Future Generation Computer Systems 33 (2014) 53–63

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Measuring the privacy of user profiles in personalized information systems☆

Javier Parra-Arnau*, David Rebollo-Monedero, Jordi Forné

Department of Telematics Engineering, Universitat Politècnica de Catalunya (UPC), C. Jordi Girona 1-3, 08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 15 June 2012
Received in revised form
19 November 2012
Accepted 4 January 2013
Available online 17 January 2013

Keywords:
Personalized information systems
User profiling
Privacy-enhancing technologies
Privacy criterion
Shannon's entropy
Kullback–Leibler divergence

ABSTRACT

Personalized information systems are information-filtering systems that endeavor to tailor information-exchange functionality to the specific interests of their users. The ability of these systems to profile users is, on the one hand, what enables such intelligent functionality, but on the other, the source of innumerable privacy risks. In this paper, we justify and interpret KL divergence as a criterion for quantifying the privacy of user profiles. Our criterion, which emerged from previous work in the domain of information retrieval, is here thoroughly examined by adopting the beautiful perspective of the method of types and large deviation theory, and under the assumption of two distinct adversary models. In particular, we first elaborate on the intimate connection between Jaynes' celebrated method of entropy maximization and the use of entropies and divergences as measures of privacy; and secondly, we interpret our privacy metric as false positives and negatives in a binary hypothesis testing.

1. Introduction

Recent years have witnessed the accelerated growth of a rich variety of personalized information systems of unprecedented sophistication, which have been integrating seamlessly into our daily lives. Examples of these systems comprise personalized Web search and news, resource tagging in the semantic Web and multimedia recommendation systems. The key enabling technology of such systems is personalization, a research area that has received great attention lately and whose aim is to tailor information-exchange functionality to the specific interests of their users. To accomplish this functionality, most personalized information systems capitalize on, or lend themselves to, the construction of profiles, either directly declared by a user, or inferred from past activity, not only of the user in question, but also from the profiles of users with whom social relationships are known to the information system.

Personalized services therefore allow users to deal with the overwhelming overabundance of information, but inevitably at the expense of privacy, especially when profiling is conducted across several information systems. Besides, the enrichment of these services with data from social networks creates additional opportunities with respect to information sharing but, at the same time, increases the user privacy risks.

But the advent of these information systems is not only changing people's habits and stressing our concerns about privacy—it is also leading to a profound transformation of the traditional business model. As a matter of fact, the technologies enabling personalization as a solution of the one-size-fits- all are contributing to unprecedented performance improvements in large business and small and medium enterprises. These technologies are having an impact not only on how products are sold but also, and more importantly, on how companies approach users in a personalized manner, attending their specific and particular needs more effectively. Amazon, for example, who invented item-to-item collaborative-filtering algorithms [1], one of the most widely used personalization techniques, is visited by more than 93 million users per day. Another example that illustrates this transformation is Facebook, which will surpass 4.27 billion dollars in revenue this year, 89% of its income will come from selling access to their data so that advertisers can personalize their digital content [2]. The information used to provide such personalization ranges from location, education, likes and interests, to friends and relationship status [3]. Pushed by these personalization techniques, online advertising is expected to grow by 10.6% each year through 2016, with $70.9 billion in global advertising during 2011.

The impact of personalized information systems on society and economy is therefore undeniable. Nowadays, personalization is present in a myriad of applications we frequently use on the Internet, when submitting queries to a Web search engine, rating products at an online store or posting tags in a collaborative

☆ The material in this paper has been published in part in the proceedings of the International Conference on Security Technology (SecTech), Jeju, South Korea, Dec. 2011.

* Corresponding author. Tel.: +34 93 401 7041.
E-mail addresses: javier.parra@entel.upc.edu (J. Parra-Arnau), david.rebollo@entel.upc.edu (D. Rebollo-Monedero), jforne@entel.upc.edu (J. Forné).

0167-739X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.future.2013.01.001

tagging system. But this is only the tip of the iceberg – in the near future a much wider spectrum of services such as personalized medicine will become a reality. However, we must not forget that the cornerstone of these current and future systems is the ability to profile users, which poses serious threats to one of our fundamental rights – the right to privacy.

1.1. The need for measuring the privacy of user profiles

A variety of privacy-enhancing technologies (PETs) have been proposed to enable the provision of new services and functionalities aimed at mitigating those privacy threats. Anonymous-communication networks [4,5], anonymous credentials [6], anonymous electronic cash [7], multiparty computation [8] and oblivious transfer protocols [9] are some examples of general-purpose PETs whose development roughly originates from the fields of security and cryptography. Unfortunately, these technologies have not yet gained wide adoption. This is because it remains unclear whether their overall benefits outweigh their typically costly deployment and/or integration, as well as the operational cost that arises due to the fact that PETs typically come with penalties in terms of utility and performance, when compared to more privacy-invasive alternatives [10].

Assessing the privacy provided by a PET is, therefore, crucial to both determine its overall benefit and compare its effectiveness with other technologies. In other words, privacy metrics, accompanied with utility metrics, provide a quantitative means of contrasting the suitability of two or more privacy-enhancing mechanisms, in terms of the privacy–utility trade-off posed. Ultimately, such metrics enable us to systematically build privacy-aware information systems by formulating design decisions as optimization problems, solvable theoretically or numerically, capitalizing on a rich variety of mature ideas and powerful techniques from the wide field of optimization engineering.

A great effort has been devoted to the investigation of privacy metrics, especially in the scenario of statistical disclosure control (SDC) [11–18]. Although some of those metrics might be applied to our context of personalized information systems, the fact is that there are few proposals specifically conceived for measuring the privacy of user profiles; and not only that, but also they are often not appropriately justified and are defined in an ad hoc manner [19–29].

1.2. Contribution and organization

This paper approaches the fundamental problem of proposing quantitative measures of the privacy of user profiles. We have established the critical importance of quantifying privacy in order to assess, compare, improve and optimize privacy-enhancing technologies. In application scenarios involving user profiles, there exists no general framework systematically leading to a formal metric, but merely ad hoc proposals for a few specific applications. The main contribution of this work identifies the need for such quantitative measures of privacy for user profiles in personalized information systems.

Bearing this need in mind, we explore the privacy risks inherent in such systems, and then provide a thorough justification of a common, generalized framework to measure those risks. Our justification relies on fundamental principles from information theory and statistics, thereby drawing intriguing links between said fields and information privacy. In practice, the impact of a privacy mechanism on information-exchange functionality, traffic and processing overhead, and general usability cannot simply be overlooked. We would like to stress that quantitative measures of privacy on the one hand, and utility on the other, allow researchers to optimize their technologies in terms of the trade-off posed by these contrasting aspects.

Specifically, we tackle two adversary models. The first model considers an attacker aimed at targeting users who deviate from the average profile of interests; and the second one contemplates an attacker whose objective is to classify a given user into a predefined group of users. Under the former model, the use of Kullback–Leibler (KL) divergence as a measure of privacy is justified by elaborating on Jaynes' rationale behind entropy-maximization methods and the method of types, a justification that we introduced in [30]. Under the latter adversary model, a riveting argument in favor of divergence stems from hypothesis testing and large deviation theory.

Section 2 illustrates the privacy concerns that arise in the motivating scenario of this work. Section 3 examines several approaches to model user profiles and specifies the adversary capabilities assumed in our interpretation of divergence as a measure of privacy. The use of divergence is justified on the one hand in Section 4 when the attacker strives to identify users, and on the other in Section 5 when the adversary endeavors to classify users. Section 6 then overviews some of the most relevant privacy criteria in the literature. Finally, conclusions are drawn in Section 7.

2. Illustration of privacy risks in personalized information systems

In this section, we carefully examine the privacy risks posed by the personalized information systems that proliferate these days on the Internet. The following example illustrates those privacy threats.

Jane Doe is about to finish a long day of work in the patent department of her law firm in New York City. It has been a pretty hectic week, due to the forthcoming, albeit still unannounced, release of a spanking new model of smartphone by Apple. This patent is by far her favorite legal case, as she enjoys keeping herself up to date on the latest technological gadgets, often browsing for them via Google search and YouTube. She also loves how, these days, online tools retrieve both intelligent search results and videos, almost anticipating her interests, undoubtedly learning from her past activity. Unsurprisingly, after health, she rated technology highest when customizing her preferences in Google News, which she accesses almost religiously every morning. Her boyfriend, a computer scientist, keeps telling her that the future of information systems lies in their personalization, by means of automated compilation of user profiles, implicitly from behavior or explicitly from declared interests. Sounds about right.

Jane is aware that her company may be tracking her work habits by monitoring the use of applications and Internet access, with tools such as Track4Win. Still, before turning off her desktop computer at work, she quickly checks a friend's post in Twitter confirming a meeting this Friday evening to chat about tomorrow's protest, organized by the Occupy Wall Street movement, against the budget cuts planned by the government. She promptly responds, and adds a link to an intriguing article on the subject in The New York Times, an American newspaper with left-wing views.

