma(2.5)$, $\sigma < 10^5$) may provide more utility than (c,l)-diversity.

[7]The choice of $t = 0.2, c = 4$ and $\epsilon = 2.5$ is not arbitrary, and reasons will be revealed by the end of this section.

• $\mathcal{A}_c^\infty(\epsilon)$ *versus t-closeness:* A similar equivalence between t-closeness and $\mathcal{A}_c^\infty(\epsilon)$ does not hold even though they seem to guard against the same adversary. There is no value for t such that t-closeness is equivalent to $\mathcal{A}_c^\infty(2.5)$. In our figure the generalization [5,2,0,0] is not 0.2-close, but it is 2.5-private against $\mathcal{A}_c^\infty(2.5)$. If we increase t further then generalization [5,1,1,0] becomes t-close before generalization [5,2,0,0]. This is because t-closeness and ϵ-privacy differ in the way they measure closeness. t-closeness uses earth movers distance, which is an average difference between two distributions. On the other hand ϵ-privacy requires pointwise closeness. Average closeness measures may allow some privacy breaches to occur. For instance, there are tables that satisfy 0.34-closeness (for instance [4,2,0,0]) that have homogeneous groups (which is usually considered a privacy breach). These tables with homogeneous groups do not satisfy ϵ-private for any ϵ.

Nevertheless, the fact that every table that satisfies 0.2-closeness also satisfies $\mathcal{A}_c^\infty(2.5)$ is no co-incidence. We show in the next section that whenever t-closeness does not allow tables with homogenous blocks to be published, then for $\epsilon_t = max_s\{\frac{p_s+t}{p_s}, \frac{p_s}{p_s-t}\}$ ($p_s = n(s)/n$ is the fraction of tuples in the table with sensitive value s), every table that satisfies $\mathcal{A}_c^\infty(\epsilon_t)$ also satisfies t-closeness. For instance, for $t = 0.2$, and $p_{wealthy} = 0.334$, $\epsilon_t = 2.5$. Again in conjunction with results in the previous section,

$\rightsquigarrow$ Strictly more useful tables can be published by considering class I $\mathcal{A}_c^\sigma(\epsilon_t)$ adversaries than t-closeness.

$\rightsquigarrow$ In addition to that guaranteeing privacy against adversaries with arbitrary prior shapes, class II adversaries (like $\mathcal{A}_-^\sigma(2.5)$, $\sigma < 10^5$) may provide more utility than t-closeness.

In summary, we have demonstrated that by considering realistic adversaries who form their beliefs from external datasets we are able to publish tables with strictly more utility than existing techniques. Moreover, for the first time, we are able to publish provably private useful generalized tables even against powerful adversaries with arbitrary priors. Finally, we uncovered an equivalence between $\mathcal{A}_d^\infty(\epsilon)$ and $(c, 2)$-diversity when $\epsilon = \frac{c+1}{2}$; we show in the next section, that is equivalence is no accident, but that ϵ-privacy captures many existing privacy definitions.

5. EMBEDDING PRIOR WORK

We set out to build a framework for defining privacy for the space of adversaries that are neither too weak nor too strong. We achieved this by modeling realistic adversaries who form their knowledge based on external data, and our experiments showed that we can publish useful tables while protecting against strong adversaries.

In this setion, we show that there are much deeper connections to prior work. Our definition is not yet another privacy definition, but it is a framework that encompasses previous definitions. Thus our work does not only span the space between prior work, but also covers the existing end points. We will show that by changing parameters in the ϵ-privacy framework, we can walk along our newly constructed bridge and visit the end points. In particular, we show how our model can instantiate ℓ-diversity, differential privacy, perfect privacy [19] and t-closeness. (Due to space constraints we do not consider (α, β)-privacy breaches and (d, γ)-privacy, which can also be instanted in the ϵ-privacy framework.) Interestingly, all these privacy definitions only protect against ∞-stubborn adversaries.

