Data & Knowledge Engineering 81–82 (2012) 46–66

Contents lists available at SciVerse ScienceDirect

Data & Knowledge Engineering

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

Optimal tag suppression for privacy protection in the semantic Web

Javier Parra-Arnau *, David Rebollo-Monedero, Jordi Forné, Jose L. Muñoz, Oscar Esparza

Dept. of Telematics Engineering, Universitat Politècnica de Catalunya, C. Jordi Girona 1-3, E-08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 10 December 2010
Received in revised form 24 July 2012
Accepted 28 July 2012
Available online 25 August 2012

Keywords:
Information privacy
Privacy-enhancing technology
Shannon entropy
Privacy–suppression trade-off
Semantic Web
Tagging systems

ABSTRACT

Leveraging on the principle of data minimization, we propose *tag suppression*, a privacy-enhancing technique for the semantic Web. In our approach, users tag resources on the Web revealing their personal preferences. However, in order to prevent privacy attackers from profiling users based on their interests, they may wish to refrain from tagging certain resources. Consequently, tag suppression protects user privacy to a certain extent, but at the cost of semantic loss incurred by suppressing tags. In a nutshell, our technique poses a trade-off between privacy and suppression. In this paper, we investigate this trade-off in a mathematically systematic fashion and provide an extensive theoretical analysis. We measure user privacy as the entropy of the user's tag distribution after the suppression of some tags. Equipped with a quantitative measure of both privacy and utility, we find a close-form solution to the problem of optimal tag suppression. Experimental results on a real-world tagging application show how our approach may contribute to privacy protection.

1. Introduction

The World Wide Web constitutes the largest repository of information in the world. Since its invention in the nineties, the form in which information is organized has evolved substantially. At the beginning, Web content was classified in directories belonging to different areas of interest, manually maintained by experts. These directories provided users with accurate information, but as the Web grew they rapidly became unmanageable. Although they are still available, they have been progressively dominated by the current search engines based on Web crawlers, which explore new or updated content in a methodic, automatic manner. However, even though search engines are able to index a large amount of Web content, they may provide irrelevant results or fail when terms are not explicitly included in Web pages. A query containing the keyword *accommodation*, for instance, would not retrieve pages with terms such as *hotel* or *apartment* not including that keyword.

Recently, a new form of conceiving the Web, called the *semantic Web* [1], has emerged to address this problem. The semantic Web, envisioned by Tim Berners-Lee in 2001, is expected to provide Web content with a conceptual structure so that information can be interpreted by machines. For this to become a reality, the semantic Web requires to explicitly associate meaning with resources on the Web. A widely spread manner to accomplish this is by means of *semantic tagging*.

One of the major benefits of associating concepts with Web pages is clearly the semantic interoperability in Web applications. In addition, tagging will allow these applications to decrease the interaction with users, to obtain some form of semantic distance between pages and to ultimately process pages whose content is nowadays only understandable by humans. In a nutshell, the semantic Web lies the foundation for a future scenario where intelligent software agents will be able to automatically book flights for us, update our medical records at our request and provide us with personalized answers to particular queries, without the hassle of exhaustive literal searches across myriads of disorganized data [2]. In the meantime, we can enjoy some instances, although limited in scope, of this new conception of the Web, namely the tagging systems that have proliferated over the last

* Corresponding author. Tel.: +34 93 401 7027.
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), jose.munoz@entel.upc.edu (J.L. Muñoz), oscar.esparza@entel.upc.edu (O. Esparza).

0169-023X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.datak.2012.07.004

years. Some examples include *BibSonomy* [3], *CiteULike* [4], *Delicious* [5] and *StumbleUpon* [6], where users add short, usually one-word descriptions of resources they find when browsing the Web.

Despite the many advantages the semantic Web is bringing to the Web community, the continuous tagging activity prompts serious privacy concerns. The tags submitted by users to semantic Web servers could be used not only by these servers but also by any privacy attacker capable of collecting this information, to extract an accurate snapshot of user interests or *user profiles* [7,8], containing sensitive information such as health-related issues, political preferences, salary or religion. This could be the case of the tagging systems mentioned above and many other applications where tags are used to build user profiles, normally in the form of some kind of histogram or tag cloud.

1.1. Hard privacy vs. soft privacy

Philosophers, scholars and jurists have endeavored to conceptualize privacy since the *right-to-be-alone* definition given by Samuel Warren and Louis Brandeis in the late nineteenth century [9]. Although many admit that this task is virtually impossible [10], the privacy research literature [11] recognizes the distinction between *hard privacy* and *soft privacy*. Hard privacy, which may be regarded as *data minimization*, relies upon the assumption that users mistrust communicating entities and thus strive to reveal as little private information as possible. On the other hand, soft privacy assumes that users entrust their private data to an entity, which is thereafter responsible for the protection of their data. In the literature, numerous attempts to protect privacy have followed the traditional method of anonymous communications [12–15], which is fundamentally built on the assumptions of soft privacy. Unfortunately, anonymous-communication systems are not completely effective [16–19], they normally come at the cost of infrastructure, and assume that users are willing to trust other parties. However, even in those cases where we could trust an entity completely, that entity could eventually be legally enforced to reveal the information they have access to [20]. A discussion on the shortcomings of the main approaches regarding hard and soft privacies is provided in Section 2, and later, in more detail, in Section 5.1.

In this paper, we present a privacy-enhancing technology (PET), namely *tag suppression*, which capitalizes on the principle of data minimization. Despite the fact that the proposed strategy and the state of the art of anonymous-communication systems rely upon different assumptions, we would like to emphasize that both alternatives are not mutually exclusive and, more importantly, that users could benefit from the synergy of our approach and other systems providing soft privacy. As a matter of fact, there are examples in the literature in which techniques providing hard privacy may complement anonymous-communication systems perfectly. One example of this could be the use of dummy messages in combination with the traditional mix networks proposed in Refs. [12,21]. The study of the impact of a possible application of tag suppression in other privacy-protecting systems is out of the scope of the present work.

1.2. Contribution and plan of this paper

In this paper, we propose a privacy-enhancing mechanism that has the purpose of hindering privacy attackers in their efforts to profile users on the basis of the tags they specify. In our approach, users tag resources on the Web revealing their personal preferences. However, in order to avoid being accurately profiled, they may wish to refrain from tagging some of those resources. Consequently, tag suppression protects user privacy to a certain degree without having to trust an external entity, but at the cost of some processing overhead and, more importantly, the semantic loss incurred by suppressing tags.

The theoretical analysis of the inherent trade-off between privacy and suppression is precisely the main object of this paper. Specifically, we present a mathematical formulation of optimal tag suppression in the semantic Web. We propose an information-theoretic criterion to measure user privacy, namely the Shannon entropy of the user's tag distribution after the suppression of certain tags, and justify it with the rationale behind entropy-maximization methods. Accordingly, we formulate and solve an optimization problem modeling the privacy–suppression trade-off. The theoretical analysis presented here may also be applied to a wide range of user-generated data, rather than semantic tags. For example, one may conceive the suppression of queries in information retrieval or the elimination of ratings in the domain of recommendation systems.

In addition, we experimentally evaluate the extent to which our technique contributes to privacy protection in a real-world tagging application. Namely, we apply tag suppression to *BibSonomy*, a popular tagging system for sharing bookmarks and publications, and show, in a series of experiments, how our approach enables its users to enhance their privacy.

Section 2 explores some relevant approaches related to privacy and the semantic Web. Section 3 describes our privacy-enhancing mechanism, some considerations about the user profile model and the adversary capabilities, and ultimately a formulation of the trade-off between privacy and suppression. Section 4 presents a detailed theoretical analysis of the optimization problem characterizing the privacy-suppression trade-off. In addition, this section shows a simple but insightful example that illustrates the formulation and theoretical analysis argued in the previous sections. Section 5 compares tag suppression with other approaches providing soft privacy, and presents an experimental evaluation of our technique in *BibSonomy*. Conclusions are drawn in Section 6.

2. State of the art

A number of approaches have been suggested to preserve user privacy in the semantic Web, most of them focused on privacy policies. In the traditional Web, the majority of Web sites interact with users to provide them with privacy policies, and allowing them to find out how their private information will be managed. Unfortunately, users do not frequently understand [22] or even

read [23] privacy policies. The platform for privacy preferences (P3P) is created to deal with this situation and provides a framework with informed online interactions. More accurately, when a Web site supports the P3P, it establishes a set of policies to define how user's private information will be used. Users, in turn, set their own privacy policies to determine what kind of personal information they are willing to disclose to the Web sites they browse. Accordingly, when a user browses a Web site, P3P compares both the Web site's and the user's privacy policies. If they do not match, P3P informs the user about this situation and consequently they decide how to proceed. In the semantic Web, this process is intended to be carried out by autonomous agents. In this regard, several policy languages to define privacy and security requirements have been proposed. In Ref. [24], the authors suggest a new semantic policy language based on the resource description framework (RDF) schema to express access control requirements over concepts defined in ontologies. In Ref. [25], privacy and authentication policies are incorporated into the descriptions of an ontology called ontology Web language for services (OWL-S). Furthermore, the authors implement algorithms for the requester to verify the provider's adherence to policies.

In the scenario of collaborative tagging, Ref. [26] illustrates the privacy risks posed by those systems where users tag resources such as videos or pictures. With this aim, the authors propose an algorithm capable of inferring the geographical location of such resources, on the basis of tag descriptions as well as certain visual features. Of special interest is Ref. [27], which combines tagging with trusting methodologies to enforce privacy protection in the domain of eHealth. In this same line, Ref. [28] presents a framework that enhances social tag-based applications capitalizing on trust policies and user preferences. The authors propose a multi-layer system where users indicate the appropriateness of the tags posted by other users, and assign trust levels to users, depending on the type of relationship within the social network.

In the context of private information retrieval (PIR), users send general-purpose queries to an information service provider. In this scenario, query forgery, which consists in accompanying genuine with false queries, appears as an approach to guarantee user privacy to a certain extent at the cost of traffic and processing overhead. Precisely, in Ref. [29] we investigate the trade-off between privacy and the additional traffic overhead in a mathematically systematic fashion. Building on the simple principle of query forgery, several PIR protocols, mainly heuristic, have been proposed and implemented. In Refs. [30,31], a solution is presented, aimed to preserve the privacy of a group of users sharing an access point to the Web while surfing the Internet. The authors propose the generation of fake transactions, i.e., accesses to a Web page to hinder eavesdroppers in their efforts to profile the group. Privacy is measured as the similarity between the actual profile of a group of users and that observed by privacy attackers [30]. Specifically, the authors use the cosine measure, as frequently used in information retrieval [32], to capture the similarity between the group genuine profile and the group apparent profile. Based on this model, some experiments are conducted to study the impact of the construction of user profiles on the performance [33]. In line with this, recent surveys with a greater focus on anonymous Internet search include Refs. [34,35]. Further, some simple, heuristic implementations in the form of add-ons for popular browsers have started to appear [36,37]. More recently, Ref. [38] proposes a model that describes the building blocks of a privacy-protecting architecture relying on data perturbation. The authors suggest measuring privacy as a conditional entropy and evaluate, according to this specific metric, several approaches that propose generating false queries as an obfuscation mechanism.

Yet another strategy for anonymous tagging could be built on the principle of user collaboration, not unlike the protocols for *k*-anonymous location-based services (LBSs) [39] and for other forms of anonymity through collaboration [40,21]. Additionally, we could conceive the adoption of trusted third parties (TTPs), or even digital credentials [41–43], in order to enable anonymous and pseudonymous tagging. More specifically, we could make use of the anonymous-communication systems already introduced in Section 1.1. In this sense, a number of approaches have been proposed during the last two decades. Most of them are based on Chaum's mix networks [12], which aimed to address *traffic analysis*, i.e., the process of intercepting and examining messages in order to infer any information from patterns in communication. Essentially, a TTP called mix collects messages from a number of senders and forwards them to their intended receivers, possibly other mixes, rearranging them with the express purpose of hiding the correspondence between inputs and outputs. Messages sent to mixes are encrypted using public-key cryptography, in a layered fashion when several mixes are involved. There are several anonymous-communication proposals based on the idea of mix networks. They can be roughly classified into high-latency and low-latency systems. The former introduce significant delay to attain a high degree of anonymity against traffic analysis [44,45]. Naturally, their main drawback is that they are hardly applicable to real-time interactive tasks such as tagging, Web browsing or online chat.

In an attempt to address this limitation, low-latency systems were proposed. To attain a higher degree of anonymity, rather than increasing latency disproportionately, these systems simply benefit from the networking of a combination of several mixes frequently accessed by a significantly large population of users. Quoting [46], "All mix systems require that messages to be anonymized should be relayed through a sequence of trusted intermediate nodes". Some of these systems are based on peer-to-peer communications [47,48], under the assumption of a very large interconnected population of trusted users who know how to reach each other, and who also act as mixes. The most popular approaches are *onion routing* [13,14] and its second-generation version, TOR [15]. Onion routing uses a single data structure encrypted in a layered fashion to build an anonymous circuit. Alternatively, TOR uses an incremental path-building design, where a client who wishes to communicate with a server negotiates session keys with each successive hop in the circuit.