They are meeting at "Café Lalo", a famous café on the Upper West Side. During the half-hour bus ride to that location, Jane uses her iPhone to log into Facebook, to find the lovely pictures of her cousin's newborn baby. She politely types a cheerful comment in the album congratulating the happy family. Over the last few months, she and her boyfriend have been seriously considering having a baby, although she wishes her job at the law firm would offer a better work–life balance. Still a few bus stops to go, giving her ample time to discover a couple of new Web sites on childbearing, one of them showing Facebook's "like" button, which she immediately presses almost as a reflex response. Of course, her action will be diligently reflected back in her profile. In a way,

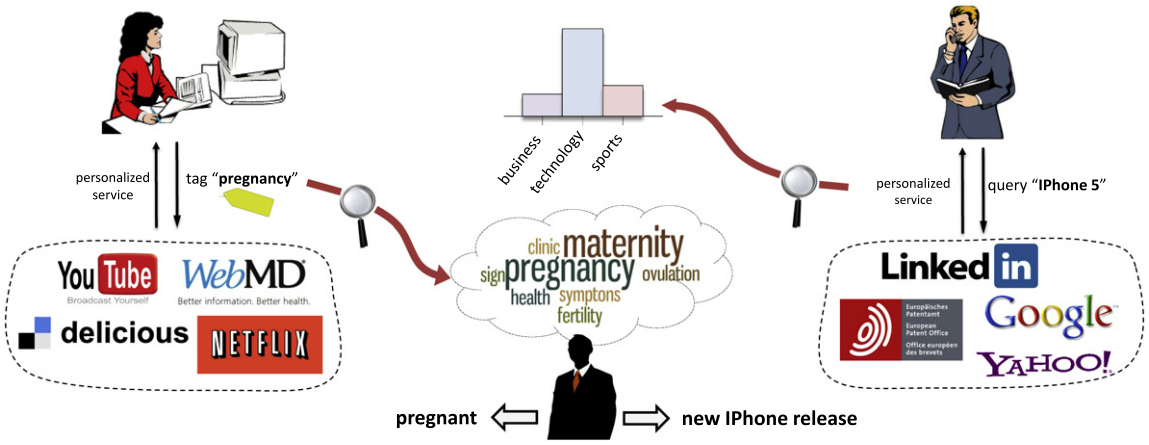

Fig. 1. During an Internet session through various personalized information systems, users leave innumerable traces of sensitive information which, especially in combination, pose serious risks, not only to their own privacy, but also to the privacy of others.

social networks are personalized information systems, reactively and proactively providing media tailored to their users' profiles of interests, built on the basis of their social interactions. She also notes a new friend request in Facebook, coming from a coworker in the human resources department. Even though their relationship is strictly professional, she finally accepts the request out of courtesy.

Comfortably seated in the café, while waiting for her friend, Jane continues using her smartphone to turn to Delicious, a social Web service where millions share and tag their favorite bookmarks. Luckily, she comes across a bookmark pointing to a site advertising an interesting job opportunity, also in the area of patents, in a law firm with more flexible hours, which she tags with the description "work–life balance", having her plans to get pregnant in mind. However, she is not sure whether she has to seriously consider this job opportunity since she is unfamiliar with both the law firm and the bookmark's author. Her friend arrives a few minutes late, but they both have a pleasant evening.

Little does Jane know that, during her Internet expedition from Google search to Delicious, passing by Google News, Twitter and Facebook, among other sites, she has left innumerable traces of sensitive information which, especially in combination, pose a serious risk, not only to her own privacy, but also to the privacy of others. Hypothetically speaking, Google could correlate queries on smartphones with patents and Jane's declared interests on technology news, with IP addresses, presumably targeting her computer at work, from which she recently posted a detailed CV in LinkedIn, and thus learning about her occupation. Gathering additional evidence confirming a surge in query activity on the subject from similar sources, Google could be led to infer that Apple is likely to release the new iPhone 5, and retaliate by moving forward the new Android version.

Also hypothetically, someone in the department of human resources in Jane's law firm, which has started considering her promotion, could have attempted to become friends and inspect her Facebook profile to deduce the existence of a statistical chance of her having pregnancy plans. Further, her Twitter account is indicative of leftist views that might conflict with the political convictions of the company management. The fact that she uses a pseudonym in Delicious may not prevent the computer specialist in the human resources department from correlating users with tags related to law, patents, smartphones, pregnancy and the political Occupy Wall Street movement to guess her actual identity, and find out about her interest in job positions with a better work–life balance. Not to mention the monitoring of her work habits and activity profile with Track4Win. Any of this could presumably endanger her promotion or even her current position. Some of these privacy risks are conceptually depicted in Fig. 1.

3. Adversary models and privacy-enhancing strategies

In this section, first we shall examine some approaches to model user profiles. Afterwards, we shall describe some considerations on the adversary capabilities. Those considerations will result in the definition of two objectives for a privacy attacker, namely *identification* and *classification*. The distinction between these two objectives will enable us to justify our privacy criterion under two different perspectives later in Sections 4 and 5. The last part of this section will be devoted to the presentation of several data-perturbative approaches for the privacy protection of user profiles and the illustration of the privacy–utility trade-off they pose.

3.1. User profile model

In the motivating scenario of this work, a user submits queries to a Web search engine, clicks on news links in a personalized news recommendation system, and assigns tags to resources (e.g., photos, videos and bookmarks) on the Web, all according to his/her profile of interests. The information conveyed, i.e., queries, news clicked and tags, allows those systems to extract a profile of interests or *user profile*, which turns to be essential in the provision of personalized services.

In the context of personalized information systems, user profiles are frequently modeled as histograms. For example, collaborative tagging systems commonly represent profiles by using tag clouds, which, in essence, may be regarded as histograms. Recall that a tag cloud is a visual depiction in which tags are weighted according to their frequency of use. Those two possible representations for user profiles, tag clouds and histograms, are, in fact, simultaneously used in popular tagging systems such as BibSonomy, CiteULike and Delicious.

In the scenario of personalized recommendation systems, we also find examples of profiles modeled as histograms, especially in content-based recommenders [31] such as Jinni. Of particular interest is the case of Google News, where news are classified into a predefined set of topic categories; and accordingly, users are modeled by their distribution of clicks on news, i.e., as histograms of relative frequencies of clicks within that set of categories [32]. In this same spirit, recent privacy-protecting approaches in the scenario of recommendation systems also propose using histograms of absolute frequencies for modeling user profiles [33,34].

Motivated by these examples and inspired by other works in the field [23–26,28,29], in this paper we justify and interpret a privacy criterion under the assumption that user profiles are modeled as probability mass functions (PMFs), that is, as histograms of

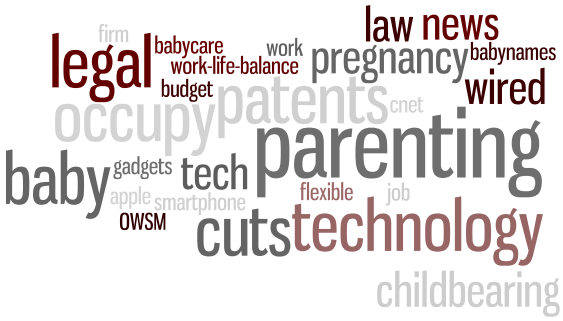

Fig. 2. User profile modeled as a tag cloud in a collaborative tagging system. The tags posted by users are frequently depicted as tag clouds, not only in those tagging systems, but also in multimedia recommendation systems such as Jinni. Recall that a tag cloud is a visual representation where the font size of each tag is proportional to its frequency of use.

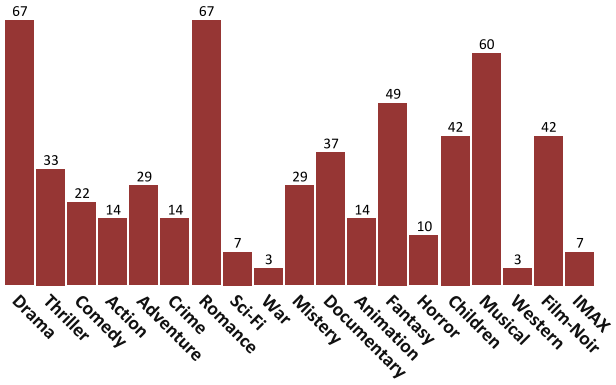

Fig. 3. User profile modeled as a histogram of absolute frequencies of ratings within a set of predefined movie genres. Many personalized information systems use this kind of representation, or slight variations of this idea, to model user interests.

relative frequencies of user data (e.g., queries, news clicks and tags) within a set of categories of interest. Therefore, our user profile model is in line with the representations used in numerous tagging systems and personalized recommendation systems. Fig. 2 shows an example of user profile that could perfectly resemble the case of Jane Doe described in Section 2. Fig. 3, on the other hand, depicts the profile of a user as shown in Movielens, a movie recommender.