5.1 $(c, 2)$-Diversity

$(c, 2)$-Diversity requires that the most frequent sensitive attribute appears in at most $\frac{c}{c+1}$ fraction of the tuples in every anonymous group. The privacy guaranteed by this metric is equivalent to considering the following class III adversary.

Sensitive information: For every individual, $\Phi(u)$ contains one predicate $t_u[S] = s$ for every $s \in S$.

Adversary Model: An ∞-stubborn adversary who prior shape is uniform; i.e., the adversary assumes that all individuals are drawn from $\overrightarrow{p^\star}$, where for all $s \in S$, $p^\star(s) = 1/|S|$. Moreover, the adversary does not know exact information about any individual ($B = \emptyset$).

ϵ-Privacy protecting against the above adversary guarantees the same privacy as $(c, 2)$-diversity if and only if

$$\epsilon = \min\left(k \cdot \frac{c}{c+1}, (c+1) \cdot \frac{k-1}{k}\right) \quad (6)$$

$$= \frac{c+1}{2}, \text{ when } k = 2 \quad (7)$$

This is precisely the equivalence we uncovered in the experimental section. Variants of this metric, like $(c, 2)$-diversity [16] and personalized privacy [23] can be captured by changing the set of sensitive predicates. Since we do not consider adversarial background knowledge in the form of negation statements and implications, ϵ-privacy does not yet capture (c, ℓ)-diversity and its variants [18, 7].

5.2 ϵ-Differential Privacy

ϵ-differential privacy requires that an adversary should not be able to distinguish between two input tables that differ only in one tuple using the published data (in fact, by any possible dataset that could be published). Using Claim 3 from Dwork et al. [10], we can show that equivalent privacy can be guaranteed using the following setting of ϵ-privacy.

Sensitive Information: Differential privacy does not distinguish between sensitive and non-sensitive attributes. Hence, for every individual, $\Phi(u)$ contains one predicate $t_u \in D$, for every D a subset of the domain of tuples.

Adversary Model: The adversary has exact information about all but one individuals in the table; i.e., $|B| = |T|-1$. Clearly, no generalization guarantees differential privacy for any ϵ, $T_{pub} - B$ would only contain one tuple. Moreover, the adversary is ∞-stubborn with an arbitrary prior shape.

5.3 Perfect Privacy

Perfect privacy is a very stringent privacy condition which requires that no information be disclosed about any sensitive predicate [19]. More formally, perfect privacy is preserved if for every sensitive predicate $\phi(u)$, the adversary's prior belief about the truth of $\phi(u)$ is *equal* to the adversary's posterior belief about $\phi(u)$ after seeing the published table, no matter what the adversarial prior beliefs are. We state an equivalent formulation of 1-privacy that provides the same privacy protection as perfect privacy.

Sensitive Information: Perfect privacy does not distinguish between sensitive and non-sensitive attributes. Hence, for every individual, $\Phi(u)$ contains one predicate $t_u \in D$, for every D a subset of the domain of tuples.

Adversary Model: An ∞-stubborn adversary whose prior shape is arbitrary.

One interesting question is whether generalization can guarantee 1-privacy against a weaker formulation of perfect

privacy. We can show that if the adversary's prior shape is fixed (say, $\sigma(s)/\sigma$) and known to the data publisher then 1-privacy can be guaranteed by a generalization if for every q anonymous group and every $s \in S$, $\frac{n(q,s)}{n(q)} = \frac{\sigma(s)}{\sigma}$. However, publishing this table provides no utility; since we get no new useful information about the distribution of the sensitive attribute. If the prior shape is arbitrary 1-privacy cannot be guaranteed even against finitely stubborn adversaries.

5.4 t-Closeness

t-closeness requires that the distribution of the sensitive attribute in each anonymous group is close (based on Earth Mover;s Distance) to the distribution of the sensitive attribute in the whole table. The privacy guaranteed by this metric can be equivalently guaranteed by the following class III adversary.

Sensitive information: The same as $(c, 2)$-diversity.

Adversary Model: An ∞-stubborn adversary ($B = \emptyset$) whose prior shape matches the sensitive attribute distribution in the whole table; i.e., $\forall s \in S$, $p_s = \sigma(s)/\sigma = n(s)/n$.