Lastly, we would like to remark that, despite the simplicity of query forgery, an analogous *tag forgery* would clearly not be convenient for the semantic Web, which is the motivating application of our work. Submitting a tag implies the construction of conceptual relations, a much more complex process than just sending a false query to a service provider. Therefore, users might not be willing to manually tag Web content they are not interested in. However, even though automatic mechanisms for autonomous tag forgery might be considered, they might lead to qualitative or quantitative semantic distortion. Section 5.1 examines in detail some of the approaches described in this section and compares them with our tag suppression technique.

3. Privacy protection in the semantic Web via tag suppression

Relying upon the assumptions of hard privacy, tag suppression is a PET that has the purpose of preventing privacy attackers from profiling users on the basis of the tags they specify. Conceptually, our approach protects user privacy to a certain extent, by dropping those tags that make a user profile show bias towards certain categories of interest. From a practical perspective, our tag suppression technique is conceived to be implemented as a software application running on the users' local machine. The software implementation is then responsible, on the one hand, for warning the user when their privacy is being compromised, and on the other, for helping them decide which tags should be eliminated and which should not.[1] Consequently, our approach guarantees user privacy to a certain degree without having to trust an external entity, but at the cost of some local processing overhead and, more importantly, the semantic loss incurred by suppressing tags.

In this section, first we propose a model of user profile and afterwards describe our assumptions about the adversary capabilities. Then, we justify a quantitative measure of the privacy of this profile. These considerations lead us to formulate the problem of choosing a suppression strategy as a multi-objective optimization problem that takes into account both privacy and suppression rate.

3.1. User profile model

In our scenario, a user browsing the Web assigns tags to resources of a very different nature (e.g., photos, videos, publications and bookmarks) according to their personal preferences. The user therefore contributes to categorise that content, but this is inevitably at the expense of revealing their user profile.

In Section 1, we already mentioned that tagging systems commonly represent user profiles by using some kind of histogram or tag cloud, which, in essence, are equivalent representations. Recall that a tag cloud is a visual depiction in which tags are weighted according to their frequency of use. These two possible representations for user profiles, histograms and tag clouds, are simultaneously used in popular tagging systems such as *BibSonomy*, *CiteULike* and *Delicious*. According to these examples, and as suggested in Refs. [50,49,29], we propose a first-approximation, mathematically-tractable model of user profile as a probability mass function (PMF), that is, a histogram of relative frequencies of tags within a predefined set of categories of interest. Consequently, our user profile model is in line with the representations used in numerous tagging systems to visualize the tags posted by users. An example of user profile may be found in Section 4.6. In addition, Section 5.2.3 shows the profile of real user in *BibSonomy*, a tagging system for sharing bookmarks and publications.

3.2. Adversary model

Our technique is built on the conceptually-simple principle of tag suppression. Under this principle, a user may wish to tag some resources and refrain from tagging some others to enable the resulting user profile, as observed from the outside, approach the uniform profile. We shall refer to this resulting user profile as the *apparent* user profile.

Bearing in mind this consideration and the user profile model described in Section 3.1, we assume a rudimentary adversary model in which users submitting tags are observed by a passive attacker who is able to ascertain which tags are associated to which resources. Namely, this could be the case of the semantic Web server storing the tags submitted by users, or any privacy attacker able to crawl through this information. In addition, we may contemplate the definition of the profile of a user tagging across several of these Web servers. In this case, we may also suppose that an attacker has the ability to link several profiles across different servers.

Last but not least, we also assume that the privacy attacker is unable to discern whether a particular user is adhered to the proposed privacy strategy, and therefore cannot estimate their tag suppression rate.

3.3. Privacy metric

We use an information-theoretic quantity to reflect the intuition that an attacker will be able to compromise user privacy as long as the apparent user profile diverges from the uniform profile. Namely, we measure user privacy as the Shannon entropy of the apparent user distribution. Recall [51] that the entropy of a PMF s is defined as

$$\mathrm{H}(s) = -\sum_i s_i \log_b s_i,$$

where b is the base of the logarithm used. Common values of b are 2, e and 10. In those cases, the units of entropy are *bit*, *nat* and *dit*, respectively.

In addition, recall that entropy may be interpreted as a measure of the uncertainty of the outcome of a random variable distributed according to such PMF, and that may be regarded as a special case of Kullback–Leibler (KL) divergence [51]. Concretely, the KL divergence between two probability distributions s and p, that is, $\mathrm{D}(s||p)$ is essentially equivalent to the entropy of s in the special case when p becomes the uniform distribution. According to this, we may establish a connection between our privacy criterion and the privacy metric proposed in Ref. [29], in which s represents the user's apparent distribution and p the population distribution. In other

[1] A complete description of an architecture implementing this mechanism may be found in Ref. [49].

words, the criterion in Ref. [29] is a slight generalization of the criterion proposed in this paper. However, our privacy measure significantly simplifies the architecture of a practical implementation since the population distribution, which in principle could be difficult to estimate, is not required to evaluate the user privacy. In any case, the mathematical analysis in Section 4 can be readily extended to divergence.

Another interpretation of entropy stems from the observation that a privacy attacker will have actually gained some information about a user whenever their interests are significantly concentrated on a subset of categories. In other words, a user without any apparent interest in any category hides their preferences from an attacker. A richer argument may be found in Ref. [52], where the authors establish some interesting connections between Jaynes' rationale on entropy-maximization methods [53,54] and the use of entropies and divergences as measures of privacy. The key idea is that the method of types establishes an approximate monotonic relationship between the likelihood of a PMF in a stochastic system and its entropy. Loosely speaking, the higher the entropy of a profile, the more likely it is, the more users behave similarly.

Furthermore, we would like to emphasize that, although our privacy criterion is based on a fundamental quantity in information theory, the convergence of these two fields is by no means new. As a matter of fact, Shannon's work in the fifties introduced the concept of *equivocation* as the conditional entropy of a private message given an observed cryptogram [55], later used in the formulation of the problem of the wiretap channel [56,57] as a measure of confidentiality. In addition, recent studies [58,59] reassert the suitability and applicability of the concept of entropy as a measure of privacy. Specifically, the authors propose to measure the degree of anonymity observable by an attacker as the entropy of the probability distribution of possible senders of a given message.

3.4. Formulation of the trade-off between privacy and suppression

Section 3.1 explained how certain semantic Web applications represent user profiles. In particular, we mentioned that this information is normally displayed using histograms or tag clouds. Now, we provide a more formal description of user profiles and approach the problem of tag suppression in a mathematical manner.

We model user *tags* as random variables (r.v.'s) taking on values on a common finite alphabet of categories or topics, namely the set $\{1,\ldots,n\}$ for some $n\in\mathbb{Z}^+$. This model allows us to describe user profiles by means of a PMF, leading to an equivalent representation than that used in tagging systems. Accordingly, we define q as the probability distribution of the tags of a particular *user* and $\sigma\in[0,1)$ as a *tag suppression* rate, which is the ratio of suppressed tags to total tags that the user is willing to eliminate. Concordantly, we define the user's *apparent* tag distribution s as $\frac{q-r}{1-\sigma}$ for some suppression strategy $r=(r_1,\ldots,r_n)$ satisfying $0\leq r_i\leq q_i$ and $\sum r_i=\sigma$ for $i=1,\ldots,n$. Conceptually, the user's apparent tag distribution may be interpreted as the result of, on the one hand, the suppression of certain tags from the actual user profile, that is, $q-r$, and one the other, the subsequent normalization by $\frac{1}{1-\sigma}$ so that $\sum_i s_i=1$. The information about which tags should be suppressed is encoded in the tag suppression strategy r. Namely, the component r_i is the relative frequency of tags that our mechanism suggests eliminating in the category i.

According to the justification provided in Section 3.3, we use Shannon's entropy [51] to measure user privacy. In particular, our privacy metric is the entropy of the user's apparent tag distribution s. Consistently with this measure, now we define the *privacy–suppression* function

$$\mathcal{P}(\sigma)=\max_{\substack{r\\ 0\leq r_i\leq q_i\\ \sum r_i=\sigma}} \mathrm{H}\left(\frac{q-r}{1-\sigma}\right), \tag{1}$$

which characterizes the optimal trade-off between privacy and suppression, and formally expresses the intuitive reasoning behind tag suppression: the higher the tag suppression rate σ, the higher the uncertainty in terms of the entropy of the apparent distribution, and the higher the user privacy.

For simplicity, we shall use natural logarithms throughout the paper and refer to $\log_e$ as ln, particularly because all bases produce equivalent optimization objectives.

4. Optimal tag suppression

In this section, we shall analyze the fundamental properties of the privacy–suppression function (1) defined in Section 3.4, and present a closed-form solution to the maximization problem. Our theoretical analysis only considers the case when all given probabilities are strictly positive:

$$q_i>0 \text{ for all } i=1,\ldots,n. \tag{2}$$

This assumption will be properly justified in Section 4.2. We shall suppose further, now without loss of generality, that

$$q_1\leq\cdots\leq q_n. \tag{3}$$

Before proceeding with the mathematical analysis, it is immediate from the definition of the privacy-suppression function that its initial value is $\mathcal{P}(0)=\mathrm{H}(q)$. The behavior of $\mathcal{P}(\sigma)$ for $0<\sigma<1$ is characterized by the theorems presented in this section. The notation used throughout this section is summarized in Table 1.

Table 1
Description of the variables used in our notation.

Symbol	Description
n	Number of categories of interest into which tags are classified
q	The *actual* user profile is the genuine profile of interests
r	A *suppression strategy* is an n-tuple with the percentage of tags that the user should eliminate in each category
s	The *apparent* user profile is the perturbed profile, as observed from the outside, resulting from the elimination of certain tags
u	Uniform profile across the n tag categories
H	User privacy is measured as the *Shannon's entropy* of the apparent user profile
σ	The *tag suppression rate* is the percentage of tags that the user is willing to suppress
$\mathcal{P}(\sigma)$	Function modeling the privacy–suppression trade-off
σ_{crit}	The *critical suppression* is the suppression rate beyond which the privacy–suppression function attains its maximum value or critical privacy

4.1. Monotonicity and quasiconcavity

Our first theoretical characterization, namely Lemma 4.1, investigates two elementary properties of the privacy–suppression trade-off. The lemma in question shows that the trade-off is nondecreasing and quasiconcave. The importance of these two properties is that they confirm the evidence that an optimal tag suppression strategy will never lead to a degradation in privacy protection. In other words, an increase in the tag suppression rate does not lower the entropy of the apparent profile.

Lemma 4.1. *The privacy-suppression function $\mathcal{P}(\sigma)$ is nondecreasing and quasiconcave.*

Proof. First, let $0\leq\sigma<\sigma'\leq 1$. Based on the solution r to the maximization problem corresponding to $\mathcal{P}(\sigma)$, consider the tag suppression strategy r' given by the equation

$$\frac{q-r'}{1-\sigma'}=\frac{q-r}{1-\sigma}.$$

The feasibility of r' may be checked, on the one hand, by observing that the constraints $0\leq r_i'\leq q_i$ are equivalent to $0\leq\frac{q_i-r_i'}{1-\sigma'}\leq\frac{q_i}{1-\sigma'}$ for $i=1,\ldots,n$. According to the implicit definition of r', we may rewrite these constraints as $0\leq\frac{q_i-r_i}{1-\sigma}\leq\frac{q_i}{1-\sigma'}$. Given that r is feasible, the left-hand inequality is satisfied. The right-hand inequality is also verified by simply noting that $\frac{q_i}{1-\sigma}<\frac{q_i}{1-\sigma'}$. On the other hand, it is immediate to check that $\sum_i r_i'=\sigma$.

Once we have confirmed that r' is feasible, we now turn to prove the first part of the lemma. Since the feasibility of r' does not necessarily imply that r' is a maximizer of the problem corresponding to $\mathcal{P}(\sigma')$, it follows that $\mathcal{P}(\sigma')\geq\mathrm{H}\left(\frac{q-r'}{1-\sigma'}\right)=\mathcal{P}(\sigma)$, and consequently, that the privacy-suppression function is nondecreasing.

Finally, the quasiconcavity of the privacy-suppression function is directly proved by the fact that $\mathcal{P}(\sigma)$ is a nondecreasing function of σ. □

The quasiconcavity of the privacy-suppression function (1) guarantees its continuity on the interior of its domain, namely (0,1), but it is fairly straightforward to verify, directly from the definition of $\mathcal{P}(\sigma)$ and under the positivity assumption (2), that continuity also holds at the interval endpoint 0.

4.2. Critical suppression

The following theorem will confirm the intuition that there must exist a tag suppression rate beyond which *critical privacy* is achievable, in the sense that the privacy–suppression function attains its maximum value, that is, $\mathcal{P}(\sigma)=\ln n$. Precisely, this *critical suppression* is

$$\sigma_{\text{crit}}=1-n\min_i q_i=1-nq_1$$

according to the labeling assumption (3). From the above, it is interesting to note that σ_{crit} becomes worse (closer to one) with worse (smaller) ratio $\frac{q_1}{u_1}=nq_1$.