3.2. Adversary model

In Section 1.1 we stressed the need for privacy metrics as the only way to evaluate, compare and design privacy-protecting mechanisms. When measuring the level of privacy provided by a PET, however, it is essential to specify the concrete assumptions about the adversary, that is, its capabilities, properties or powers; this is known as the *adversary model*. The importance of such a model lies in the fact that the level of privacy provided is measured with respect to it. In other words, if the assumptions change, so does the metric.

In our scenario of personalized information systems, we assume that the set of potential privacy attackers encompasses any entity capable of eavesdropping the information users convey to providers. Accordingly, both service providers and network operators are deemed potential attackers. But since this information, frequently in the form of ratings, tags and comments, is often publicly available to other users of the system, any entity able to collect this information is taken into consideration in our adversary model.

Under this assumption, and as a response to the privacy threats illustrated in Section 2, a user may counter by submitting false ratings, refraining from posting certain tags or comments, or in general, adopting any data-perturbative mechanism that enhances their privacy. As a result, the attacker observes a perturbed version of the genuine profile and is unaware or ignores the fact that the observed profile does not reflect the actual interests of the user. Hereafter, we shall refer to these two profiles as the *actual* profile and the *apparent* profile.

That said, now we contemplate two possible, mutually-exclusive objectives for the attacker:

- On the one hand, we consider the attacker strives to target users that deviate from the average profile of interests. We refer to this objective as *identification*, not because the attacker wishes to ascertain the identity of a user (e.g., name, social security number, etc.), but because the adversary aims to discriminate the user from the whole population of users.
- On the other hand, we assume that the attacker's goal is to classify a user into a predefined group of users. To conduct this *classification*, the attacker contrasts the user's profile with the profile representative of a particular group.

These two objectives, together with the assumptions above, constitute the adversary models upon which our privacy metric builds. In Sections 4 and 5, we shall justify a criterion that, contrary to what was said at the beginning of this section, may be applied to both adversary models, consistently with the objectives of identification and classification. The application of these two models, however, will lead us to consider KL divergence as a measure of *privacy risk* against identification, and a measure of *privacy gain* against classification.

3.3. Privacy-enhancing strategies

The literature of privacy abounds with examples of technologies aimed at protecting users against adversaries like these. However, while these technologies may enhance user privacy to a certain extent, they inevitably come at the expense of a loss in data utility. In a nutshell, PETs pose an inherent trade-off between the contrasting aspects of privacy on the one hand, and utility on the other. In this subsection, we elaborate on several data-perturbative strategies for the protection of user profiles, and examine the privacy–utility trade-off they pose.

An interesting approach to provide a distorted version of a user's profile of interests consists in query forgery. The underlying principle, conceptually illustrated in Fig. 4, boils down to accompanying original queries or query keywords with bogus ones. By adopting this data-perturbative strategy, users prevent privacy attackers from profiling them accurately based on their queries, without having to trust either the service provider or the network operator, but clearly at the cost of traffic overhead. In other words, inherent to query forgery is the existence of a trade-off between privacy and additional traffic. Precisely, [23] studies how to optimize the introduction of forged queries in the scenario of information retrieval. Particularly, the authors measure privacy risk as the relative entropy between the perturbed profile and the population's distribution, and propose *query redundancy*, i.e., the ratio of forged queries to total queries, as utility metric. Building on this principle, several protocols, mainly heuristic, have been proposed and implemented, with various degrees of sophistication [35,36,20].

Clearly, the perturbation of user profiles for privacy preservation may be carried out not only by means of the insertion of bogus activity, but also by suppression. An example of this latter kind of perturbation may be found in [24], where the authors propose the elimination of tags as a privacy-enhancing strategy in the scenario of the semantic Web. On the one hand, this strategy allows users to enhance their privacy to a certain degree, but on the other

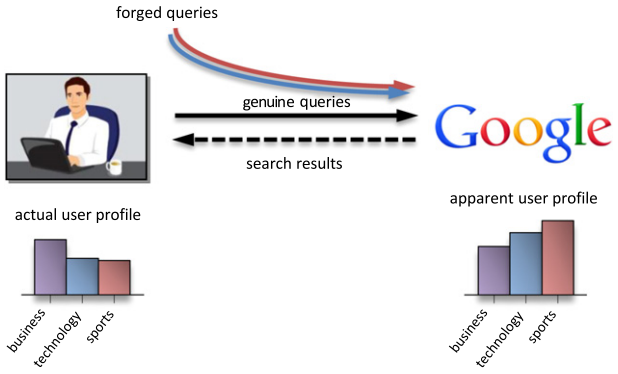

Fig. 4. Query forgery in personalized Web search. A user submits false queries, accompanied with genuine queries, to perturb his actual profile of interests. By adopting query forgery, the adversary, possibly the service provider itself, observes a distorted version of his profile. We refer to this profile as the apparent user profile.

it comes at the cost of a degradation in the semantic functionality of the Web, as tags have the purpose of associating meaning with resources. Precisely, [28] investigates mathematically the privacy–utility trade-off posed by the suppression of tags, measuring privacy as the Shannon's entropy of the perturbed profile and utility as the percentage of tags users are willing to eliminate. Intimately related to this work is [29], where the impact of tag suppression is assessed experimentally in the context of resource recommendation and parental control, in terms of percentages regarding missing tags on resources on the one hand, and in terms of false positives and negatives on the other.

The combined use of both strategies, that is, forgery and suppression, is studied in the scenario of personalized recommendation systems [25]. With the adoption of those strategies, users may wish to submit false ratings to items that do not reflect their preferences, and/or refrain from rating certain items they have an opinion on. Another approach, now from the perspective of collaborative peers, suggests that two or more users exchange a portion of their queries before submitting them, in order to obfuscate their respective interest profiles versus the network operator or external observers [26]. The idea of privacy through multiple user collaboration has also been investigated in the form of a number of protocols, such as [37,38].

Yet another example illustrating the trade-off between privacy and utility arises in the field of anonymous communications, where the goal is to conceal who is communicating with whom against an adversary who observes the inputs and outputs of the anonymous-communication channel. In this field, mixes [4,39,40] are a basic building block for implementing anonymous-communication channels. Essentially, mixes perform cryptographic operations on messages, and delay and reorder them with the aim of hindering the linking of inputs and outputs based on timing information. While delaying messages may provide such unlinkability, clearly it has an impact on the usability of the system, and therefore imposes a cost on the system. Put another way, there is a compromise between unlinkability (anonymity) and delay (utility).

4. Measuring privacy against identification

Next, we shall proceed with our first interpretation of KL divergence as a privacy criterion. Both in this section and in Section 5, the information-theoretic arguments and justifications in favor of our metric will be expounded in a systematic manner, following the points sketched in Fig. 5. In the section at hand, we shall interpret divergence and entropy under the assumptions of the adversary model defined in Section 3.2, in the special case when the attacker's objective is to identify a user in the sense of discriminating this user from all other users; this interpretation corresponds to the first branch of the tree in Fig. 5, which we term *identification*. For that purpose, we shall adopt the perspective of Jaynes' celebrated *rationale on entropy maximization methods* [41], which builds upon the *method of types* [42, Section 11], a powerful technique in large deviation theory whose fundamental results we also explore in this section.

The first part of this section, Section 4.1, reviews some basic concepts of information theory. Afterwards, Section 4.2 tackles an important question. Suppose we are faced with a problem, formulated in terms of a model, in which a probability distribution plays a major role. In the event this distribution is unknown, we wish to assume a feasible candidate. What is the most likely probability distribution? In other words, what is the "probability of a probability" distribution? We shall see that a widespread answer to this question relies on choosing the distribution *maximizing the Shannon entropy*, or, if a reference distribution is available, the distribution *minimizing the KL divergence* with respect to it, commonly subject to feasibility constraints determined by the specific application at hand.

Our review of the maximum entropy method is crucial because it is unfortunately not always known in the privacy community. As we shall see in the last part of this section, Section 4.3, the key idea is to model a user profile as a probability distribution, as considered in Section 3.1, apply the maximum entropy method to measure the likelihood of a user profile either as its entropy or as its divergence with respect to the population's average profile, and finally take that likelihood as a measure of anonymity.

4.1. Statistical and information-theoretic preliminaries

This section establishes notational aspects, and, in order to make the presentation of our privacy criterion suited to a wider audience, recalls key information-theoretic concepts assumed to be known in the remainder of the paper. The measurable space in which a *random variable* (r.v.) takes on values will be called an *alphabet*, which, with a mild loss of generality, we shall always assume to be finite. We shall follow the convention of using uppercase letters for r.v.'s, and lowercase letters for particular values they take on. The PMF p of an r.v. X is essentially a *relative histogram* across the possible values determined by its alphabet.