Unlike $(c, 2)$-diversity or differential privacy, there is no setting of ϵ for which ϵ-privacy is equivalent to t-closeness. This is because while ϵ-privacy uses point-wise closeness between distributions, t-closeness uses earth movers distance to measure the closeness of distributions. As a consequence, a table that satisfied t-closeness may have *homogeneous* anonymous groups where all the individuals have the same sensitive value. This is usually considered a breach of privacy, and ϵ-privacy will never allow such a table to be published.

Though the two conditions are not equivalent, under some restrictions t-closeness implies ϵ-privacy.

LEMMA 1. *Any generalized table that does not contain a homogeneous group and satisfies t-closeness, also satisfies ϵ_t-privacy against an ∞-stubborn adversary whose prior shape matches the sensitive attribute distribution in the whole table, if,* $\epsilon_t \geq max_{s \in S} \left(\frac{p_s + t}{p_s}, \frac{p_s}{p_s - t} \right)$

6. RELATED WORK

Recent research has focused on formally defining privacy. We have already discussed k-anonymity, ℓ-diversity, t-closeness, and differential privacy in depth.

Variants of k-anonymity and ℓ-diversity have received copious attention in terms of anonymization algorithms [13], richer data semantics [23], and improved adversarial models [18, 6]. The latter papers bound the worst cases adversarial knowledge that can break an anonymized dataset. (ρ_1, ρ_2)-privacy and γ-amplification [11] bound the point-wise distance between the adversary's prior belief in a property and the adversary's posterior belief in that property after seeing a randomized version of the data. Other approaches [1] ensure (ρ_1, ρ_2)-breaches only on some parts of the data. (d, γ)-privacy [21] is a probabilistic privacy definition for data publishing in which all tuples are considered independent and the privacy is guaranteed by bounding the prior $P(t)$ and the posterior $P(t|D)$ after seeing the published data D.

7. CONCLUSIONS

We introduced ϵ-privacy, a new privacy framework that allows us for the first time to bridge weak and strong adversaries. Our evaluation of ϵ-privacy for generalization shows that it gives practically useful tradeoffs between privacy and utility. The relationship of ϵ-privacy to previous privacy definitions gives interesting insights and opens up directions for future work.

One interesting avenue for future work is to consider correlations between sensitive and non-sensitive values. If an adversary has different priors based on the non-sensitive attributes, we can prove that an expression similar to Thm 3.1 in [16] should be used to compute p^{in}. Analyzing privacy in this case is an important next step. Another interesting avenue for future work is applying ϵ-privacy to other anonymization algorithms.

Acknowledgments. This material is based upon work supported by the DHS under Grant 2006-CS-001-000001, by NYSTAR under Agreement Number C050061, and by the NSF under Grants 0627680 and 0121175. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors.

8. REFERENCES

[1] R. Agrawal, R. Srikant, and D. Thomas. Privacy preserving olap. In *Proceedings of the 23th ACM SIGMOD Conference on Management of Data*, June 2004.

[2] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. From statistics to beliefs. In *National Conference on Artificial Intelligence AAAI*, 1992.

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

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 2003.

[5] MPC Minnesota Population Center. http://ipums.org/.

[6] B. Chen, K. Lefevre, and R. Ramakrishnan. Privacy skyline: Privacy with multidimensional adversarial knowledge. In *VLDB*, 2007.

[7] Bee-Chung Chen, Lei Chen, Raghu Ramakrishnan, and David R. Musicant. Learning from aggregate views. In *ICDE*, 2006.

[8] B. de Finetti. Funzione caratteristica di un fenomeno aleatorio. *Mathematice e Naturale*, 1931.

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

[10] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

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

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

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

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

[15] A. Machanavajjhala, J. Gehrke, and M. Goetz. Using priors to model realistic adversaries. Technical report, Cornell University, 2009.

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

[17] A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vihuber. Privacy: From theory to practice on the map. In *ICDE*, 2008.