Theorem 4.2 (Critical suppression). *Let u be the uniform distribution on $\{1,\ldots,n\}$, that is, $u_i=1/n$. For all $\sigma\in[0,1)$, if $\sigma\geq\sigma_{\text{crit}}$, then $\mathcal{P}(\sigma)=\mathrm{H}(u)=\ln n$. In addition, the optimal tag suppression strategy is $r^*=q-u(1-\sigma)$, for which the user's apparent distribution and the uniform's match. Conversely, if $\sigma<\sigma_{\text{crit}}$, then $\mathcal{P}(\sigma)<\ln n$.*

Proof. We consider only the nontrivial case when $q \neq u$, which implies that $q_1 < 1/n$ and, consequently, $\sigma_{\text{crit}} > 0$. To confirm this implication, assume $q \neq u$ and suppose now that $q_1 \geq 1/n$. Taking into account the labeling assumption (3) and the fact that q is a probability distribution in the sense that $\sum_i q_i = 1$, we arrive at the contradiction that q must be the uniform distribution. Given that $q_1 < 1/n$, it immediately follows that $\sigma_{\text{crit}} > 0$. The converse, that is, $\sigma_{\text{crit}} > 0$ implies $q \neq u$, is easily checked by noting that when $q_1 < 1/n$, q cannot be, by definition, the uniform distribution. On the other hand, the positivity assumption (2) ensures that $\sigma_{\text{crit}} < 1$.

Once we have determined the interval of values in which σ_{crit} is defined, we now proceed to confirm the feasibility of r^*. It is clear from its form that $\sum_i r_i^* = \sigma$, thus it suffices to verify that $0 \leq r_i^* \leq q_i$. First, observe that the right-hand inequality is satisfied for all i as $\sigma < 1$. Secondly, note that requiring that $r_i^* = q_i - \frac{1}{n}(1-\sigma) \geq 0$ for all i is equivalent to $\sigma \geq 1 - nq_i$, and finally to

$$\sigma \geq \max_i 1 - nq_i = 1 - n \min_i q_i,$$

as assumed in the theorem. Interestingly, observe that the expression for the critical suppression is independent of the privacy criterion assumed. To complete the first part of the proof, it is immediate to check that the proposed r^* maximizes the user privacy, since the uniform distribution maximizes entropy.

Now it remains to prove that $\mathcal{P}(\sigma) < \ln n$ when $\sigma < \sigma_{\text{crit}}$. To this end, note that the KL divergence between the user's apparent distribution and the uniform's is

$$\mathrm{D}(s\|u) = \sum_i s_i \ln \frac{s_i}{u_i} = \ln n - \mathrm{H}(s),$$

as informally argued in Section 3.3. But the information inequality [51] asserts that $\mathrm{D}(s\|u) \geq 0$, with equality if, and only if, $s = u$ for all i. Hence, when $\sigma < \sigma_{\text{crit}}$, the solution s to the optimization problem corresponding to $\mathcal{P}(\sigma)$ satisfies that $s \neq u$, and therefore $\mathcal{P}(\sigma) = \mathrm{H}(s) < \ln n$. □

After routine manipulation, we may write the optimal solution at exactly the critical suppression as

$$r_i^* = q_i - q_1,$$

equal to zero if, and only if, $q = u$. Owing to the fact that we are dealing with relative rather than absolute frequencies, it is not surprising that $r_1^* = 0$ at $\sigma = \sigma_{\text{crit}}$. More generally, in accordance with the labeling assumption (3) observe that only the first components of r^* may vanish. Fig. 1 conceptually illustrates the results stated by Lemma 4.1 and Theorem 4.2.

Before proceeding further with our theoretical analysis, we would like to remark that our assumption about the strictly positiveness of q is conveniently made, albeit not without loss of generality, to guarantee that the critical privacy is attained for a suppression $\sigma < 1$, as proved in Theorem 4.2.

4.3. Closed-form solution

Our last theorem, Theorem 4.4, will provide a closed-form solution to the maximization problem involved in the definition of the privacy–suppression function (1). This solution will be obtained from a resource allocation lemma, namely Lemma 4.3, which addresses an extension of the usual water filling problem. Even though Lemma 4.3 provides a parametric-form solution, fortunately, we shall be able to proceed towards an explicit closed-form solution, albeit piecewise.

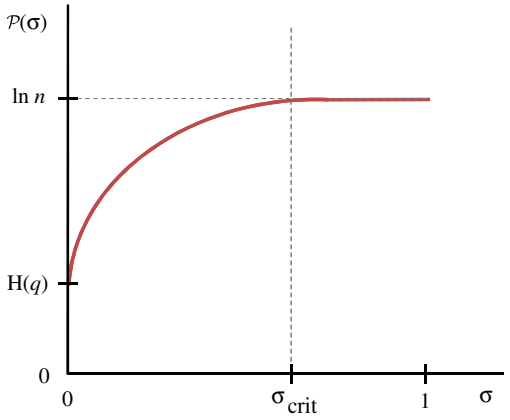


Fig. 1. Conceptual plot of the privacy–suppression function.

More specifically, this lemma considers the allocation of resources $x_1,\ldots,x_n$ minimizing the sum $\sum_i f_i(x_i)$ of convex cost functions on the individual resources. Resources are assumed to be nonnegative, upper bounded by positive thresholds b_i, and to amount to a total of $\sum_i x_i = t$, for some $t>0$. The well-known water-filling problem [60, §5.5] may be regarded as a special case when resources are not upper bounded and $f_i(x_i) = -\ln(\alpha_i + x_i)$, for $\alpha_i > 0$.

Lemma 4.3 (Resource allocation). *For all $i=1,\ldots,n$, let $f_i : [0,b_i)\to\mathbb{R}$ be twice differentiable on $[0,b_i)$, with $f_i'' > 0$, and hence strictly convex. Additionally, assume that $\lim_{x_i\to b_i^-} f_i'(x_i) = \infty$. Because $f_i'' > 0$, f_i' is strictly increasing, and, interpreted as a function from $[0,b_i)$ to $f_i'([0,b_i))$, invertible. Denote the inverse by $f_i'^{-1}$. Consider the following optimization problem in the variables $x_1,\ldots,x_n$:*

$$\begin{aligned} &\text{minimize} && \sum_{i=1}^n f_i(x_i) \\ &\text{subject to} && 0\le x_i\le b_i, \text{for all } i, \\ &&& \text{and } \sum_{i=1}^n x_i = t, \text{for some } t>0. \end{aligned}$$

i. *The solution to the problem exists, is unique and of the form $x_i^* = \max\{0, f_i'^{-1}(\nu)\}$, for some $\nu\in\mathbb{R}$ such that $\sum_i x_i^* = t$.*
ii. *Suppose further, albeit without loss of generality, that $f_n'(0)\le\cdots\le f_1'(0)$. Then, either $f_i'(0)<\nu\le f_{i-1}'(0)$ for $i=2,\ldots,n$, or $f_i'(0)<\nu$ for $i=1$, and for the corresponding index i,*

$$x_j^* = \begin{cases} f_j'^{-1}(\nu), & j = i,\ldots,n \\ 0, & j = 1,\ldots,i-1 \end{cases},$$

and

$$\sum_{j=1}^n x_j^* = \sum_{j=i}^n f_j'^{-1}(\nu) = t.$$

Proof. The existence and uniqueness of the solution is a consequence of the fact that we minimize a strictly convex function over a compact set. Systematic application of the Karush–Kuhn–Tucker (KKT) conditions [60] leads to the Lagrangian cost

$$\mathcal{L} = \sum f_i(x_i) - \sum \lambda_i x_i + \sum \mu_i(x_i - b_i) - \nu(\sum x_i - t),$$

which must satisfy $\frac{\partial \mathcal{L}}{\partial x_i} = 0$, and finally to the conditions

$$\begin{aligned} 0\le x_i\le b_i, \textstyle\sum x_i = t &\quad \text{(primal feasibility)}, \\ \lambda_i,\mu_i\ge 0 &\quad \text{(dual feasibility)}, \\ \lambda_i x_i = 0, \mu_i(x_i - b_i) = 0 &\quad \text{(complementary slackness)}, \\ f_i'(x_i) - \lambda_i + \mu_i - \nu = 0 &\quad \text{(dual optimality)}. \end{aligned}$$

Since $\lim_{x_i\to b_i^-} f_i'(x_i) = \infty$, it follows from the dual optimality condition that $x_i < b_i$. But then, the complementary slackness condition implies that $\mu_i = 0$, and consequently, we may rewrite the dual optimality condition as $f_i'(x_i) = \lambda_i + \nu$. By eliminating the slack variables λ_i, we finally obtain the simplified condition $f_i'(x_i)\ge\nu$. In addition, observe that since $f_i'(x_i) = \lambda_i + \nu$, the complementary slackness condition implies that $\left(f_i'(x_i) - \nu\right)x_i = 0$. In short, we may rewrite the dual optimality and the complementary slackness conditions equivalently as

$$\begin{aligned} f_i'(x_i)\ge\nu &\quad \text{(dual optimality)}, \\ \left(f_i'(x_i) - \nu\right)x_i = 0 &\quad \text{(complementary slackness)}. \end{aligned}$$

Now, we proceed to directly solve these equations. To this end, recall that, since $f_i'' > 0$, f_i' is strictly increasing. Consider, first, the case when $f_i'(0)\ge\nu$, or equivalently, $f_i'^{-1}(\nu)\le 0$. Suppose that $x_i > 0$, so that by complementary slackness, $f_i'(x_i) = \nu\le f_i'(0)$, contradicting the fact that f_i' is strictly increasing. Consequently, $x_i = 0$.

Consider now the opposite case, that is, when $f_i'(0) < \nu$, or equivalently $f_i'^{-1}(\nu) > 0$. In this case, the only conclusion consistent with the dual optimality condition is $x_i > 0$. But then, it follows from the complementary slackness condition that $f_i'(x_i) = \nu$, or equivalently, $x_i = f_i'^{-1}(\nu)$. This could be interpreted as a Pareto equilibrium. Specifically, for all positive resource $x_i > 0$, the

marginal ratios of improvements $f_i'(x_i)$ must all be the same. Otherwise, minor allocation adjustments on the resources could improve the overall objective. In summary,

$$x_i = \max\left\{0, f_i^{'-1}(v)\right\},$$

which proves claim (i) in the lemma.

In order to verify (ii), observe that whenever $v \leq f_{i-1}^{'}(0) \leq \cdots \leq f_1^{'}(0)$ holds for some $i=2,\ldots,n$, then $f_{i-1}^{'-1}(v),\ldots,f_1^{'-1}(v) \leq 0$, and thus $x_{i-1}=\cdots=x_1=0$. Note that the index $i=n+1$ is not permitted, since the zero solution, that is, $x_i=0$ for all $i=1,\ldots,n$, contradicts the primal feasibility condition $\sum_i x_i = t$. □

Next, we shall provide a close-form solution for the privacy-suppression function. However, before presenting the theorem in question, we shall introduce some notation. Let $\bar{Q}_i = \sum_{j=i+1}^{n} q_j$ denote the complementary cumulative distribution function. In addition, define

$$\sigma_i = \bar{Q}_i - q_i(n-i),$$

for $i=1,\ldots,n$, and, conveniently, define $\sigma_0=1$. Note that $\sigma_n=0$, that $\sigma_1 = 1-nq_1 = \sigma_{\text{crit}}$, and consistently with Theorem 4.2, the solution in this theorem at $\sigma=\sigma_{\text{crit}}$ becomes $\frac{q_j - r_j^*}{1-\sigma} = \frac{1}{n}$, for $j=1,\ldots,n$. Further, define

$$\tilde{q} = \left(q_1,\ldots,q_{i-1},\frac{\bar{Q}_{i-1}}{n-i+1},\ldots,\frac{\bar{Q}_{i-1}}{n-i+1}\right),$$
$$\tilde{r} = \left(0,\ldots,0,\frac{\sigma}{n-i+1},\ldots,\frac{\sigma}{n-i+1}\right),$$

a distribution in the probability simplex in $\mathbb{R}^n$, and an n-tuple representing a tag suppression strategy, respectively.

Theorem 4.4. *For any $i=2,\ldots,n$, $\sigma_i \leq \sigma_{i-1}$, with equality if, and only if, $q_i = q_{i-1}$. For any $i=1,\ldots,n$ and any $\sigma \in [\sigma_i, \sigma_{i-1}]$, the optimal suppression strategy is*

$$r_j^* = \begin{cases} 0 & , \quad j=1,\ldots,i-1 \\ q_j - \dfrac{\bar{Q}_{i-1}-\sigma}{n-i+1}, & j=i,\ldots,n \end{cases},$$

and, consequently, the corresponding optimal user's apparent tag distribution is

$$s_j^* = \begin{cases} \dfrac{q_j}{1-\sigma} & , \quad j=1,\ldots,i-1 \\ \dfrac{\bar{Q}_{i-1}-\sigma}{(1-\sigma)(n-i+1)}, & j=i,\ldots,n \end{cases}$$

Accordingly, the corresponding, maximum entropy yields the privacy-suppression function

$$\mathcal{P}(\sigma) = \mathrm{H}\left(\frac{\tilde{q}-\tilde{r}}{1-\sigma}\right).$$

Proof. From the definition of σ_i and under the labeling assumption (3), it is immediate to check the monotonicity of these suppression thresholds.