Informally, we shall occasionally refer to the function p by its value $p(x)$. The *expectation* of an r.v. X will be written as EX, concisely denoting $\sum_x x\,p(x)$, where the sum is taken across all values of x in its alphabet.

We adopt the same notation for information-theoretic quantities used in [42]. Concordantly, the symbol H will denote entropy and D relative entropy or KL divergence. We briefly recall those concepts for the reader not intimately familiar with information theory. All logarithms are taken to base 2. The *entropy* $H(p)$ of a discrete r.v. X with probability distribution p is a measure of its uncertainty, defined as

$$H(X) = -E\,\log p(X) = -\sum_x p(x)\log p(x).$$

Given two probability distributions $p(x)$ and $q(x)$ over the same alphabet, the *KL divergence* or *relative entropy* $D(p\parallel q)$ is defined as

$$D(p\parallel q) = E_p\,\log\frac{p(X)}{q(X)} = \sum_x p(x)\log\frac{p(x)}{q(x)}.$$

The KL divergence is often referred to as *relative entropy*, as it may be regarded as a generalization of the entropy of a distribution, relative to another. Conversely, entropy is a special case of KL

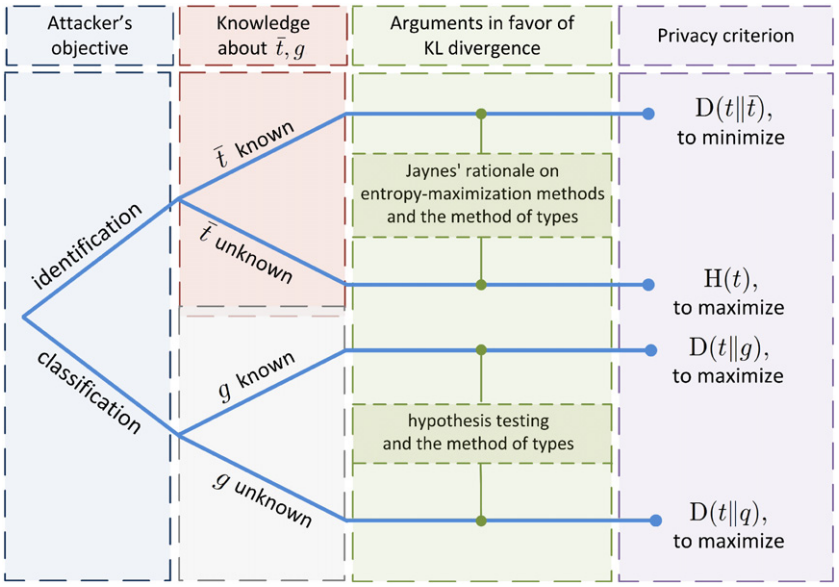

Fig. 5. Summary of our interpretations of KL divergence and Shannon's entropy as measures of privacy. This figure illustrates, at a conceptual level, the assumptions upon which our privacy criterion builds. First, we follow Jaynes' rationale behind entropy-maximization methods to justify divergence and entropy when the attacker's goal is to *identify* users. The knowledge of the population's distribution $\bar{t}$ determines whether the metric to be used is divergence or entropy. Secondly, when the attacker aims at *classifying* a user as a member of a particular group, our arguments in favor of divergence stem from hypothesis testing and the method of types. In the special case when the group profile g is unknown to the user, they may wish to maximize the divergence between the actual profile q and the perturbed, observed profile t, in order to avoid being classified as they actually are.

divergence, as for a uniform distribution u on a finite alphabet of cardinality n,

$$D(p \parallel u) = \log n - H(p). \tag{1}$$

The *cross entropy* of two probability distributions $p(x)$ and $q(x)$ over the same alphabet is defined as

$$H(p \parallel q) = -E_p \log q(X) = -\sum_x p(x) \log q(x),$$

from whence it follows that

$$H(p \parallel q) = H(p) + D(p \parallel q).$$

Although the KL divergence is not a distance in the mathematical sense of the term, because it is neither symmetric nor satisfies the triangle inequality, it does provide a measure of discrepancy between distributions, in the sense that $D(p \parallel q) \geqslant 0$, with equality if, and only if, $p = q$. On account of this fact, relation (1) between entropy and KL divergence implies that $H(p) \leqslant \log n$, with equality if, and only if, $p = u$. Simply put, *entropy maximization* is a special case of *divergence minimization*, attained when the distribution taken as the optimization variable is identical to the *reference distribution*, or as "close" as possible, should the optimization problem appear accompanied with *constraints* on the desired space of candidate distributions.

4.2. Rationale behind the maximum entropy method

A wide variety of models across diverse fields have been explained on the basis of the intriguing principle of entropy maximization. A classical example in physics is the Maxwell–Boltzmann probability distribution $p(v)$ of particle velocities V in a gas [43,44] of known temperature. It turns out that $p(v)$ is precisely the probability distribution maximizing the entropy, subject to a constraint on the temperature, equivalent to a constraint on the average kinetic energy, in turn equivalent to a constraint on EV^2. Another well-known example, in the field of electrical engineering, of the application of the maximum entropy method, is Burg's spectral estimation method [45]. In this method, the power spectral density of a signal is regarded as a probability distribution of power across frequency, only partly known. Burg suggested filling in the unknown portion of the power spectral density by choosing that maximizing the entropy, constrained on the partial knowledge available. More concretely, in the discrete case, when the constraints consist in a given range of the cross-correlation function, up to a time shift k, the solution turns out to be a kth order Gauss–Markov process [42]. A third and more recent example, this time in the field of natural language processing, is the use of log-linear models, which arise as the solution to constrained maximum entropy problems [46] in computational linguistics.

Having motivated the maximum entropy method, we are ready to proceed to describe Jaynes' attempt to justify, or at least interpret it, by reviewing the method of types of large deviation theory, a beautiful area lying at the intersection of statistics and information theory. Let $X_1, \ldots, X_k$ be a sequence of k independent and identically distributed (i.i.d.) drawings of an r.v. uniformly distributed in the alphabet $\{1, \ldots, n\}$. Let k_i be the number of times symbol $i = 1, \ldots, n$ appears in a sequence of outcomes $x_1, \ldots, x_k$, thus $k = \sum_i k_i$. The *type* t of a sequence of outcomes is the relative proportion of occurrences of each symbol, that is, the *empirical distribution* $t = \left(\frac{k_1}{k}, \ldots, \frac{k_n}{k}\right)$, not necessarily uniform. In other words, consider tossing an n-sided fair dice k times, and seeing exactly k_i times face i. In [41], Jaynes points out that

$$H(t) = H\left(\frac{k_1}{k}, \ldots, \frac{k_n}{k}\right) \simeq \frac{1}{k} \log \frac{k!}{k_1! \cdots k_n!} \quad \text{for } k \gg 1.$$

Loosely speaking, for large k, the size of a *type class*, that is, the number of possible outcomes for a given type t (permutations with repeated elements), is approximately $2^{kH(t)}$ in the exponent. The fundamental rationale in [41] for selecting the type t with maximum entropy $H(t)$ lies in the approximate equivalence between entropy maximization and the maximization of the number of possible outcomes corresponding to a type. In a way, this justifies the infamous *principle of insufficient reason*, according to which, one may expect an approximately equal relative frequency $k_i/k = 1/n$ for each symbol i, as the uniform distribution maximizes the entropy. The principle of entropy maximization is extended to include constraints also in [41].

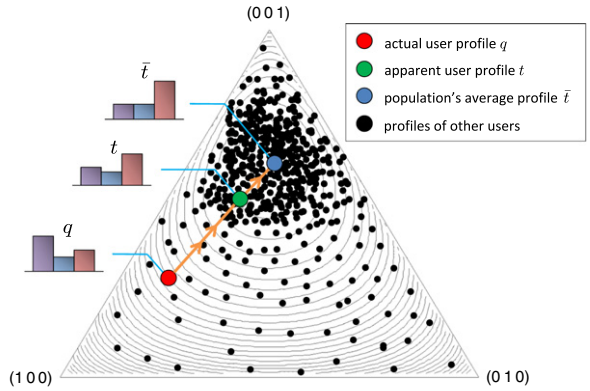

Fig. 6. A privacy attacker aims at distinguishing a particular user among the population of users. Under Jaynes' rationale, KL divergence may be regarded as a measure of user profile density. This is, certainly, under the assumption that this particular user does not know the distribution of profiles here depicted. Accordingly, the user adopts some perturbative strategy whereby the observed profile t gets close, in terms of divergence, to the average population's distribution $\bar{t}$. As a result, the apparent profile becomes more common, getting lost in the crowd, and thus thwarting the attacker's intention.