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

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

[20] J. B. Paris. *The Uncertain Reasoner's Companion*. Cambridge University Press, 1994.
[21] V. Rastogi, D. Suciu, and S. Hong. The boundary between privacy and utility in data publishing. Technical report, University of Washington, 2007.
[22] Yufei Tao, Xiaokui Xiao, Jiexing Li, and Donghui Zhang. On anti-corruption privacy preserving publication. In *ICDE*, pages 725–734, 2008.
[23] Xiaokui Xiao and Yufei Tao. Personalized privacy preservation. In *SIGMOD*, 2006.
[24] Xiaokui Xiao and Yufei Tao. Output perturbation with query relaxation. In *VLDB '08: Proceedings of the 34th international conference on Very large data bases*, 2008.

APPENDIX

A. DERIVING PRIVACY CONDITIONS

Due to space constraints, we present a proof sketch for Theorem 1. Proofs of Theorems 2, 3 and 4 follow along the same lines. We only derive the privacy condition when $B = \emptyset$ due to space constraints. We refer the reader to our technical report for the complete versions of the proofs [15]. Given T_{pub}, the adversary can find out the q anonymous group that contains and individual, say Rachel's, record (since $t_{Rachel}[Q] = q$ is public information). Given no other information, the adversary cannot distinguish between individuals within an anonymous group. Hence, we can equivalently reason about the privacy of a q-anonymous group.

ϵ-Privacy is guaranteed if for every individual u (with $t_u[Q] = q$) and for every S_ϕ such that $\phi(u) \in \Phi(u)$, $p^{in}/p^{out} \leq \epsilon$ and $(1 - p^{out})/(1 - p^{in}) \leq \epsilon$.
p^{in} and p^{out} can be computed as follows:

$$p^{in}(T_{pub}, q, S_\phi, \overrightarrow{\sigma}, \emptyset) = \sum_{s \in S_\phi} \frac{n(q,s)}{n(q)} \quad (8)$$

$$p^{out}(T_{pub}, q, s, \overrightarrow{\sigma}, \emptyset) \geq \sum_{s \in S_\phi} \frac{n(q,s) + \sigma(s) - 1}{n(q) + \sigma - 1} \quad (9)$$

Given, T_{pub}, an adversary would not know the exact composition of $T_{pub} - \{t_u\}$. If $t = (q, s)$ then $p^{out} = \frac{n(q,s)+\sigma(s)-1}{n(q)+\sigma-1}$, else $p^{out} = \frac{n(q,s)+\sigma(s)}{n(q)+\sigma-1}$. Since we are only interested in how much larger p^{in} is than p^{out}, it is enough to compute the smallest value p^{out} can attain.

Equation 9 leverages the duality between the parameters of the Dirichlet distribution and prior data seen by the adversary. Note that, if the adversary's prior $\overrightarrow{\sigma}$ is interpreted as counts in another dataset disjoint from T_{pub}, the posterior probability that $t[S] = s$ when the t is not in the published table can be interpreted as the fraction of tuples with $S = s$ in all the data that the adversary has seen until now. Moreover, Equation 9 has the following properties:
Prior beliefs vs published data: The adversary's belief about the distribution of the sensitive attribute is a combination of his prior and the statistics in the published table.
Anonymity vs Privacy: When $\sigma < \infty$, as the number of individuals in each q-anonymous group increases, the difference between posterior beliefs when the individual is in or out of the published table shrinks.