Now, we proceed to prove the rest of the theorem for the nontrivial case $\sigma \in (0,1)$. Using the definition of entropy, we may write the objective function in the (original) optimization problem (1) as $-\mathrm{H}(s) = \sum_i s_i \ln s_i$, with $s_i = \frac{q_i - r_i}{1-\sigma}$, since the maximization of entropy is equivalent to the minimization of negative entropy. Recall that r is optimal for the original problem if, and only if, r is optimal for the scaled problem. After this convenient, straightforward transformation, the objective function exposes the structure of the privacy-suppression optimization problem as a special case of the resource allocation lemma, Lemma 4.3. Specifically, the functions $f_i(r_i) = s_i \ln s_i$ of r_i are twice differentiable on $[0,q_i)$, and satisfy $f_i^{''} > 0$ and $\lim_{r_i \to q_i^-} f_i^{'}(r_i) = \infty$. Further, the equality constraint in Eq. (1) becomes $\sum_i r_i = \sigma$. In this special case, $f_i^{'}(r_i) = -\frac{1}{1-\sigma}\left(\ln \frac{q_i - r_i}{1-\sigma} + 1\right)$ and

$$f_i^{'-1}(v) = q_i - (1-\sigma)e^{-(1-\sigma)v-1},$$

the solution for r_i when $r_i > 0$.

The labeling assumption (3) is equivalent to the assumption that $f_n^{'}(0) \leq \cdots \leq f_1^{'}(0)$ in the lemma, since $f_i^{'}(0) = -\frac{1}{1-\sigma}\left(\ln \frac{q_i}{1-\sigma} + 1\right)$ is a strictly decreasing function of q_i. From the second part of the lemma,

$$\sigma = \sum_{j=i}^{n} f_j^{'-1}(v) = \bar{Q}_{i-1} - (n-i+1)(1-\sigma)e^{-(1-\sigma)v-1},$$

and hence,

$$v = -\frac{1}{1-\sigma}\left(\ln\frac{\bar{Q}_{i-1}-\sigma}{(1-\sigma)(n-i+1)}+1\right).$$

Now it suffices to substitute v into $f_i^{'}(v)$ in order to obtain the expression for the non-zero optimal suppression strategy r_j in the theorem. The optimal user's apparent tag distribution s is easily derived from this expression.

Next, we shall confirm the interval of values of σ in which it is defined. To this end, observe that the condition $f_i^{'}(0)<v$ in the lemma, is equivalent to

$$-\frac{1}{1-\sigma}\left(\ln\frac{q_i}{1-\sigma}+1\right)<-\frac{1}{1-\sigma}\left(\ln\frac{\bar{Q}_{i-1}-\sigma}{(1-\sigma)(n-i+1)}+1\right),$$

and finally, after routine algebraic manipulation, to

$$\sigma > \bar{Q}_i - q_i(n-i).$$

One could proceed to carry out an analogous analysis on the upper bound condition $v\leq f_{i-1}^{'}(0)$ of the lemma to determine the interval of values of σ in which the solution is defined. However, it is simpler to realize that because a unique solution will exist for each σ, then the intervals resulting from imposing $f_i^{'}(0)<v\leq f_{i-1}^{'}(0)$ must be contiguous and nonoverlapping, hence, of the form $(\sigma_i,\sigma_{i-1}]$. Further, since $\mathcal{P}(\sigma)$ is continuous on [0,1), one may write the intervals as $[\sigma_i,\sigma_{i-1}]$ in lieu of $(\sigma_i,\sigma_{i-1}]$.

To complete the proof, we shall express the privacy-suppression function in terms of the optimal user's apparent tag distribution, that is, $\mathcal{P}(\sigma) = -\sum_{j=1}^{n} s_j \ln s_j$. We split the sum into two parts, namely,

$$-\sum_{j=1}^{i-1}\frac{q_j}{1-\sigma}\ln\frac{q_j}{1-\sigma}-\sum_{j=i}^{n}\frac{\bar{Q}_{i-1}-\sigma}{(1-\sigma)(n-i+1)}\ln\frac{\bar{Q}_{i-1}-\sigma}{(1-\sigma)(n-i+1)},$$

where we observe that the terms in the second sum do not depend on j. From this expression, it is straightforward to identify the terms of $\mathcal{P}(\sigma)$ as the entropy of the distribution

$$\left(\frac{q_1}{1-\sigma},\ldots,\frac{q_{i-1}}{1-\sigma},\frac{\bar{Q}_{i-1}-\sigma}{(1-\sigma)(n-i+1)},\ldots,\frac{\bar{Q}_{i-1}-\sigma}{(1-\sigma)(n-i+1)}\right),$$

precisely the distribution $\frac{\tilde{q}-\tilde{r}}{1-\sigma}$, given at the end of the theorem. □

The optimal tag suppression strategy in Theorem 4.4 is interpreted as follows. On the one hand, only tags corresponding to the categories $j=i,\ldots,n$ are suppressed, which is not surprising because, precisely, these are the categories with the highest probabilities, or roughly speaking, with probabilities furthest away from the uniform distribution. On the other, the optimal user's apparent tag distribution within those categories does not depend on j, and hence they all have the same probability. Further, consistently with the fact that we are dealing with relative frequencies, the components of the apparent distribution belonging to the categories $j=1,\ldots,i-1$ are obtained by normalizing the genuine user distribution. Fig. 2 captures this intuitive analysis by illustrating a simple example with $n=4$ categories. Namely, this figure shows a user with an actual profile q who is willing to

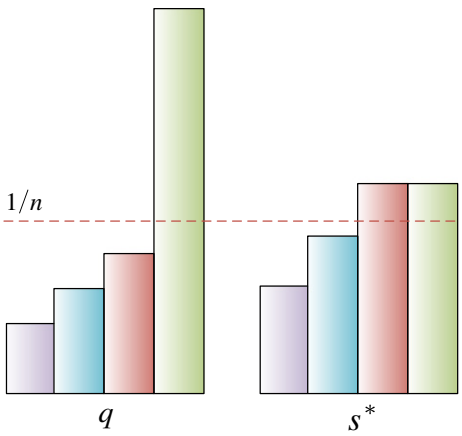


Fig. 2. A user's tag distribution q and their corresponding apparent tag distribution s^* after an optimal suppression of tags.

accept a tag suppression rate $\sigma\in[\sigma_3,\sigma_2]$ causing that a privacy attacker observe an optimal user's apparent profile s^* significantly different from q, specially in those categories with the highest ratio $\frac{q_j}{u_j}=\frac{q_j}{1/n}$.

A number of conclusions can be drawn from the results obtained in this last theorem. The following two sections will be focused on the analysis of the behavior of the privacy-suppression function at low suppression rates and high privacy.

4.4. Low-suppression case

This section investigates the privacy–suppression $\mathcal{P}(\sigma)$ in the case when $\sigma\simeq 0$.

Proposition 4.5 (Low suppression). *In the nontrivial case when $q\neq u$, there exists a positive integer i with suppression thresholds satisfying $0=\sigma_n=\cdots=\sigma_i<\sigma_{i-1}$. For all $\sigma\in[0,\sigma_{i-1}]$, the optimal tag suppression strategy r^* contains $n-i+1$ nonzero components, and the slope of the privacy–suppression function at the origin is $\mathcal{P}'(0)=\mathrm{H}(q)+\ln q_n$.*

Proof. The hypothesis $q\neq u$ implies that $n>1$, and the existence of a positive integer i enabling us to rewrite the labeling assumption (3) as

$$q_1\leq\cdots\leq q_{i-1}<q_i=\cdots=q_n,$$

and to express q_j as $\frac{\bar{Q}_{i-1}}{n-i+1}$, for $j=i,\ldots,n$. On account of Theorem 4.4,

$$0=\sigma_n=\cdots=\sigma_i<\sigma_{i-1}\leq\cdots\leq\sigma_1,$$

and for all $\sigma\in[0,\sigma_{i-1}]$, we have that

$$\mathcal{P}(\sigma)=\mathrm{H}\left(\frac{\tilde{q}-\tilde{r}}{1-\sigma}\right).$$

It is routine to check that

$$\mathcal{P}'(0)=-\sum_{j=1}^{i-1}q_j\ln q_j-\sum_{j=i}^{n}q_j\ln\frac{\bar{Q}_{i-1}}{n-i+1}+\ln\frac{\bar{Q}_{i-1}}{n-i+1}=-\sum_{j=1}^{n}q_j\ln q_j+\ln q_n,$$

where the last equality follows from the fact that $q_i=\cdots=q_n$, as shown previously. □

Now we define the *relative increment factor*

$$\delta=\frac{\mathcal{P}'(0)}{\mathcal{P}(0)}=1+\frac{\ln q_n}{\mathrm{H}(q)}.$$

The results from Proposition 4.5 allows us to approximate the privacy–suppression function at $\sigma\simeq 0$ as

$$\mathcal{P}(\sigma)\simeq\mathrm{H}(q)+\sigma(\mathrm{H}(q)+\ln q_n)$$

or, in terms of the relative increment,

$$\frac{\mathcal{P}(\sigma)-\mathrm{H}(q)}{\mathrm{H}(q)}\simeq\delta\sigma. \tag{4}$$

In conceptual terms, q_n characterizes the privacy gain at low suppression, together with $\mathrm{H}(q)$, in contrast to the fact that the ratio $\frac{q_1}{1/n}$ determines σ_{crit}, the minimum suppression rate for which the critical privacy is achievable, as defined in Section 4.2. We mentioned in that section that $q_1<1/n$ in the nontrivial case when $q\neq u$. An entirely analogous argument shows that $q_n\geq 1/n$, with equality if, and only if, $q=u$, since the opposite, that is, $q_i<1/n$, leads to a contradiction. This result allows us to conclude that $\delta<1$, unless $q=u$, for which, unsurprisingly, δ becomes zero. In other words, the relative privacy gain (4) is lower than the suppression introduced. Namely, the privacy increment at low suppression rates becomes less noticeable with smaller q_n, for a fixed $\mathrm{H}(q)$.

4.5. High-privacy case

Next, we shall analyze the case when $\sigma \simeq \sigma_{\text{crit}}$ and consequently the privacy–suppression function attains its maximum value. To this end, consider the index $i=2$ just to check that, whenever $\sigma \in [\sigma_2, \sigma_{\text{crit}}]$, for $q \neq u$,

$$\mathcal{P}(\sigma) = \mathrm{H}\left(\frac{\left(q_1, \frac{1-q_1}{n-1}, \ldots, \frac{1-q_1}{n-1}\right) - \left(0, \frac{\sigma}{n-1}, \ldots, \frac{\sigma}{n-1}\right)}{1-\sigma}\right) < \ln n.$$

In addition, we are implicitly assuming that $q_1 \neq q_2$, so that, by virtue of Theorem 4.4, $\sigma_2 < \sigma_{\text{crit}}$. Consequently, we skip an empty interval and we may express the privacy–suppression function as

$$\mathcal{P}(\sigma) = -\frac{q_1}{1-\sigma} \ln \frac{q_1}{1-\sigma} - \frac{1-q_1-\sigma}{1-\sigma} \ln \frac{1-q_1-\sigma}{(1-\sigma)(n-1)}.$$

From this expression, it is routine to conclude that $\mathcal{P}'(\sigma_{\text{crit}}) = 0$ and $\mathcal{P}''(\sigma_{\text{crit}}) = -\frac{1}{q_1^2 n^2 (n-1)}$, and finally,

$$\mathcal{P}(\sigma) \simeq \ln n + \frac{1}{2}\mathcal{P}''(\sigma_{\text{crit}})(\sigma - \sigma_{\text{crit}})^2.$$

We would like to remark that the fact that $\mathcal{P}(\sigma)$ admits a quadratic approximation for $\sigma \simeq \sigma_{\text{crit}}$, with $\mathcal{P}'(\sigma_{\text{crit}}) = 0$, may be determined directly from the fundamental properties of Fisher information [51]. Recall that for a family of distributions f_θ indexed by a scalar parameter θ, $\mathrm{D}(f_\theta \| f_{\theta'}) \simeq \frac{1}{2}\mathrm{I}(\theta')(\theta' - \theta)^2$, where $\mathrm{I}(\theta') = \mathrm{E}\left(\frac{\partial}{\partial \theta'} \ln f_{\theta'}\right)^2$ is Fisher information. Denote by $s^*_\sigma = \frac{q - r^*}{1-\sigma}$ the family of optimal apparent distributions, indexed by the suppression rate. Theorem 4.2 guarantees that $s^*_{\sigma_{\text{crit}}} = u$, thus we may write $\mathcal{P}(\sigma) = \mathrm{H}(s^*_\sigma) = \ln n - \mathrm{D}(s^*_\sigma \| s^*_{\sigma_{\text{crit}}})$. Under this formulation, it is clear that the Fisher information associated with the suppression rate is $\mathrm{I}(\sigma_{\text{crit}}) = -\mathcal{P}''(\sigma_{\text{crit}})$.