Obviously, since all possible permutations count equally, the argument only works for uniformly distributed drawings, which is somewhat circular. A more general argument [42, Section 11], albeit entirely analogous, starts with a prior knowledge of an arbitrary PMF $\bar{t}$, not necessarily uniform, of such samples $X_1, \ldots, X_k$. Because the empirical distribution or type T of an i.i.d. drawing is itself an r.v., we may define its PMF $p(t) = P\{T = t\}$; formally, the PMF of a random PMF. Using indicator r.v.'s, it is straightforward to confirm the intuition that $ET = \bar{t}$. The general argument in question leads to approximating the probability $p(t)$ of a type class, a fractional measure of its size, in terms of its relative entropy, specifically $2^{-kD(t\|\bar{t})}$ in the exponent, i.e.,

$$D(t \parallel \bar{t}) \simeq -\frac{1}{k} \log p(t) \quad \text{for } k \gg 1,$$

which encompasses the special case of entropy, by virtue of (1). Roughly speaking, the likelihood of the empirical distribution t exponentially decreases with its KL divergence with respect to the average, reference distribution $\bar{t}$.

In conclusion, the most likely PMF t is that minimizing its divergence with respect to the reference distribution $\bar{t}$. In the special case of uniform $\bar{t} = u$, this is equivalent to maximizing the entropy, on account of (1), possibly subject to constraints on t that reflect its partial knowledge or a restricted set of feasible choices.

4.3. Measuring the privacy of user profiles

We are finally equipped to justify, or at least interpret, our proposal to adopt Shannon's entropy and KL divergence as measures of the privacy of a user profile. Before we dive in, we must stress that the use of entropy as a measure of privacy, in the widest sense of the term, is by no means new. Shannon's work in the fifties introduced the concept of *equivocation* as the conditional entropy of a private message given an observed cryptogram [47], later used in the formulation of the problem of the wiretap channel [48,49] as a measure of confidentiality. More recent studies [50,51] rescue the suitable applicability of the concept of entropy as a measure of privacy, by proposing to measure the degree of anonymity observable by an attacker as the entropy of the probability distribution of possible senders of a given message. More recent work has taken initial steps in relating privacy to information-theoretic quantities [23,18,52,15,28,29].

In the context of this paper, an intuitive justification in favor of entropy maximization is that it boils down to making the apparent user profile as uniform as possible, thereby hiding a user's particular bias towards certain categories of interest. But a much richer argumentation stems from Jaynes' rationale behind entropy maximization methods [41,53], more generally understood under the beautiful perspective of the method of types and large deviation theory [42, Section 11], which we motivated and reviewed in the previous subsection.

Under Jaynes' rationale on entropy maximization methods, the entropy of an apparent user profile, modeled by a relative frequency histogram of categorized user data (e.g., queries, ratings or tags) may be regarded as a measure of privacy, or perhaps more accurately, anonymity. The leading idea is that the method of types from information theory establishes an approximate monotonic relationship between the likelihood of a PMF in a stochastic system and its entropy. Loosely speaking and in our context, the higher the entropy of a profile, the more likely it is, and the more users behave according to it. Under this interpretation, entropy is a measure of anonymity, *not* in the sense that the user's identity remains unknown, but only in the sense that higher likelihood of an apparent profile, believed by an external observer to be the actual profile, makes that profile more common, hopefully helping the user go unnoticed, less interesting to an attacker whose objective is to target peculiar users. This is of course in the absence of a probability distribution model for the PMFs, viewed abstractly as r.v.'s themselves; if available, that distribution of profiles would be the measure of anonymity to be used, in the same sense of user profile density.

If an aggregated histogram of the population were available as a reference profile, the extension of Jaynes' argument to relative entropy, that is, to the KL divergence, would also give an acceptable measure of anonymity. Recall from Section 4.1 that KL divergence is a measure of discrepancy between probability distributions, which includes Shannon's entropy as the special case when the reference distribution is uniform. Conceptually, a lower KL divergence hides discrepancies with respect to a reference profile, say the population's, and there also exists a monotonic relationship between the likelihood of a distribution and its divergence with respect to the reference distribution of choice, which enables us to deem KL divergence as a measure of anonymity in a sense entirely analogous to the above mentioned. To sum up, our justification of divergence as a measure of anonymity builds upon these two fundamental ideas:

- user profile density may be regarded as a measure of anonymity;
- through Jaynes' rationale, KL divergence may be interpreted as a measure of user profile density.

Fig. 6 illustrates these ideas by means of a simple but insightful example. The figure in question shows a distribution of profiles in the probability simplex, in the case when profiles are modeled across $n = 3$ categories of interest, e.g., business, technology and sports. Note, however, that the justification of divergence provided in this section presumes that this information is not at the disposal of users. If available, users would certainly use it as a measure of anonymity. In this figure, we also represent the actual profile of a particular user, their apparent profile, and the average population's profile. Besides, we plot the contours of the divergence between a point in the simplex and the reference distribution $\bar{t}$, that is, $D(\cdot \parallel \bar{t})$. Under Jaynes' rationale, this particular user perturbs their actual profile in such a way that the resulting profile approaches, in terms of KL divergence, the population's profile. In doing so, the apparent profile gets lost in the crowd, thus hindering privacy attackers in their efforts to distinguish this user from other users.

Last but not least, we would like to emphasize that, under the assumptions this justification relies on, i.e., an adversary aimed at discriminating a given user from the population of users, KL divergence is, in fact, a measure of privacy *risk* or, more accurately,

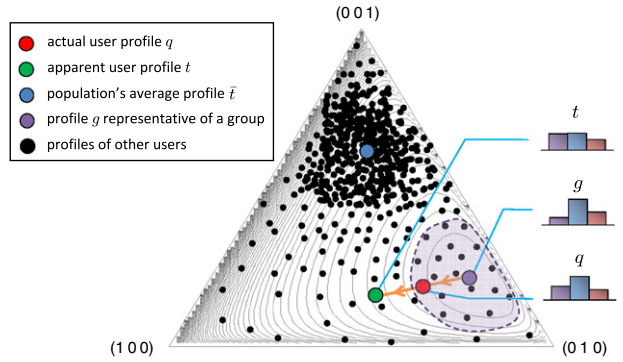

Fig. 7. A user distorts their actual profile q to counter an attacker who strives to classify this user as belonging to a particular group. Under our interpretation of divergence through hypothesis testing, the probability of being classified as a member of that group decreases as t moves away, in terms of divergence, from the profile g representative of said group.

anonymity *loss*. This contrasts with the interpretation given in Section 5, where the assumption of an attacker operating as a classifier leads us to consider KL divergence as a measure of privacy *gain*.

5. Measuring privacy against classification

In Section 4, we interpreted KL divergence and Shannon's entropy as privacy criteria, under the assumption that the attacker attempted to target users who deviated from the average profile. In this section, however, we justify our metric under the premise that the attacker strives to classify a particular user into a predefined group. Put differently, the attacker's objective boils down to a classification problem. The justification provided in this section corresponds to the branch called *classification* in the tree of Fig. 5.

The use of KL divergence as a classifier is justified by its extensive application in the fields of speech and image recognition, machine learning, data mining, and also in information security [54–60]. Just as an illustrative example, in the context of grid computing, [60] makes use of KL divergence to classify an incoming traffic of packets as a denial-of-service attack traffic. Nonetheless, a more elaborated and rich justification in favor of KL divergence as a classifier stems from hypothesis testing [42, Section 11] and the method of types of large deviation theory. More precisely, we shall interpret our privacy metric as false positives and negatives when an attacker applies a binary hypothesis test to find out whether a sequence of observed data (e.g., ratings, tags, or queries) belongs to a predefined group of users or not.

Let H be a binary r.v. representing two possible hypothesis about the distribution of an r.v. X. Precisely, $H = 1$ with probability θ and $H = 2$ with probability $1 - \theta$, and X conditioned on H has PMF g when $H = 1$ and g' when $H = 2$. Let $\left(X_j\right)_{j=1}^{k}$ be k i.i.d. drawings of this reference r.v. X and let t denote the type or empirical distribution of a k-tuple of their observed values $\left(x_j\right)_{j=1}^{k}$. Recall that the maximum a posteriori (MAP) of a finite-alphabet r.v. is its most likely value. It can be shown [42] that:

(i) The log-likelihood

$$-\frac{1}{k}\log P\left\{\left(X_j\right)_{j=1}^{k} = \left(x_j\right)_{j=1}^{k}\middle| H\right\} = \begin{cases} H(t \parallel g) & \text{if } H = 1. \\ H(t \parallel g') & \text{if } H = 2. \end{cases}$$

(ii) The MAP estimate $\hat{H}_{\text{MAP}}$ of the hypothesis H from the observed sequence $\left(X_j\right)_j$ is determined by the Neyman–Pearson criterion, namely $\hat{H}_{\text{MAP}} = 1$ if, and only if,

$$D(t \parallel g) \leqslant D(t \parallel g') + \gamma, \tag{2}$$

with $\gamma = \frac{1}{k}\log\frac{\theta}{1-\theta}$, and $\hat{H}_{MAP} = 2$ otherwise.

