# Query Profile Obfuscation by Means of Optimal Query Exchange between Users

David Rebollo-Monedero, Jordi Forné, and Josep Domingo-Ferrer, *Fellow*, *IEEE*

**Abstract**—We address the problem of query profile obfuscation by means of partial query exchanges between two users, in order for their profiles of interest to appear distorted to the information provider (database, search engine, etc.). We illustrate a methodology to reach mutual privacy gain, that is, a situation where both users increase their own privacy protection through collaboration in query exchange. To this end, our approach starts with a mathematical formulation, involving the modeling of the users' apparent profiles as probability distributions over categories of interest, and the measure of their privacy as the corresponding Shannon entropy. The question of which query categories to exchange translates into finding optimization variables representing exchange policies, for various optimization objectives based on those entropies, possibly under exchange traffic constraints.

**Index Terms**—Profile obfuscation, private information retrieval, privacy via user collaboration, entropy, information theory.

## 1 Introduction

User profiling is very common in information retrieval systems, for example in Web search engines. The information provider (IP) operating the information retrieval system can derive substantial marketing benefits from improved knowledge of the user interest profiles. For the users themselves, profiling can result in improved and more targeted search results; for example, thanks to profiling, Amazon is able to present to each customer a list of books which are likely to interest her. In the more general setting of electronic commerce, profiling customers by their interests is essential to customer relationship management.

The negative side of user profiling is that the interests and the query history of users may contain information considered as private. For example, if a user has looked up a certain disease, it can be inferred that either the user, or someone close to her, suffers from that disease. This is not an academic speculation; in 2006, 20 million queries submitted by 658,000 users of the AOL search engine were publicly disclosed; AOL claimed queries to be properly protected against user reidentification. Two *New York Times* journalists identified a user after studying the released queries [2].

A number of techniques has been proposed in order to counter profiling, or give users a choice against it, with various degrees of suitability, depending on the scenario of application, in terms of privacy, complexity, traffic overhead, infrastructure requirements, and requirements of trust on various parts involved. An overview of these techniques will be provided later on in the background section, Section 2. At this point, we would like to continue motivating our contribution with a specific collaborative strategy, which shall be the object of our study.

Concretely, we focus here on a generic peer-to-peer (P2P) single-hop system, in which users submit queries to an untrusted IP. A purely conceptual depiction of this system appears in Fig. 1. The IP might attempt to profile users according to their interests, possibly by a semantic analysis that first assigns their queries to predefined categories of interest, and then builds a histogram of interests over an extended period of time. An example scenario where this privacy risk is particularly relevant is undoubtedly Internet search.

In principle, two or more users could exchange a portion of their queries before submitting them, in order to obfuscate their respective interest profiles versus the IP or external observers. For additional security in the special case when collaborating users do not trust each other, queries could be encrypted to be readable only by the IP. Clearly, the strategy described could be combined with many other privacy-protecting mechanisms, such as those based on anonymizing proxies, to further reinforce user privacy. Note, however, that users may feel inclined to prefer one method or combination over another, according to what parties, users or intermediaries, they choose to place their trust on.

There are a number of questions that arise naturally from any attempt of implementation of our proposal, which contemplates the potential for mutual privacy gain. First of all, we need to establish a mathematical model representing each user's behavior or profile of interests, along with a measure of privacy. Once privacy becomes quantifiable, so does any query exchange policy, in terms of the privacy gain for each user, and such policies may even be obtained as a mathematical solution to multiobjective optimization problems.

- *D. Rebollo-Monedero and J. Forné are with the Department of Telematics Engineering, Universitat Politècnica de Catalunya, Campus Nord, Módulo C5, Despacho S102A, C. Jordi Girona 1-3, Barcelona E-08034, Spain. E-mail: {david.rebollo, jforne}@entel.upc.edu.*
- *J. Domingo-Ferrer is with the Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Av. Països Catalans 26, Tarragona E-43007, Catalonia and the UNESCO Chair in Data Privacy. E-mail: josep.domingo@urv.cat.*

*Manuscript received 5 Apr. 2011; revised 16 Dec. 2011; accepted 27 Dec. 2011; published online 17 Jan. 2012.*
*For information on obtaining reprints of this article, please send e-mail to: tdsc@computer.org, and reference IEEECS Log Number TDSCSI-2011-04-0062.*
*Digital Object Identifier no. 10.1109/TDSC.2012.16.*