Lastly, we would like to note that the observation at the end of Section 4.2 that $r^*_1 = 0$ at $\sigma = \sigma_{\text{crit}}$ is consistent with the fact that σ_{crit} is the endpoint of the interval corresponding to the solution for r^* with $n-1$ nonzero components in Theorem 4.4.

4.6. Numerical example

In this section, we show some numerical results for a simple but insightful example that will illustrate the formulation presented in Section 3.4 and the theoretical analysis argued in Section 4. The evaluation of our privacy-enhancing mechanism in a real-world application is presented later in Section 5.

In this practical example, we shall consider three categories and assume that the user's distribution is $q = (0.100, 0.200, 0.700)$, thus fulfilling both the positivity and the labeling assumptions (2, 3). On account of Theorem 4.4, the suppression thresholds are $\sigma_3 = 0$, $\sigma_2 = 0.500$ and $\sigma_1 = \sigma_{\text{crit}} = 0.700$. In addition, the initial privacy value is $\mathcal{P}(0) \simeq 0.8018$, which is the privacy level achieved by a user who is not willing to accept the suppression of any tag. Furthermore, Sections 4.4 and 4.5 allow us to characterize the behavior of the privacy–suppression function for $\sigma = 0$ and $\sigma = \sigma_{\text{crit}}$. Concretely, the first and second order approximations are determined by the quantities $\mathcal{P}'(0) \simeq 0.4451$ and $\mathcal{P}''(\sigma_{\text{crit}}) \simeq -5.56$. All these results are captured in Fig. 3, where the privacy–suppression function $\mathcal{P}(\sigma)$ is represented. Namely, the optimization problem involved in the definition of this function has been computed theoretically, by simply applying Theorem 4.4, and numerically.[2]

After observing the behavior of the optimal trade-off curve between privacy and suppression, now we turn to examine the optimal apparent tag distribution for a set of suppression rates. To this end, the user's distribution q, the optimal apparent distribution s^* and the uniform distribution u are represented in the probability simplexes shown in Fig. 4. In addition, the contours of the entropy $\mathrm{H}(\cdot)$ of a distribution in the simplex are depicted. More interestingly, this figure also shows the region, highlighted in dark blue, which corresponds to all the possible apparent tag distributions, not necessarily optimal, for a given suppression rate. Namely, this feasible region results from the intersection of the set $\left\{s = \frac{q-r}{1-\sigma} \,|\, 0 \leq r_i \leq q_i, \sum_i r_i = \sigma\right\}$, and the probability simplex.

We now turn our attention to Fig. 4(a), where a suppression $\sigma \in [\sigma_3, \sigma_2]$ has been selected to check that, according to the notation of Theorem 4.4, r^* has $n - i + 1 = 1$ nonzero components. Geometrically, this places the solution s^*, not entirely unexpectedly, at one vertex in the feasible region. In addition, observe that a suppression of 10% increases the user privacy to a 5.8% of the original privacy $\mathrm{H}(q)$. This confirms an interesting result obtained in Section 4.4, where we concluded that the relative increment factor δ for low-suppression rates was lower than the suppression introduced. In Fig. 4(b) the suppression rate is on the interval $[\sigma_2, \sigma_{\text{crit}}]$, leading to an optimal suppression strategy r^* with $n - i + 1 = 2$ nonzero components. In this case, the solution s^* is placed on one edge of the feasible region. Additionally, note that a suppression of 55% increments the user privacy to a 33% of its original value. The case in which $\sigma = \sigma_{\text{crit}}$ and thus user privacy attains its maximum value is depicted in Fig. 4(c). When this happens, r^* still has $n - i + 1 = 2$ nonzero components. Precisely, note that $r^*_3 = q_3 - q_1$ and $r^*_2 = q_2 - q_1$, which perfectly agree with the results obtained at the end of Section 4.2. Finally, the case when $\sigma > \sigma_{\text{crit}}$, which certainly does not make

[2] The numerical method chosen is the interior-point optimization algorithm [60] implemented by the Matlab R2009a function `fmincon`.

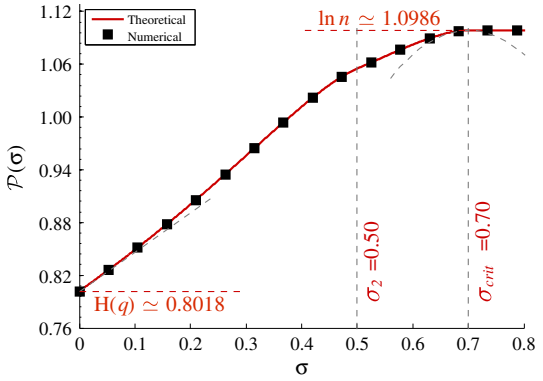
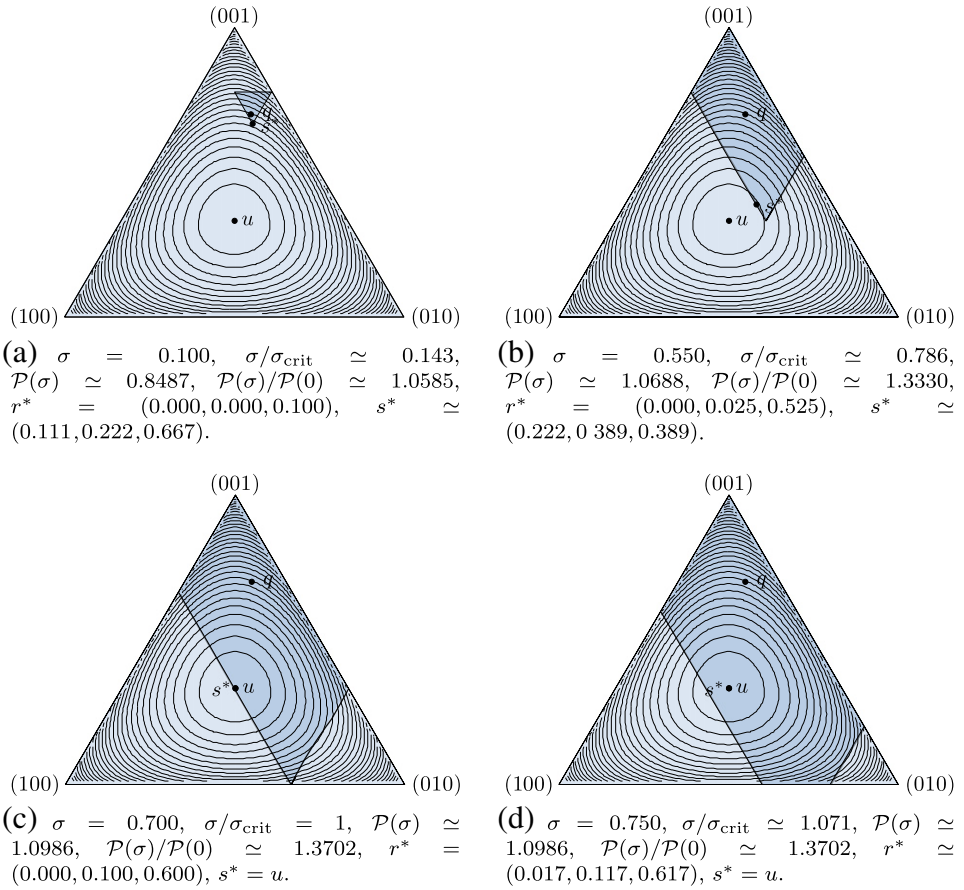


Fig. 3. Optimal trade-off curve between privacy and suppression, and the corresponding approximations and suppression thresholds for $q=(0.100, 0.200, 0.700)$.

sense, is shown in Fig. 4(d). In this particular case, r^* has $n-i+1=3$ nonzero components and s^* falls into the interior of the feasible region.

5. Evaluation

The purpose of this section is twofold. First, we compare our tag suppression technique with some of the most prominent approaches in the area of anonymous-communication systems. Secondly, we evaluate the impact that the application of our technique would have on a real-world tagging system.



(a) $\sigma = 0.100$, $\sigma/\sigma_{\text{crit}} \simeq 0.143$, $\mathcal{P}(\sigma) \simeq 0.8487$, $\mathcal{P}(\sigma)/\mathcal{P}(0) \simeq 1.0585$, $r^* = (0.000, 0.000, 0.100)$, $s^* \simeq (0.111, 0.222, 0.667)$.

(b) $\sigma = 0.550$, $\sigma/\sigma_{\text{crit}} \simeq 0.786$, $\mathcal{P}(\sigma) \simeq 1.0688$, $\mathcal{P}(\sigma)/\mathcal{P}(0) \simeq 1.3330$, $r^* = (0.000, 0.025, 0.525)$, $s^* \simeq (0.222, 0\ 389, 0.389)$.

(c) $\sigma = 0.700$, $\sigma/\sigma_{\text{crit}} = 1$, $\mathcal{P}(\sigma) \simeq 1.0986$, $\mathcal{P}(\sigma)/\mathcal{P}(0) \simeq 1.3702$, $r^* = (0.000, 0.100, 0.600)$, $s^* = u$.

(d) $\sigma = 0.750$, $\sigma/\sigma_{\text{crit}} \simeq 1.071$, $\mathcal{P}(\sigma) \simeq 1.0986$, $\mathcal{P}(\sigma)/\mathcal{P}(0) \simeq 1.3702$, $r^* \simeq (0.017, 0.117, 0.617)$, $s^* = u$.

Fig. 4. Probability simplexes showing u, q and s^* for several interesting values of σ.

5.1. Anonymous-communication systems vs. tag suppression

In Sections 1.1 and 2, we briefly commented on the different nature of the assumptions upon which anonymous-communication systems and tag suppression rely. Particularly, we argued that users of anonymous communications entrust their private data to a TTP which is then responsible for protecting their privacy; our approach, however, is built on the supposition that users mistrust communicating entities, e.g., the semantic Web server itself, but also the Internet service provider (ISP), intermediary routers and firewalls, or users of the same local area network. In this section, we shall elaborate more on these differences by examining the most relevant contributions in the area of anonymous communications. The ultimate purpose is to compare them qualitatively with our technique.

In essence, an anonymous-communication system is a system that, working on the top of Internet protocols, aims at hiding the correspondence between users exchanging messages. Since the introduction of Chaum's *mixes* in 1981 [12], which were the first anonymous-communication mechanism, a wide and rich variety of approaches have been proposed to achieve the same goal [61–66,13–15]. Many of these approaches are variations in the design of the original mix, which, despite the time elapsed, remains a key building block of anonymous-communication systems [67].

Fundamentally, mix systems and all their variations are nodes that collect and forward messages so that it is unfeasible for an attacker to link an outgoing message to its corresponding input message. Among these systems, one of the most important varieties is a family of mixes known as *threshold pool mixes* [61]. The leading idea is that this type of mixes collects a number of incoming messages, stores them on the internal memory of the mix, and outputs some of them at a time when the number of messages kept in the memory meets a certain threshold. In order to eliminate any correlation between outgoing and incoming messages, the mix modifies the flow of messages by using two strategies, namely the delay and reordering of messages.

Another important group of pool mixes outputs messages based on time [62]. Essentially, these *timed* mixes forward all messages kept in the memory every fixed interval of time called timeout. The major advantage of these mixes is that the delay experienced by messages is upper bounded, in contrast to the case of threshold pool mixes. The flip side is that the unlinkability between incoming and outgoing messages may be seriously compromised when the number of messages arriving in that interval of time is small. Motivated by this, some of the current mix designs implement a combination of the strategies based on threshold and those based on time [68]. An alternative to pool mixes are the mixes based on the concept of *stop-and-go*, known as *continuous* mixes [63], where users specify the time that their messages will be stored in the mix. Finally, we would like to mention the use of networks of mixes [64,65], an approach that has been thoroughly studied in the literature. The reason is evident — on the one hand, routing messages through several mixes makes it more difficult for an attacker to track messages, and on the other hand, it improves the availability of the anonymous-communication system.

While all these approaches may provide unlinkability to a certain extent, this is at the cost of delaying messages, which affects the usability of these systems and hence imposes a cost on them. In other words, mix systems pose a trade-off between anonymity and utility, as in our tag suppression technique. But in addition to this trade-off, mix systems require the deployment of infrastructure and, more importantly, assume that users are disposed to trust them. Nevertheless, even though they were completely trustful, they could not prevent the recipient of those messages, i.e., the Web server, from profiling users and ultimately identifying them. Recall that mixes just provide unlinkability in the sense that an *external* attacker cannot ascertain the correspondence between input and output messages.