Even if the prior probability θ is unknown or if the hypothesis is not modeled as an r.v., for any $\gamma \in \mathbb{R}$, criterion (ii) still optimizes the trade-off between the probabilities of false positives and false negatives, in the sense that one of these errors is minimized for a fixed value of the other. In short, γ parametrizes the trade-off curve in the error plane.

Our interpretation contemplates the scenario where an attacker knows, or is able to estimate, both the apparent distribution t of a given user, and the distribution g representing a group into which this user does *not* want to be categorized. Accordingly, the attacker observes a sequence of k i.i.d. queries, and attempts to ascertain whether they belong to a member of that group. More accurately, the attacker considers the *hypothesis testing* between two alternatives, namely whether the data have been drawn according to g, hypothesis $\mathscr{H}_1$, or g', hypothesis $\mathscr{H}_2$, where g' may represent the complement of the sensitive group at hand, or any other group.

Define the *acceptance region* $\mathscr{A}_k$ as the set of sequences of observed data over which the attacker decides to accept $\mathscr{H}_1$. Concordantly, consider the following two probabilities of decision error:

(a) the probability of a false negative $\alpha_k = g(\bar{\mathscr{A}_k})$, defined as the probability of accepting $\mathscr{H}_2$ when $\mathscr{H}_1$ is true,
(b) and the probability of a false positive $\beta_k = g'(\mathscr{A}_k)$, defined as the probability of accepting $\mathscr{H}_1$ when $\mathscr{H}_2$ is true.

Above, $\bar{\mathscr{A}_k}$ denotes the complement of $\mathscr{A}_k$. $g(\mathscr{A}_k)$, for example, represents the probability of all data sequences in $\mathscr{A}_k$, i.i.d. according to g, and similarly for $g'(\bar{\mathscr{A}_k})$. Hence, α_k is the probability that the attacker mistakenly classifies the user as not belonging to the group, and β_k the probability of the attacker incorrectly assuming that the user does belong to it.

According to the preliminaries in this section, an intelligent attacker would perform a Neyman–Pearson test (2) to infer whether the user belongs in fact to the group, in an optimal fashion, that is, minimizing the classification error α_k for a given error β_k, or vice versa. In the event that a suitable representation g' of the alternative group is unavailable, or that a simpler approach is deemed preferable, the user shall strive to counter such an intelligent attacker by merely maximizing the discrepancy $D(t \parallel g)$ between the apparent profile t and the representation g of the sensitive group to avoid.

Fig. 7 provides an example that illustrates our justification of divergence as a measure of privacy against classification. Particularly, this figure plots a distribution model for profiles in the simplex of probability, under the assumption that user profiles are represented across $n = 3$ categories of interest, exactly as in Fig. 6. We also depict the actual profile q of a particular user, their apparent profile t and the profile g representative of a group into which this user does not want to be classified. The contours correspond to the divergence $D(\cdot \parallel g)$ between a point in the simplex and the group profile g. The figure in question also shows the region of the simplex that leads the attacker to classify a user as belonging to this particular group.

Last but not least, we would like to stress that the justifications provided in this section are clearly under the premise that the user knows the distribution g. An alternative to the absence of this information is assuming $g = q$, that is, considering the user as the group into which they do not want to be classified. Building on this assumption, the user's strategy consists in maximizing $D(t \parallel q)$. Conceptually, this reflects the situation in which a user does not want the perturbed, observed profile resemble their actual profile. The resulting privacy measure, i.e., the divergence between the apparent user profile and the actual user profile, is in line with other criteria in the literature that suggest measuring privacy by using some measure of similarity or distance between these

two profiles [19–21]. Fig. 5 illustrates the assumptions about the adversary model and the information-theoretic arguments that we have followed to justify and interpret KL divergence and Shannon's as privacy criteria.

6. Related work

In this section we give a broad overview of the most relevant privacy criteria in the literature. In the context of SDC, a *microdata set* is defined as a database table whose records carry information concerning individual respondents. Specifically, this set contains key attributes, that is, attributes that, in combination, may be linked with external information to reidentify the respondents to whom the records in the microdata set refer. Examples include job, address, age and gender, height and weight. In addition, the data set contains confidential attributes with sensitive information on the respondent, such as health, salary and religion.

A common approach in SDC is microaggregation, which consists in clustering the data set into groups of records with similar tuples of key attributes values, and replacing these tuples in every record within each group by a representative group tuple. One of the most popular privacy criteria in database anonymization is k-anonymity [11,12], which can be achieved through the aforementioned microaggregation procedure. This criterion requires that each combination of key attribute values be shared by at least k records in the microdata set. However, the problem of k-anonymity, and of enhancements [61,13,14,62] such as l-diversity, is their vulnerability against skewness and similarity attacks [63].

In order to overcome these deficiencies, yet another privacy criterion was considered in [15]: a dataset is said to satisfy t-closeness if for each group of records sharing a combination of key attributes, a certain measure of divergence between the within-group distribution of confidential attributes and the distribution of those attributes for the entire dataset does not exceed a threshold t. It is worth mentioning that the underlying principle of this criterion, that is, capturing the incremental gain in the adversary's knowledge, was originally introduced in [64]. An average-case version of the worst-case t-closeness criterion, using the Kullback–Leibler divergence as a measure of discrepancy, turns out to be equivalent to a mutual information, and lends itself to a generalization of Shannon's rate-distortion problem [52,18]. A simpler information-theoretic privacy criterion, not directly evolved from k-anonymity, consists in measuring the degree of anonymity observable by an attacker as the entropy of the probability distribution of possible senders of a given message [50,51]. Lastly, [17] analyzes privacy for interactive databases, where a randomized perturbation rule is applied to a true answer to a query, before returning it to the user. Consider two databases that differ only by one record, but are subject to a common perturbation rule. Conceptually, the randomized perturbation rule is said to satisfy the ϵ-*differential privacy* criterion if the two corresponding probability distributions of the perturbed answers are similar, according to a certain inequality.

With regard to metrics for user profiles, [19,20] propose a mechanism aimed to preserve the privacy of a group of users sharing an access point to the Web while surfing the Internet. Specifically, the authors suggest generating fake transactions, i.e., accesses to a web page to hinder eavesdroppers in their efforts to profile the group. Accordingly, privacy is measured as the similarity between the genuine profile of the group and that observed from the outside. More accurately, they use cosine measure, as frequently done in information retrieval [65], to capture the similarity between the group genuine profile and the group apparent profile. Another metric for profiles is [21], which quantifies privacy as a weighted version of the Euclidean distance between the genuine profile and that observed by the recommender. In addition, [22] measures privacy as the Shannon entropy of the exposed profile, normalized by the entropy of the genuine profile.

In [23], we formulate the trade-off between privacy and utility in query forgery for online search as an optimization problem, also measuring privacy as a KL divergence, and measuring utility as the ratio of forged queries to total queries. In the scenario of the semantic Web, we investigate tag suppression as a privacy-enhancing technology [28,29]. Privacy is measured as the entropy of the observed profile resulting from the elimination of certain tags, and utility is quantified as the degradation in semantic functionality and accuracy in parental control filtering.

Finally, we would like to note that the metrics for profiles mentioned above are certainly connected to the adversary models assumed in this work. On the one hand, [19–21] may be understood under the perspective of an attacker who strives to classify users, and under the assumption that these users do not know the profile representative of the group they do not wish to be categorized into. And on the other hand, [22,23,28,29] may be interpreted in the special case when the attacker's objective is to identify users.

7. Concluding remarks

Recent years have witnessed the accelerated growth of a rich variety of information–communication technologies of unprecedented sophistication, which endeavors to tailor information-exchange functionality to the specific interests of their users. Examples of these technologies include personalized Web search, resource tagging in the semantic Web, and multimedia recommendation systems. Most of them build upon, or lend themselves to, the creation of user profiles, which, by themselves but especially when combined across several information services, pose evident privacy and security risks.

Numerous privacy-enhancing technologies have been proposed to mitigate those risks. Unfortunately, these technologies have not yet gained wide adoption, because, frequently, their effectiveness remains unclear as well as their penalties in terms of utility. In this state of affairs, privacy metrics, together with utility metrics, help pave the way for their adoption, as the only manner to evaluate and compare them.

The literature of statistical disclosure control abounds with examples of privacy metrics. Although some of them might be applied to the motivating scenario of this work, the fact is that there are few proposals specifically conceived for measuring the privacy of user profiles and, what is more important, they are not appropriately justified.

To the best of our knowledge, our work is the first to rigorously justify a measure of the privacy of user profiles. The proposed metric is KL divergence, an information-theoretic quantity that we interpret under two distinct adversary models. First, we consider an attacker that strives to target users who deviate from the average profile of interests; and secondly, we contemplate an attacker whose objective is to classify a given user into a predefined group of users.

For the former model, the use of KL divergence is justified by elaborating on Jaynes' rationale behind entropy-maximization methods and the method of types of large deviation theory. Under this interpretation, divergence is a measure of privacy risk, or more accurately, anonymity loss. Only in this case, the uniform profile is of particular interest, as it translates into an entropy-maximization problem.