Theorem 1 (when $B = \emptyset$) follows from simple arithmetic.

$$\frac{p^{in}}{p^{out}} \leq \frac{n(q, S_\phi)}{n(q, S_\phi) + \sigma(q, S_\phi) - 1} \cdot \frac{n(q) + \sigma(q) - 1}{n(q)} \leq \epsilon$$

$$\text{iff,} \quad \delta \geq 1 \quad \text{or} \quad \frac{n(q, S_\phi)}{n(q)} \leq \frac{\epsilon}{1 - \delta} \cdot \frac{\sigma(q, S_\phi) - 1}{\sigma(q)}$$

$$\frac{1 - p^{out}}{1 - p^{in}} \leq \left(1 - \frac{n(q, S_\phi) + \sigma(q, S_\phi) - 1}{n(q) + \sigma(q) - 1}\right) \Big/ \left(\frac{n(q) - n(q, S_\phi)}{n(q)}\right) \leq \epsilon$$

$$\text{iff,} \quad \frac{n(q, S_\phi)}{n(q)} \leq 1 - \frac{1}{\epsilon' + \delta} + \frac{1}{\epsilon' + \delta} \cdot \frac{\sigma(q, S_\phi) - 1}{\sigma(q)}$$

■

Proof of Equation 8.

$$p^{in}(T_{pub}, q, s, \overrightarrow{\sigma}, \emptyset) = \frac{Pr[t = (q,s) \wedge t \in T_{pub}]}{\sum_{s' \in S} Pr[t = (q,s') \wedge T_{pub}]} \quad (10)$$

The probability of seeing the dataset T_{pub} equals the probability of seeing the histogram $H = (\ldots, n(q,s), \ldots)$ from n draws using some $\overrightarrow{p}$ drawn from the distribution $D(\overrightarrow{\sigma})$.

$$Pr[T_{pub}] = \int_{\overrightarrow{p}} Pr[T_{pub}|\overrightarrow{p}] \cdot Pr[\overrightarrow{p}|D(\overrightarrow{\sigma})] d\overrightarrow{p}$$
$$= \frac{n!}{\prod n(q,s)!} \frac{\prod \Gamma(n(q,s)+\sigma(q,s))}{\prod \Gamma(\sigma(q,s))} \frac{\Gamma(\sigma)}{\Gamma(n+\sigma)} \quad (11)$$

The probability of seeing the dataset T_{pub} where $t = (q,s)$ equals the product of (a) the probability of $t = (q,s)$ given $\overrightarrow{p}$, and (b) the probability of seeing the histogram $H = (\ldots, n(q,s) - 1, \ldots)$ from $n - 1$ draws from $\overrightarrow{p}$, for some $\overrightarrow{p}$ drawn from $D(\overrightarrow{\sigma})$. Let $T'_{pub} = T_{pub} - \{t\}$.

$$Pr[t = (q,s) \wedge T_{pub}]$$
$$= \int_{\overrightarrow{p}} Pr[t = (q,s)|\overrightarrow{p}] \cdot Pr[T'_{pub}|\overrightarrow{p}] \cdot Pr[\overrightarrow{p}|D(\overrightarrow{\sigma})] d\overrightarrow{p}$$
$$= \frac{n(q,s)}{n(q)} \times Pr[T_{pub}] \quad (12)$$

Thus using Equations 10 and 12 we get the result. ■

Proof of Equation 9.

$$p^{out}(T_{pub}, q, s, \overrightarrow{\sigma}, \emptyset) = \frac{Pr[t = (q,s) \wedge t \notin T_{pub}]}{\sum_{s' \in S} Pr[t = (q,s') \wedge t \notin T_{pub}]}$$

$$Pr[t = (q,s) \wedge T_{pub} \wedge t \notin T_{pub}]$$
$$= \frac{\sigma(q,s)}{\sigma} \int_{\overrightarrow{p}} Pr[T_{pub}|\overrightarrow{p}] \cdot Pr[\overrightarrow{p}|D(\overrightarrow{\sigma'})] d\overrightarrow{p} \quad (13)$$

where $\sigma' = \sigma + 1$, $\sigma'(q,s) = \sigma(q,s) + 1$, and for all other $(q', s') \in Q \times S$, $\sigma'(q', s') = \sigma(q', s')$. From Equation 11

$$Pr[t = (q,s) \wedge T_{pub} \wedge t \notin T_{pub}]$$
$$= \frac{n(q*,s) + \sigma(q*,s)}{n + \sigma} \times Pr[T_{pub}] \quad (14)$$

Thus from Equations 13 and 14 we get the result. ■