In an attempt to overcome some of these limitations, namely the delay introduced by mixes, Refs. [14,13] propose an architecture called *onion routing*. In such approach, messages are routed through a network of nodes, similarly to the scenario of mix networks, but with the difference that messages are not delayed. Specifically, when a user wishes to send a message, they submit it first to one of these nodes. Then, the node encrypts the message in a layered fashion and chooses the intermediate nodes to reach the recipient. Afterwards, each of these intermediate nodes peels off a layer of encryption and forwards the resulting message to the next node in the route. In the end, the last node delivers the message to the recipient. Considering how the system works, we may conclude that the functionality of the nodes essentially boils down to relaying messages. Clearly, this is in contrast to the case of mix systems, where messages are also delayed. An improvement of this system is the second-generation version of onion routing, Tor [15], which has been available to Internet users since 2002. Despite being an improvement on onion routing, Tor nodes do not delay messages either, rendering the system susceptible to traffic analysis based on timing comparisons and other more sophisticated attacks [16–19]. Although these systems reduce the delay inherent in mixes, they suffer exactly from the same limitations in terms of infrastructure, trustworthiness and privacy protection.

Abandoning the idea of mixes and onion routing, a conceptually-simple approach to anonymous tagging consists in a TTP acting as an intermediary between the user and the Web server. In this scenario, the server cannot know the user ID, but merely the identity of the TTP itself involved in the communication. Alternatively, the TTP may act as a pseudonymizer by supplying a pseudonym ID' to the server, but only the TTP knows the correspondence between the pseudonym ID' and the actual user ID. A convenient twist to this approach is the use of digital credentials [41–43] granted by a trusted authority, namely digital content proving that a user has sufficient privileges to carry out a particular transaction without completely revealing their identity. The main advantage is that the TTP need not be online at the time of service access to allow users to access a service with a certain degree of anonymity. Unfortunately, these approaches do not prevent the Web server from inferring the real identity of a user by colluding with the network operator or some entity involved in the communication. In addition, all TTP-based solutions require that users shift their trust from the Web server to another party, possibly capable of collecting tags from different applications, which ultimately might facilitate user profiling via crossreferencing. In the end, traffic bottlenecks are a potential issue with TTP solutions.

Apart from the systems based on TTPs, there exist a myriad of alternatives based on user collaboration [39,40,21,66]. One of the most popular is *Crowds* [66], which contemplates that a group of users wanting to browse the Web will collaborate to submit

their requests. With this purpose, a user who decides to send a request to a Web server, selects first a member of the group at random and then forwards the request to it. When this member receives the request, it flips a biased coin to determine whether to send the request to another member or to submit it to the Web server. This process is repeated until the request is finally relayed to the intended destination. As a result, the Web server and any of the members forwarding the request cannot ascertain the identity of the true sender, that is, the member who initiated the request. While this approach does not require the use of a TTP, its main limitation lies in the assumption that a number of users will participate in the protocol. However, even though it was possible, this solution could not protect user privacy against the collusion of all participants. Another important drawback is the additional traffic intrinsic to this forwarding mechanism.

Unlike the alternatives examined in this section, our technique appears as a simple approach in terms of infrastructure requirements, as users need not trust an external entity, the network operator nor other users. Our technique, which falls into the category of data minimization, enables users to protect their privacy against the collusion of any passive attackers, but at the cost of semantic loss incurred by suppressing tags. Precisely, this privacy-utility trade-off also appears in anonymous-communication systems and collaborative approaches. In these two cases, the degradation in utility is the delay introduced by mixes and the traffic overhead incurred by a forwarding strategy, respectively. Table 2 summarizes the major conclusions of this section.

5.2. Experimental analysis

In this section, we analyze the extent to which our technique enables users to enhance their privacy in a real-world tagging application. Our analysis also contemplates the impact that the suppression of tags has on the semantic functionality of this application, but tackles this in a more simplified manner, by using a tractable measure of utility, namely the tag suppression rate. With this aim, Section 5.2.1 first examines the data set that we used to conduct the experimental evaluation. To make user profiles tractable, Section 5.2.2 describes a complete methodology for mapping tags into a small set of meaningful categories of interest. Finally, Section 5.2.3 presents the experimental results.

5.2.1. Data set

We applied the proposed technique to *BibSonomy* [3], a popular social bookmarking and publication-sharing system. In particular, we experimented with the data set retrieved by the Knowledge & Data Engineering Group at the University of Kassel [69]. The data set in question comprises those bookmarks and publications tagged by approximately two thousand users. The information is organized in the form of triples (*username*, *resource*, *tag*), each one modeling the action of a user who associates a resource, being a bookmark or a publication, with a tag. Our data set contains 671,807 of these triples, which were posted from Jan. 1989 to Dec. 2007, and includes 1921 users, 206,941 resources and 58,755 tags. It is worth mentioning that no preprocessing was done, although usernames were anonymized.

5.2.2. Tag categorization

The representation of a user profile as a normalized histogram across these 58,755 tags is clearly an inappropriate approach for our experiments; not only because of the intractability of the profile, but also because it makes it difficult to have a quick overview of the user interests. For example, for users posting the tags “welfare”, “Dubya” and “campaign” it would be preferable to have a higher level of abstraction that enables us to conclude, directly from the inspection of their profiles, that they are interested in politics. In this sense, the categorization of tags allows us to represent user profiles in a tractable manner, on the basis of a reduced set of meaningful categories of interest, consistently with the model assumed in Section 3.1. With this spirit, next we proceed to describe the methodology that we followed to cluster the tags into categories. The categorization process is in line with other works on this field [70,71].

Table 2
Comparison among some popular anonymous-communication systems and our privacy-enhancing technique.

Approaches	Technique	Privacy	Disadvantages
Mixes [12,61–65]	TTP	Soft	o Delay experienced by messages, o Users must trust an external entity, o Vulnerable to collusion attacks, o Infrastructure requirements
Onion routing [13–15]	TTP	Soft	o Traffic analysis and more sophisticated attacks, o Users must trust an external entity, o Vulnerable to collusion attacks, o Infrastructure requirements
Anonymizer, pseudonymizer, credentials [41–43]	TTP	Soft	o Users must trust an external entity, o Vulnerable to collusion attacks, o Traffic bottlenecks
Crowds and similar [66,39,40,21]	User collaboration	Soft	o Numerous users must collaborate, o Vulnerable to collusion attacks, o Traffic overhead
Tag suppression	Data minimization	Hard	o Semantic loss incurred by suppressing tags

To accomplish this categorization, first we carried out some preprocessing to filter out those tags considered as spam. For this purpose, we dropped the tags with a number of characters over 26, which in our data set represented the 99th percentile. In addition, we eliminated those posts with no tags. After this simple preprocessing, the number of triples reduced to 665,052, and, consequently, the number of users, resources and tags to 1916, 206,697, and 50,900, respectively.

In a second stage, we aimed at identifying clusters or groups of semantically similar tags. As frequently done in the literature, we performed a clustering analysis based on the *co-occurrence* between tags, that is, the number of times each pair of tags simultaneously appears in a same resource. Specifically, we modeled the relationships among tags as a matrix of co-occurrences c_{ij}, where each entry with $i \neq j$ corresponds to the co-occurrence between tags i and j, and each entry in the diagonal is a self-occurrence, i.e., the absolute frequency of appearance of a tag. Note that, clearly, this is a symmetric matrix and that each row (column) describes one tag in terms of the semantic similarity to the other tags.

In an attempt to concentrate on the significant relationships among these tags, we eliminated those rows satisfying $\sum_j c_{ij} < t$, for a certain threshold t. Similarly, we dropped those columns fulfilling an equivalent condition. In this regard, observe that, the higher the threshold, the lower the number of resulting tags, and thus the lower the total number of triples. Since we aimed at preserving at least 80% of the triples, and at the same time, we required the resulting tags to have a strong co-occurrence, we chose $t = 100$. In conclusion, after this filtering process the number of users, resources, tags and triples became 1737, 190,478, 5057 and 540,904, respectively.

Once we filtered the co-occurrence matrix, we proceeded to use a well-known clustering algorithm to group all these tags into categories. But before applying this algorithm, we first required to specify a measure of similarity among tags. Recall that we modeled tags both as rows and columns of a matrix, that is, vectors. As often done in the literature, we employed the cosine metric [72], a simple and robust measure of similarity between vectors. Equipped with this measure, we applied Lloyd's algorithm.[3] As a result, we clustered the 5057 tags into 5 categories, which gave us a granularity level sufficiently aggregated as to avoid having user profiles with many empty categories. Afterwards, the resulting categories were classified in increasing order of popularity of their tags, with the aim of satisfying the labeling assumption (3). Although this classification does not necessarily imply that all user profiles meet this condition, in our experiments we shall ultimately rearrange the categories of each individual profile to fulfill it. Lastly, the tags in each category were sorted in decreasing order of proximity to the centroid.[4]

In a last stage, and on account of the positivity assumption (2), we eliminated those users who did not tag across all categories. In addition, we dropped users with an activity level lower than 50 tags, since it would have been difficult to calculate a reliable estimate of their profiles with such a few tags. Accordingly, the number of users, resources and triples became 209, 144,904 and 447,203, respectively.

5.2.3. Results

In this section, we examine the extent to which our technique contributes to privacy preservation. For this purpose, first we explore how a particular user in our data set benefits from the application of an optimal tag suppression strategy; and secondly, we analyze the effect of this optimal suppression when the whole population of users enhances their privacy by using a common tag suppression rate.

As detailed in previous sections, tag suppression requires that a user specify a rate indicating the fraction of tags they are disposed to eliminate. Based on this suppression rate and the user profile across the $n = 5$ categories obtained in Section 5.2.2, our approach solves the optimization problem (1). The result of this optimization is a suppression strategy r^*, that is, an n-tuple containing the percentage of tags that the user should eliminate in each category. In our first series of experiments, we select a particular user in our data set[5] and compute this suppression strategy in the special case when the user specifies $\sigma = \sigma_{\text{crit}}$. Both the actual profile of the user in question and the optimal strategy are plotted in Fig. 5, where it is shown one of the theoretical results obtained in Section 4.2, namely the fact that $r_i^* = q_i - q_1$ for any category i. In addition, Fig. 6 illustrates the optimal trade-off curve between privacy and suppression, which we calculated theoretically and numerically. The suppression thresholds σ_i shown in this figure indicate the suppression rates beyond which the components $j = i,\ldots, n$ of the apparent profile s have the same probability. This effect is observed in Fig. 7, where we represent s precisely for these interesting values of σ.

The second set of experiments contemplates a scenario where all users apply our technique by using a common tag suppression rate. Under this assumption, Fig. 9 shows the privacy protection achieved by these users in terms of percentile curves (10th, 25th, 50th, 75th and 90th) of relative privacy gain. Noteworthy is the fact that certain users obtain privacy gains between 100% and 235%, although, clearly, at the cost of high suppression rates. Another eye-opening finding is the distribution of the suppression thresholds σ_i plotted in Fig. 8. Recall that we also refer to σ_1 as the critical suppression σ_{crit}. Particularly, we observe that 86.6% of users have $\sigma_1 \in [0.9,1)$, whereas the remaining percentage of users lie in the interval $[0.7,0.9)$. In practice, this means that all users will require a high suppression rate for their profiles to become completely uniform. Although this might be certainly controversial, this is not a poor performance of our mechanism, but a consequence of the stringent privacy requirement imposed by such uniformity. As a matter of fact, the distributions of σ_i, for $i = 2, 3, 4$, indicate that the components s_j with $j = i,\ldots, n$ may be uniform at a significantly lower cost. For example, 32.5% of users have 3 out of 5 components evenly balanced for a suppression rate below 68%.

[3] Lloyd's algorithm [73], which is normally referred to as $k-$means in the computer science community, is a popular iterated algorithm for grouping data points into a set of k clusters.

[4] The complete results of this clustering are available to other researchers at http://sites.google.com/site/javierparraarnau/publications.

[5] This specific user is identified by the number 633 in Ref. [69].

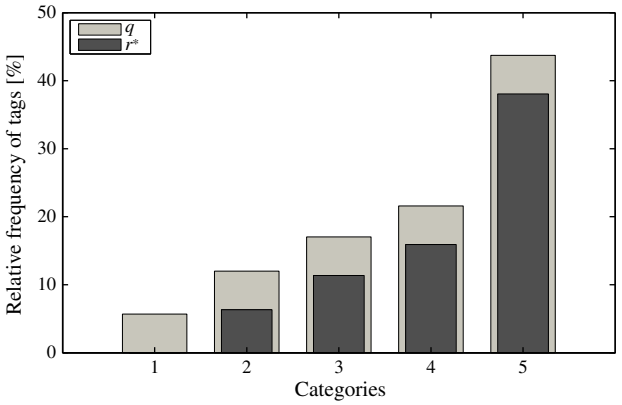


Fig. 5. In this figure, we plot the actual user profile q of the particular user considered in Section 5.2.3, who posted a total of 1075 tags across all categories. Additionally, we plot the optimal suppression strategy r^* for $\sigma=\sigma_{\text{crit}}\approx 0.7163$, that is, the percentage of tags that the user should refrain from tagging in each category in order to achieve the uniform profile.

In closing, the results shown in this section illustrate how our mechanism perturbs the user profile observed from the outside and how this perturbation enables users to protect their privacy to a certain degree.