For the latter adversary model, our privacy criterion is supported by its extensive use in fields such as speech and image recognition, machine learning, data mining and information security. But a richer argument stems from hypothesis testing and

the method of types, which enables us to interpret our criterion as false positives and negatives. Under this perspective, divergence is a measure of privacy gain.

In a nutshell, in this paper we have accomplished the following goals:

- Jaynes' rationale for entropy maximization has been applied to other scientific areas – for example spectral estimation – to both justify and interpret a variety of models and algorithms—continuing with the example of spectral estimation, Burg's method. The application of the same celebrated rationale and its extension to relative entropy, now to the field of information privacy, is one of the novel, exciting contributions of this work. An analogous application is made for the method of hypothesis testing in the field of statistics.
- By doing so, we argue in favor of the use of KL divergence to measure profile privacy, along with conceptual insight into its information-theoretic, statistical meaning.
- Further, we introduce completely new models tying up the notions of profiling and profile classification with information theory.
- The value of these contributions stems from the fact that drawing a connection between information theory and information privacy, at the level of privacy metrics in mathematical modeling of privacy-enhancing mechanisms, opens the door for further application of powerful, mature concepts form the former field to the latter, and transitively, fields related to the former such as data compression and convex optimization, as we illustrate here with concrete examples.

Acknowledgments

We would like to thank the anonymous referees for their helpful comments. This work was partly supported by the Spanish Government through projects Consolider Ingenio 2010 CSD2007-00004 "ARES", TEC2010-20572-C02-02 "Consequence" and by the Government of Catalonia under grant 2009 SGR 1362.

References

[1] G. Linden, B. Smith, J. York, Amazon.com recommendations: item-to-item collaborative filtering, IEEE Internet Comput. Mag. 7 (1) (2003) 76–80.
[2] E. Protalinski, Facebook to double revenue to $4.27 billion, 89% is from ads (Dec. 2011), URL http://www.zdnet.com/blog/facebook/facebook-to-double-revenue-to-427-billion-89-is-from-ads/3877.
[3] Facebook for business, URL http://www.facebook.com/business/ads.
[4] D. Chaum, Untraceable electronic mail, return addresses, and digital pseudonyms, Commun. ACM 24 (2) (1981) 84–88.
[5] R. Dingledine, N. Mathewson, P. Syverson, Tor: the second-generation onion router, in: Proc. Conf. USENIX Secur. Symp., Berkeley, CA, 2004, p. 21.
[6] V. Benjumea, J. López, J.M.T. Linero, Specification of a framework for the anonymous use of privileges, Telemat. Inform. 23 (3) (2006) 179–195.
[7] G. Fuchsbauer, D. Pointcheval, D. Vergnaud, Transferable constant-size fair e-cash, in: Proc. Cryptology, Netw. Secur., CNS, Springer-Verlag, 2009, pp. 226–247.
[8] P. Bogetoft, D.L. Christensen, I. Damgård, M. Geisler, T. Jakobsen, M. Krøigaard, J.D. Nielsen, J.B. Nielsen, K. Nielsen, J. Pagter, M. Schwartzbach, T. Toft, Financial Cryptography and Data Security, Springer-Verlag, 2009, pp. 325–343 (Chapter Secure Multiparty Computation Goes Live).
[9] A. Rial, B. Preneel, Optimistic fair priced oblivious transfer, in: Proc. Int. Conf. Cryptology Africa, AFRICACRYPT, Springer-Verlag, 2010, pp. 131–147.
[10] J. Borking, Why adopting privacy enhancing technologies (pets) takes so much time, in: S. Gutwirth, Y. Poullet, P. Hert, R. Leenes (Eds.), Proc. Comput. Priv. Data Prot., CPD, Springer-Verlag, 2011, pp. 309–341.
[11] L. Sweeney, *k*-anonymity: a model for protecting privacy, Int. J. Uncertain., Fuzz., Knowl.-Based Syst. 10 (5) (2002) 557–570.
[12] P. Samarati, Protecting respondents' identities in microdata release, IEEE Trans. Knowl. Data Eng. 13 (6) (2001) 1010–1027.
[13] T.M. Truta, B. Vinay, Privacy protection: *p*-sensitive *k*-anonymity property, in: Proc. Int. Workshop Priv. Data Manage., PDM, Atlanta, GA, 2006, p. 94.
[14] A. Machanavajjhala, J. Gehrke, D. Kiefer, M. Venkitasubramanian, *l*-diversity: privacy beyond *k*-anonymity, in: Proc. IEEE Int. Conf. Data Eng., ICDE, Atlanta, GA, 2006, p. 24.
[15] N. Li, T. Li, S. Venkatasubramanian, *t*-closeness: privacy beyond *k*-anonymity and *l*-diversity, in: Proc. IEEE Int. Conf. Data Eng., ICDE, Istanbul, Turkey, 2007, pp. 106–115.
[16] J. Brickell, V. Shmatikov, The cost of privacy: destruction of data-mining utility in anonymized data publishing, in:Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min., KDD, Las Vegas, NV, 2008, pp. 70–78.
[17] C. Dwork, Differential privacy, in: Proc. Int. Colloq. Automata, Lang. Program., Springer-Verlag, 2006, pp. 1–12.
[18] D. Rebollo-Monedero, J. Forné, J. Domingo-Ferrer, From *t*-closeness-like privacy to postrandomization via information theory, IEEE Trans. Knowl. Data Eng. 22 (11) (2010) 1623–1636. URL http://doi.ieeecomputersociety.org/10.1109/TKDE.2009.190.
[19] Y. Elovici, B. Shapira, A. Maschiach, A new privacy model for hiding group interests while accessing the Web, in: Proc. Workshop Priv. Electron. Society, ACM, Washington, DC, 2002, pp. 63–70.
[20] B. Shapira, Y. Elovici, A. Meshiach, T. Kuflik, PRAW—the model for PRivAte Web, J. Amer. Soc. Inform. Sci., Technol. 56 (2) (2005) 159–172.
[21] M. Halkidi, I. Koutsopoulos, A game theoretic framework for data privacy preservation in recommender systems, in: Proc. European Mach. Learn., Prin. Pract. Knowl. Disc. Databases, ECML PKDD, Springer-Verlag, 2011, pp. 629–644.
[22] Y. Xu, K. Wang, B. Zhang, Z. Chen, Privacy-enhancing personalized Web search, in: Proc. Int. WWW Conf., ACM, 2007, pp. 591–600.
[23] D. Rebollo-Monedero, J. Forné, Optimal query forgery for private information retrieval, IEEE Trans. Inform. Theory 56 (9) (2010) 4631–4642.
[24] J. Parra-Arnau, D. Rebollo-Monedero, J. Forné, A privacy-preserving architecture for the semantic Web based on tag suppression, in: Proc. Int. Conf. Trust, Priv. Secur. Digit. Bus., TRUSTBUS, Bilbao, Spain, 2010, pp. 58–68.
[25] J. Parra-Arnau, D. Rebollo-Monedero, J. Forné, A privacy-protecting architecture for collaborative filtering via forgery and suppression of ratings, in: Proc. Int. Workshop Data Priv. Manage., Auton. Spontaneus Secur., DPM, Leuven, Belgium, 2011, pp. 42–57.
[26] D. Rebollo-Monedero, J. Forné, J. Domingo-Ferrer, Coprivate query profile obfuscation by means of optimal query exchange between users, IEEE Trans. Depend. Secure Comput. 9 (5) (2012) 641–654. URL http://doi.ieeecomputersociety.org/10.1109/TDSC.2012.16.
[27] E. Balsa, C. Troncoso, C. Díaz, OB-PWS: obfuscation-based private Web search, in: Proc. IEEE Symp. Secur. Priv., SP, IEEE Comput. Soc., 2012, pp. 491–505.
[28] J. Parra-Arnau, D. Rebollo-Monedero, J. Forné, J.L. Muñoz, O. Esparza, Optimal tag suppression for privacy protection in the semantic Web data, Knowl. Eng. 81–82 (0) (2012) 46–66. URL http://dx.doi.org/10.1016/j.datak.2012.07.004.
[29] J. Parra-Arnau, A. Perego, E. Ferrari, J. Forné, D. Rebollo-Monedero, Privacy-preserving enhanced collaborative tagging, IEEE Trans. Knowl. Data Eng. (2013). URL http://dx.doi.org/10.1109/TKDE.2012.248.
[30] D. Rebollo-Monedero, J. Parra-Arnau, J. Forné, An information-theoretic privacy criterion for query forgery in information retrieval, in: Proc. Int. Conf. Secur. Technol., SecTech, in: Lecture Notes Comput. Sci. (LNCS), Springer-Verlag, Jeju Island, South Korea, 2011, pp. 146–154. invited paper.
[31] M.J. Pazzani, D. Billsus, Content-based recommendation systems, in: P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), The Adaptive Web, Springer-Verlag, 2007, pp. 325–341.
[32] J. Liu, P. Dolan, E.R. Pedersen, Personalized news recommendation based on click behavior, in: Proc. Int. Conf. Intell. User Interf., IUI, ACM, 2010, pp. 31–40.
[33] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, S. Barocas, Adnostic: privacy preserving targeted advertising, in: Proc. IEEE Symp. Netw. Distrib. Syst. Secur., SNDSS, 2010, pp. 1–21.
[34] M. Fredrikson, B. Livshits, RePriv: re-envisioning in-browser privacy, in: Proc. IEEE Symp. Secur. Priv., SP, 2011, pp. 131–146.
[35] T. Kuflik, B. Shapira, Y. Elovici, A. Maschiach, Privacy preservation improvement by learning optimal profile generation rate, in: User Modeling, in: Lecture Notes Comput. Sci. (LNCS), vol. 2702, Springer-Verlag, 2003, pp. 168–177.
[36] Y. Elovici, C. Glezer, B. Shapira, Enhancing customer privacy while searching for products and services on the World Wide Web, Internet Res. 15 (4) (2005) 378–399.
[37] M.K. Reiter, A.D. Rubin, Crowds: anonymity for Web transactions, ACM Trans. Inform. Syst. Secur. 1 (1) (1998) 66–92.
[38] D. Rebollo-Monedero, J. Forné, A. Solanas, T. Martínez-Ballesté, Private location-based information retrieval through user collaboration, Comput. Commun. 33 (6) (2010) 762–774. URL http://dx.doi.org/10.1016/j.comcom.2009.11.024.
[39] L. Cottrell, Mixmaster and remailer attacks, 1994, URL http://obscura.com/~loki/remailer/remailer-essay.html.
[40] G. Danezis, R. Dingledine, N. Mathewson, Mixminion: design of a type III anonymous remailer protocol, in: Proc. IEEE Symp. Secur. Priv., SP, Berkeley, CA, 2003, pp. 2–15.
[41] E.T. Jaynes, On the rationale of maximum-entropy methods, Proc. IEEE 70 (9) (1982) 939–952.
[42] T.M. Cover, J.A. Thomas, Elements of Information Theory, second ed., Wiley, New York, 2006.
[43] L. Brillouin, Science and Information Theory, Academic-Press, New York, 1962.
[44] E.T. Jaynes, Papers on Probability, Statistics and Statistical Physics, Reidel, Dordrecht, 1982.
[45] J.P. Burg, Maximum entropy spectral analysis, Ph.D. Thesis, Stanford Univ., 1975.
[46] A.L. Berger, J. della Pietra, A. della Pietra, A maximum entropy approach to natural language processing, MIT Comput. Ling. 22 (1) (1996) 39–71.

[47] C.E. Shannon, Communication theory of secrecy systems, Tech. J. Bell Syst. (1949).
[48] A. Wyner, The wiretap channel, Tech. J. Bell Syst. 54 (1975).
[49] I. Csiszár, J. Körner, Broadcast channels with confidential messages, IEEE Trans. Inform. Theory 24 (1978) 339–348.
[50] C. Díaz, S. Seys, J. Claessens, B. Preneel, Towards measuring anonymity, in: Proc. Workshop Priv. Enhanc. Technol., PET, in: Lecture Notes Comput. Sci. (LNCS), vol. 2482, Springer-Verlag, 2002, pp. 54–68.
[51] C. Díaz, Anonymity and privacy in electronic services, Ph.D. Thesis, Katholieke Univ. Leuven, Dec. 2005.
[52] D. Rebollo-Monedero, J. Forné, J. Domingo-Ferrer, From t-closeness to PRAM and noise addition via information theory, in: Priv. Stat. Databases, PSD, in: Lecture Notes Comput. Sci. (LNCS), Springer-Verlag, Istambul, Turkey, 2008, pp. 100–112.
[53] E.T. Jaynes, Information theory and statistical mechanics II, Phys. Review Ser. II 108 (2) (1957) 171–190.
[54] P.A. Olsen, S. Dharanipragada, An efficient integrated gender detection scheme and time mediated averaging of gender dependent acoustic models, in: Proc. European Conf. Speech Commun. Technol., EUROSPEECH, 2003, pp. 2509–2512.
[55] H. Printz, P. Olsen, Theory and practice of acoustic confusability, Comput. Speech Lang. 16 (1) (2002) 131–164.
[56] J. Silva, S. Narayanan, Average divergence distance as a statistical discrimination measure for hidden Markov models, IEEE Trans. Audio, Speech, Lang. Process. 14 (3) (2006) 890–906.
[57] Q. Huo, W. Li, A dtw-based dissimilarity measure for left-to-right hidden Markov models and its application to word confusability analysis, in: Proc. Interspeech, 2006, pp. 2338–2341.
[58] J. Goldberger, S. Gordon, H. Greenspan, An efficient image similarity measure based on approximations of kl-divergence between two Gaussian mixtures, in: Proc. Int. Conf. Comput. Vision, ICCV, 2003, pp. 487–493.
[59] D. Olszewski, Fraud detection in telecommunications using Kullback–Leibler divergence and latent Dirichlet allocation, in: Proc. Adap. Nat. Comput. Alg., Springer-Verlag, 2011, pp. 71–80.
[60] Q. Mei, C. Zhai, Discovering evolutionary theme patterns from text: an exploration of temporal text mining, in: Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min., KDD, ACM, 2005, pp. 198–207.
[61] X. Sun, H. Wang, J. Li, T.M. Truta, Enhanced p-sensitive k-anonymity models for privacy preserving data publishing, Trans. Data Priv. 1 (2) (2008) 53–66.
[62] H. Jian-min, C. Ting-ting, Y. Hui-qun, An improved V-MDAV algorithm for l-diversity, in: Proc. IEEE Int. Symp. Inform. Process., ISIP, Moscow, Russia, 2008, pp. 733–739.
[63] J. Domingo-Ferrer, V. Torra, A critique of k-anonymity and some of its enhancements, in: Proc. Workshop Priv. Secur. Artif. Intell., PSAI, Barcelona, Spain, 2008, pp. 990–993.
[64] Y.T. Chiang, Y.C. Chiang, T.S. Hsu, C.J. Liau, D.W. Wang, How much privacy?—a system to safe guard personal privacy while releasing databases, in: Proc. Int. Conf. Rough Sets, Curr. Trends Comput., RSCTC, in: Lecture Notes in Artif. Intell. (LNAI), vol. 2475, Springer-Verlag, 2002, pp. 226–233.
[65] W.B. Frakes, R.A. Baeza-Yates (Eds.), Information Retrieval: Data Structures & Algorithms, Prentice-Hall, 1992.

Javier Parra-Arnau was awarded the M.S. degree in Electrical Engineering by the Universitat Politècnica de Catalunya (UPC) in 2004. After finishing his degree, he gained a position as a project engineer in the communications department of an important Spanish engineering company. Four years later he joined the Information Security Group in the Department of Telematics Engineering at the UPC and continued to further develop his training. He was awarded the M.S. degree in Telematics Engineering in 2009 and decided to engage in research. He is currently a Ph.D. candidate at UPC, where he investigates information-theoretic criteria of anonymity and privacy, and mathematical models of the privacy–utility trade-off in data perturbative mechanisms.

David Rebollo-Monedero received the M.S. and Ph.D degrees in Electrical Engineering from Stanford University, in California, USA, in 2003 and 2007, respectively. His doctoral research at Stanford focused on data compression, more specifically, quantization and transforms for distributed source coding. Previously, he was an information technology consultant for PricewaterhouseCoopers, in Barcelona, Spain, from 1997 to 2000, and was involved in the Retevisión startup venture. During the summer of 2003, still as a Ph.D student at Stanford, he worked for Apple Computers with the QuickTime video codec team in California, USA. He is currently a postdoctoral researcher with the Information Security Group, in the Department of Telematics of the Universitat Politècnica de Catalunya (UPC), also in Barcelona, where he investigates the application of data compression formalisms to privacy in information systems.

Jordi Forné received the M.S. degree in Telecommunications Engineering from the Universitat Politècnica de Catalunya (UPC) in 1992, and the Ph.D. degree in 1997. In 1991, he joined the Cryptography and Network Security Group, in the Department of Applied Mathematics and Telematics. Currently, he is an associate professor of the Telecommunications Engineering School of Barcelona (ETSETB), and works with the Information Security Group, both affiliated to the Department of Telematics Engineering of UPC in Barcelona. He is a coordinator of the Ph.D. program on Telematics Engineering (holding a Spanish Quality Mention) and Director of the research Master in Telematics Engineering. His research interests span a number of subfields within information security and privacy, including network security, electronic commerce and public-key infrastructures. He has been a member of the program committee of a number of security conferences, and he is editor of the Computer Standards & Interfaces Journal (Elsevier).