6. Concluding remarks

There exists a large number of proposals for privacy protection in the semantic Web. Within these approaches, tag suppression arises as a simple data minimization strategy in terms of infrastructure requirements, as users need not trust an external entity nor the network operator. Interestingly, as commented in Section 1.1, our technique may also be used in combination with other mechanisms such as traditional anonymous-communication systems and therefore it may contribute to improve their effectiveness. Nevertheless, our approach comes at the cost of some processing overhead but more importantly at the expense of semantic loss incurred by suppressing tags. In other words, tag suppression poses an inherent trade-off between privacy and suppression.

Our main contribution is, precisely, a systematic, mathematical approach to the problem of optimal tag suppression. We measure user privacy as the entropy of the user's apparent tag distribution, after the suppression of tags, and justify it with the rationale behind entropy-maximization methods. Subsequently, we formulate and solve an optimization problem modeling the privacy–suppression trade-off. Noteworthy is the fact that our theoretical analysis may also be applied to a wide spectrum of user-generated data, rather than semantic tags. For example, one could conceive the elimination of queries in information retrieval or the suppression of ratings in the domain of recommendation systems.

In our mathematical model, we represent user tags as r.v.'s taking on values on a common finite alphabet of categories or topics. This allows us to describe user profiles as PMFs, a representation that is frequently used in popular tagging systems such as *BibSonomy*, *CiteULike* and *Delicious*. The proposed model, however, is restricted to relative frequencies, relevant against content-based attacks, but does not deal with differences in the absolute frequencies, which certainly could be exploited by traffic

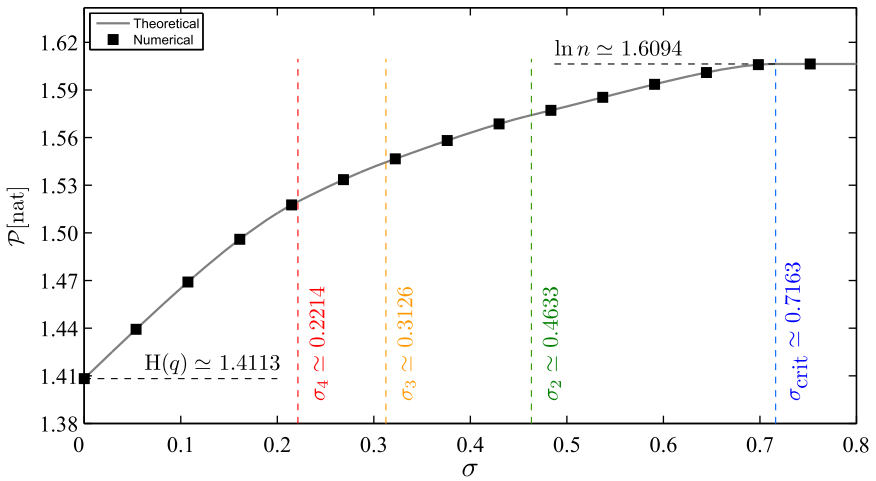


Fig. 6. We plot the privacy–suppression trade-off and the suppression thresholds for one particular user in our data set.

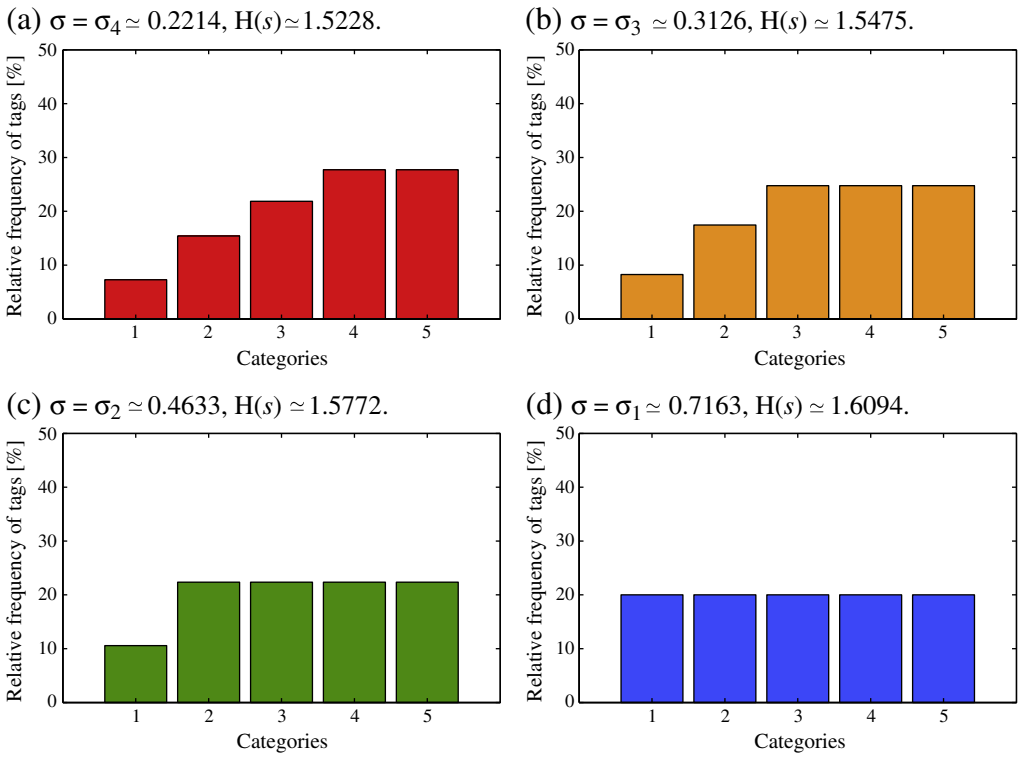


Fig. 7. We represent the apparent profile s of a particular user in the special case when the suppression rate coincides with the suppression thresholds $\sigma_i, i=1,\ldots,4$. Recall that s is the perturbed profile resulting from the elimination of tags and observed from the outside. At these interesting values of suppression, we observe how the components of s corresponding to the categories $j=i,\ldots,5$ are balanced. In the end, when the critical suppression σ_1 is attained, s becomes u and $\mathrm{H}(s)=\ln 5 \simeq 1.6094$. The actual profile of this specific user is depicted in Fig. 5.

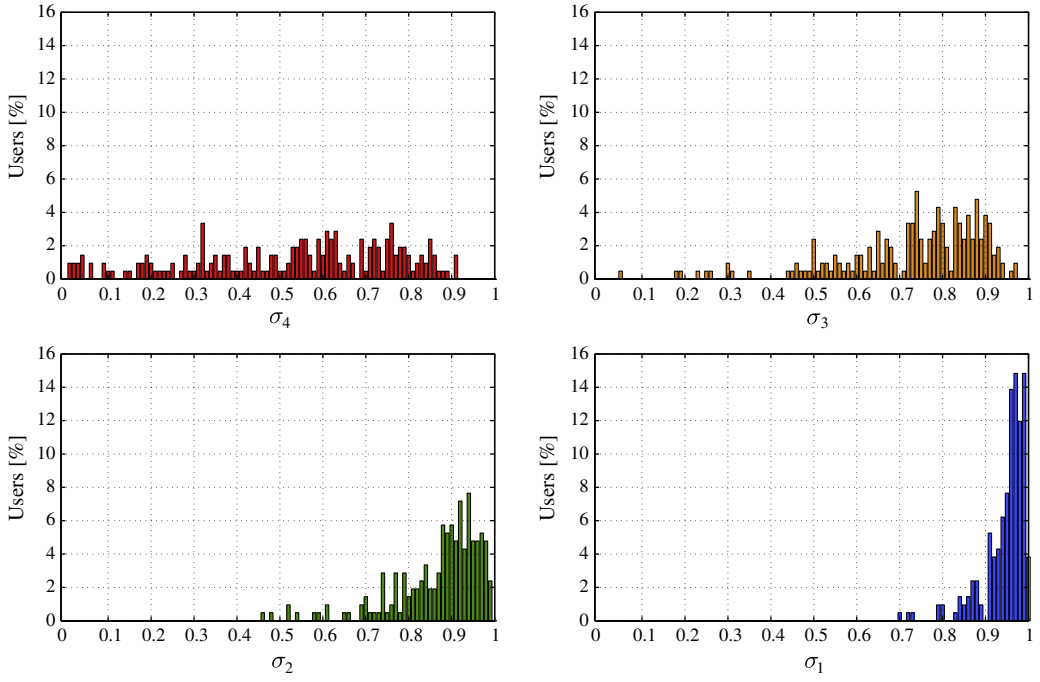


Fig. 8. We plot the distribution of the suppression thresholds σ_i, for $i=1,\ldots,4$. In the special case when $\sigma \geq \sigma_1 = \sigma_{\mathrm{crit}}$, the apparent user profile is the uniform profile across all categories.

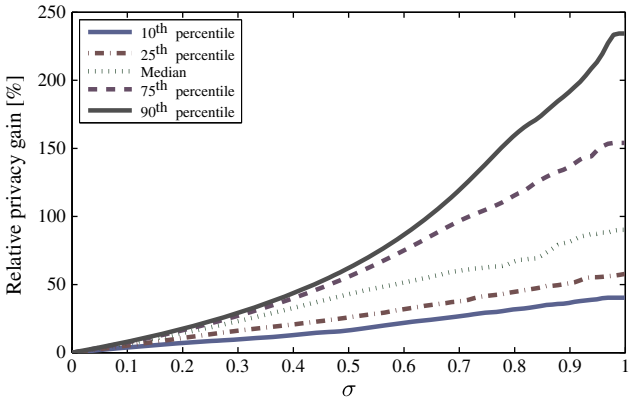


Fig. 9. We consider the case when all users in our data set protect their privacy by using a common tag suppression rate. Built on this premise, we then plot some percentiles curves of privacy gain against this common suppression rate.

analysis. Besides, we assume, on the one hand, a rudimentary adversary model where attackers are not able to estimate a particular user's tag suppression rate, and on the other, that only a small number of users adhere to this strategy.

As a result of our theoretical analysis, we present a close-form solution for the optimal tag suppression strategy and a privacy–suppression function characterizing the optimal trade-off curve. Our mathematical approach bears certain resemblance to the water-filling problem in rate-distortion theory, and is restricted to the discrete case of n tag categories. In addition, there are several analogies with Ref. [29], albeit in Section 2 we discarded forgery as a privacy-enhancing mechanism for semantic tagging, and deemed tag suppression a more suitable strategy.

Our theoretical study first proves that the privacy–suppression function $\mathcal{P}(\sigma)$ is nondecreasing and quasiconcave. Subsequently, we show that, under the positivity assumption (2), there exists a critical suppression $\sigma_{\text{crit}}<1$ beyond which the critical privacy is achievable. Specifically, this σ_{crit} only depends on the minimum ratio $\frac{q_j}{u_j}$ of probabilities between the user's tag distribution q and the uniform distribution u. More interestingly, for a given suppression σ, the suppression of tags only affects the categories $j=i,\ldots,n$, precisely those with the highest probabilities among all categories. Not unexpectedly, the number of categories exposed to suppression, that is, $n-i+1$, increases with σ. In the particular case when $\sigma=\sigma_{\text{crit}}$, only the category $i=1$ remains unchanged. With regard to the optimal user's apparent distribution, the components of s^* corresponding to the categories $j=i,\ldots,n$ have the same probability, whereas the probability of the other components is obtained by normalizing the actual user distribution.

Further, we characterize $\mathcal{P}(\sigma)$ at low suppression and high privacy. Specifically, we provide a first-order approximation for $\sigma\simeq 0$ in the nontrivial case when $q\neq u$, from which we conclude that q_n determines, together with the initial privacy value, the privacy gain at low suppression. In addition, we prove that this privacy gain is lower than the suppression introduced. Besides, we provide a second-order approximation for $\sigma\simeq\sigma_{\text{crit}}$, assuming that probabilities q_j are strictly increasing. Finally, we interpret that $\mathcal{P}'(\sigma)$ vanishes at $\sigma=\sigma_{\text{crit}}$ as a consequence of a fundamental property of the Fisher information.

Our theoretical analysis is then illustrated with a simple but insightful example. But this is not until Section 5, where we provide an experimental evaluation of our privacy-enhancing mechanism. In this section, we investigate the application of tag suppression to a real-world application and assess experimentally the extent to which our approach may help users protect their privacy. Our analysis also contemplates the impact that the suppression of tags has on the semantic functionality of this application, but tackles this in a more simplified manner, by using the tag suppression rate as a measure of utility.

Acknowledgments

We would like to thank the anonymous referees for their thorough, extremely valuable comments, which motivated major improvements on this manuscript. 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. David Rebollo-Monedero is the recipient of a Juan de la Cierva postdoctoral fellowship, JCI-2009-05259, from the Spanish Ministry of Science and Innovation.

References

[1] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, Scientific American (2001) 35–43.
[2] L. Feigenbaum, I. Herman, T. Hongsermeier, E. Neumann, S. Stephens, The semantic web in action, Scientific American 297 (2007) 90–97.
[3] Bibsonomy. URL http://www.bibsonomy.org.
[4] Citeulike. URL http://www.citeulike.org.
[5] Delicious. URL http://delicious.com/.
[6] Stumbleupon. URL http://www.stumbleupon.com.
[7] E. Michlmayr, S. Cazer, Learning user profiles from tagging data and leveraging them for personal(ized) information access, in: Proc. Workshop Tagging and Metadata for Social Inform. Org. Workshop in Int. WWW Conf., 2007.
[8] A. John, D. Seligmann, Collaborative tagging and expertise in the enterprise, in: Proc. Col. Web Tagging Workshop WWW, 2006.
[9] S. Warren, L. Brandeis, The right to privacy, Harvard Law Review 4 (5) (1890) 193–220.

[10] D.J. Solove, Understanding Privacy, Harvard Univ, Press, 2009.
[11] M. Deng, Privacy preserving content protection, Ph.D. thesis, Katholieke Universiteit Leuven—Faculty of Engineering (2010).
[12] D. Chaum, Untraceable electronic mail, return addresses, and digital pseudonyms, Communications of the ACM 24 (2) (1981) 84–88.
[13] M.G. Reed, P.F. Syverson, D.M. Goldschlag, Proxies for anonymous routing, in: Proc. Comput. Secur. Appl. Conf. (CSAC), San Diego, CA, 1996, pp. 9–13.
[14] D. Goldschlag, M. Reed, P. Syverson, Hiding routing information, in: Proc. Inform. Hiding Workshop (IH), 1996, pp. 137–150.
[15] R. Dingledine, N. Mathewson, P. Syverson, Tor: the second-generation onion router, in: Proc. Conf. USENIX Secur. Symp., Berkeley, CA, 2004, pp. 303–320.
[16] B.N. Levine, M.K. Reiter, C. Wang, M. Wright, Timing attacks in low-latency mix systems, in: Proc. Int. Financial Cryptogr. Conf., Springer-Verlag, 2004, pp. 251–265.
[17] K. Bauer, D. McCoy, D. Grunwald, T. Kohno, D. Sicker, Low-Resource Routing Attacks against Anonymous Systems, Tech. Rep, University of Colorado, 2007.
[18] S.J. Murdoch, G. Danezis, Low-cost traffic analysis of Tor, in: Proc. IEEE Symp. Secur., Priv. (SP), 2005, pp. 183–195.
[19] B. Pfitzmann, A. Pfitzmann, How to break the direct RSA implementation of mixes, in: Proc. Annual Int. Conf. Theory, Appl. of Cryptogr. Techniques (EUROCRYPT), Springer-Verlag, 1990, pp. 373–381.
[20] W.M. Grossman, Alt.scientology.war. URL www.wired.com/wired/archive/3.12/alt.scientology.war_pr.html, 1996.
[21] D. Rebollo-Monedero, J. Forné, A. Solanas, T. Martínez-Ballesté, Private location-based information retrieval through user collaboration, Computer Communications 33 (6) (2010) 762–774.
[22] A.M. Mcdonald, R.W. Reeder, P.G. Kelley, L.F. Cranor, A comparative study of online privacy policies and formats, in: Proc. Workshop Priv. Enhanc. Technol. (PET), Springer-Verlag, Seattle, WA, 2009, pp. 37–55.
[23] C. Jensen, C. Potts, C. Jensen, Privacy practices of internet users: Self-reports versus observed behavior, International Journal of Human-Computer Studies 63 (1–2) (2005) 203–227.
[24] L. Kagal, T. Finin, A. Joshi, A policy based approach to security for the semantic web, in: Proc. Int. Semantic Web Conf., 2003, pp. 402–418.
[25] L. Kagal, M. Paolucci, N. Srinivasan, G. Denker, T. Finin, K. Sycara, Authorization and privacy for semantic web services, IEEE Journal of Intelligent Systems 19 (4) (2004) 50–56.
[26] G. Friedland, J. Choi, Semantic computing and privacy: a case study using inferred geo-location, International Journal of Semantic Computing 5 (1) (2011) 79–93.
[27] K. Levy, B. Sargent, Y. Bai, A trust-aware tag-based privacy control for ehealth 2.0, in: Proc. Int. Conf. Inform. Technol. Educ. (SIGITE), ACM, New York, USA, 2011, pp. 251–256.
[28] B. Carminati, E. Ferrari, A. Perego, A multi-layer framework for personalized social tag-based applications, Data & Knowledge Engineering (2012) 101–130.
[29] D. Rebollo-Monedero, J. Forné, Optimal query forgery for private information retrieval, IEEE Transactions on Information Theory 56 (9) (2010) 4631–4642.
[30] 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.
[31] B. Shapira, Y. Elovici, A. Meshiach, T. Kuflik, PRAW—the model for PRivAte Web, Journal of the American Society for Information Science and Technology 56 (2) (2005) 159–172.
[32] W.B. Frakes, R.A. Baeza-Yates (Eds.), Information Retrieval: Data Structures & Algorithms, Prentice-Hall, 1992.
[33] T. Kuflik, B. Shapira, Y. Elovici, A. Maschiach, Privacy preservation improvement by learning optimal profile generation rate, in: User Modeling, Vol. 2702 of Lecture Notes Comput. Sci. (LNCS), Springer-Verlag, 2003, pp. 168–177.
[34] Y. Elovici, C. Glezer, B. Shapira, Enhancing customer privacy while searching for products and services on the World Wide Web, Internet Research 15 (4) (2005) 378–399.
[35] R. Puzis, D. Yagil, Y. Elovici, D. Braha, Collaborative attack on Internet users' anonymity, Internet Research 19 (1) (2009) 60–77.
[36] D.C. Howe, H. Nissenbaum, Lessons from the identity trail: privacy, anonymity and identity in a networked society, in: Ch. TrackMeNot: Resisting surveillance in Web search, Oxford Univ. Press, NY, 2009, pp. 417–436, URL http://mrl.nyu.edu/dhowe/trackmenot.
[37] V. Toubiana, SquiggleSR. URL www.squigglesr.com, 2007.
[38] E. Balsa, C. Troncoso, C. Diaz, OB-PWS: obfuscation-based private web search, in: Proc. IEEE Symp. Secur., Priv. (SP), 2012, pp. 491–505.
[39] D. Rebollo-Monedero, J. Forné, M. Soriano, Private location-based information retrieval via *k*-anonymous clustering, in: Proc. CNIT Int. Workshop Digit. Commun., Lecture Notes Comput. Sci. (LNCS), Springer-Verlag, Sardinia, Italy, 2009, pp. 421–430, invited paper.
[40] C. Chow, M.F. Mokbel, X. Liu, A peer-to-peer spatial cloaking algorithm for anonymous location-based services, in: Proc. ACM Int. Symp. Adv. Geogr. Inform. Syst. (GIS), Arlington, VA, 2006, pp. 171–178.
[41] D. Chaum, Security without identification: transaction systems to make big brother obsolete, Communications of the ACM 28 (10) (1985) 1030–1044.
[42] G. Bianchi, M. Bonola, V. Falletta, F.S. Proto, S. Teofili, The SPARTA pseudonym and authorization system, Science of Computer Programming 74 (1–2) (2008) 23–33.
[43] V. Benjumea, J. López, J.M.T. Linero, Specification of a framework for the anonymous use of privileges, Information for Telemat 23 (3) (2006) 179–195.
[44] C. Gülcü, G. Tsudik, Mixing email with Babel, in: Proc. IEEE Symp. Netw. Distrib. Syst. Secur. (SNDSS), Washington, DC, 1996, pp. 2–16.
[45] 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.
[46] G. Danezis, Statistical disclosure attacks: traffic confirmation in open environments, in: Proc. Secur., Priv., Age Uncertainty, (SEC), Athens, Greece, 2003, pp. 421–426.
[47] M.J. Freedman, R. Morris, Tarzan: a peer-to-peer anonymizing network layer, in: Proc. ACM Conf. Comput., Commun. Secur. (CCS), Washington, DC, 2002, pp. 193–206.
[48] M.J. Freedman, E. Sit, J. Cates, R. Morris, Introducing Tarzan, a peer-to-peer anonymizing network layer, in: Proc. ACM Conf. Comput., Commun. Secur. (CCS), Washingtion, DC, 2003, pp. 193–206.
[49] 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.
[50] J. Domingo-Ferrer, Coprivacy: towards a theory of sustainable privacy, in: Priv. Stat. Databases (PSD), Vol. 6344 of Lecture Notes Comput. Sci. (LNCS), Springer-Verlag, Corfu, Greece, 2010, pp. 258–268.
[51] T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd edition, Wiley, New York, 2006.
[52] 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), Lecture Notes Comput. Sci. (LNCS), Springer-Verlag, Jeju Island, South Korea, 2011, pp. 146–154, invited paper.
[53] E.T. Jaynes, On the rationale of maximum-entropy methods, Proceedings of the IEEE 70 (9) (1982) 939–952.
[54] E.T. Jaynes, Information theory and statistical mechanics II, Physical Review Series II 108 (2) (1957) 171–190.
[55] C.E. Shannon, Communication theory of secrecy systems, Bell System Technical Journal 28 (1949) 656–715.
[56] A. Wyner, The wiretap channel, Bell System Technical Journal 54 (1975).
[57] I. Csiszár, J. Körner, Broadcast channels with confidential messages, IEEE Transactions on Information Theory 24 (1978) 339–348.
[58] C. Díaz, S. Seys, J. Claessens, B. Preneel, Towards measuring anonymity, in: Proc. Workshop Priv. Enhanc. Technol. (PET), Vol. 2482 of Lecture Notes Comput. Sci. (LNCS), Springer-Verlag, 2002, pp. 54–68.
[59] C. Díaz, Anonymity and privacy in electronic services, Ph.D. thesis, Katholieke Univ. Leuven (Dec. 2005).
[60] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, UK, 2004.
[61] L. Cottrell, Mixmaster and remailer attacks. URL http://obscura.com/loki/remailer/remailer-essay.html, 1994.
[62] A. Serjantov, R.E. Newman, On the anonymity of timed pool mixes, in: Proc. Workshop Priv., Anon. Issues Netw., Distrib. Syst, Kluwer, 2003, pp. 427–434.
[63] D. Kesdogan, J. Egner, R. Büschkes, Stop-and-go mixes: providing probabilistic anonymity in an open system, in: Proc. Inform. Hiding Workshop (IH), Springer-Verlag, 1998, pp. 83–98.
[64] M. Rennhard, B. Plattner, Practical anonymity for the masses with mix-networks, in: Proc. Int. Workshop Enabling Technol.: Infra. Col. Enterprises (WETICE), IEEE Comput. Soc, 2003, pp. 255–260.
[65] G. Danezis, Mix-networks with restricted routes, in: Proc. Workshop Priv. Enhanc. Technol. (PET), Lecture Notes Comput. Sci. (LNCS), 2003, pp. 1–17.
[66] M.K. Reiter, A.D. Rubin, Crowds: anonymity for web transactions, ACM Transactions on Information and System Security 1 (1) (1998) 66–92.
[67] A. Serjantov, R. Dingledine, P. Syverson, From a trickle to a flood: active attacks on several mix types, in: Proc. Inform. Hiding Workshop (IH), Springer-Verlag, 2002, pp. 36–52.

[68] U. Möller, L. Cottrell, P. Palfrader, L. Sassaman, Mixmaster protocol—version 2, internet draft, internet eng. task force. URL http://www.freehaven.net/anonbib/cache/mixmaster-spec.txt, Jul. 2003.
[69] Knowledge and Data Engineering Group, University of Kassel: benchmark folksonomy data from BibSonomy. URL http://www.kde.cs.uni-kassel.de/bibsonomy/dumps, Dec. 2007.
[70] M. Grahl, A. Hotho, G. Stumme, Conceptual clustering of social bookmarking sites, in: Proc. Int. Conf. Knowl. Manage. (I-KNOW), Graz, Austria, 2007, pp. 356–364.
[71] L. Specia, E. Motta, Integrating folksonomies with the semantic web, in: Proc. Int. Semantic Web Conf., 2007, pp. 624–639.
[72] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, G. Stum, Evaluating similarity measures for emergent semantics of social tagging, in: Proc. Int. WWW Conf., ACM, 2009, pp. 641–650.
[73] S.P. Lloyd, Least squares quantization in PCM, IEEE Transactions on Information Theory IT-28 (1982) 129–137.

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 mathematical models dealing with the inherent trade-off between privacy and data utility in information systems.

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 Computer 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 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).

Jose L. Muñoz received the M.S. degree in telecommunication engineering from the Universitat Politècnica de Catalunya (UPC) in 1999 and the Ph.D. degree in 2003. In 2000, he joined the Information Security Group at the Department of Telematics Engineering of the UPC. His research interests include security and privacy in computer networks. Currently, he is an associate professor at the Department of Telematics Engineering of the UPC.

Óscar Esparza received the M.S. degree in telecommunication engineering from the Universitat Politècnica de Catalunya (UPC) in 1999 and the Ph.D. degree in 2004. In 2001, he joined the Information Security Group at the Department of Telematics Engineering of the UPC. His research interests include security and privacy in computer networks and mobile agent platforms. Currently, he is an associate professor at the Department of Telematics Engineering of the UPC.