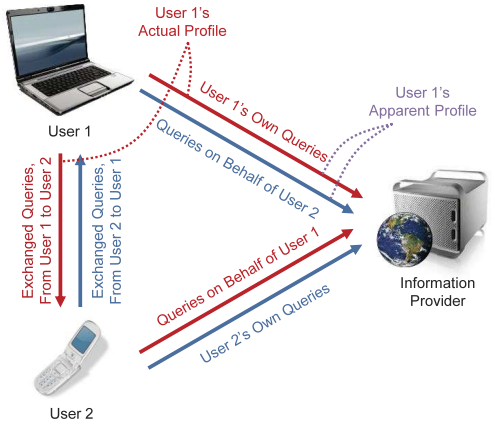



Fig. 1. A conceptual depiction of partial query exchange between two users in order to present an untrusted IP with a distorted observation of their actual profiles of interest.

### 1.1 Contribution and Organization

The intended goal and main contribution of this paper is a first step toward the modeling and analysis of the question of whether mutual privacy gain may be obtained from query exchange, and to what extent, for the special case of two users. In a nutshell, we first model user behavior in an information retrieval system by a histogram of queries assigned to predefined categories. The histogram of each user contains queries that from the point of view of any external observer, including the IP itself, appear to have been created by that user. However, the histogram contains queries originated by this user not exchanged with the other user, and queries submitted on behalf of the other user. Next, we propose to measure the privacy of such histogram, effectively the apparent user profile, as the Shannon entropy of the relative frequencies across the categories. Finally, modeling exchange policies also as categorical histograms, we quantify a number of specific query exchange strategies, and find the optimal tradeoff between the privacy of one user and the other's. The study of these strategies is mainly theoretical, but illustrated numerically, and draws upon concepts and techniques from the fields of information theory and convex and numerical optimization. We must stress that user profile perturbation, by means of query exchange, may be readily integrable with other privacy-protecting techniques in order to reinforce them, particularly pseudonymization.

Beyond the specific metric of privacy adopted or the precise mathematical characterizations and optimizations presented, an important contribution of this paper is the methodology employed, and the illustration of a connection between the still emerging field of privacy and the mature powerful ideas of information theory and convex optimization. Indeed, one of our objectives is to make this methodology, aimed at the systematical analysis of privacy problems, more widely known in the privacy community. An earlier contribution in this direction was the optimization of query forgery methods also for query profile obfuscation in [44], with which this work bears some mathematical analogies at the formulation level.

The necessity of limiting our scope does not allow us to delve, beyond what is essential in our first theoretical approach to the problem, into attacker models, privacy metrics and thoroughly developed practical protocols for its implementation. In a sense, the entropy regions characterized by our theoretical analysis, are humbly reminiscent of Shannon's entropy bounds for lossless data compression, which set the boundaries ideally attainable by practical source codes, albeit ultimately elusive.

The next section, Section 2, is devoted to the state of the art on antiprofiling techniques. Section 3 presents a mathematical formulation of the problem of optimal query exchange, where we argue in favor of Shannon's entropy as a measure of privacy. A number of query exchange strategies are proposed in Section 5, along with several criteria for selecting them in realistic application scenarios. A couple of running examples are examined in Section 6. Section 7 discusses further privacy and security considerations, and conclusions are drawn in Section 8.

## 2 Background

The literature on information retrieval provides numerous solutions to user privacy [16], some of which we touch upon, often extensible to scenarios other than the ones intended. After surveying the state of the art in such antiprofiling techniques, we proceed to review the idea of coprivacy, related to the notion of mutual privacy gain explored in this work.

Cryptographic methods for private information retrieval (PIR) enable a user to privately retrieve the contents of a database, indexed by a memory address sent by the user, in the sense that it is not feasible for the database provider to ascertain which of the entries was retrieved [1], [10], [11], [26], [41]. Simply put, PIR provides protocols whereby the IP does not learn what item the user is retrieving. Yet, these protocols make a number of problematic assumptions, some of which we would like to mention. First, information theoretically secure PIR requires the user to "touch" all items to avoid leaking any clues about the item the user is interested in. Second, the IP is assumed to cooperate in PIR protocols, which is dubious given that the IP normally wishes to profile users. Third, PIR assumes that the IP stores items in a vector and that the user somehow manages to find the address of the desired item. More generally, these protocols are effectively limited to query-response functions in the form of a finite lookup table of precomputed answers. Last but not least, PIR is commonly burdened with a significant computational overhead, although recent progress has been made in this direction [40]. See [19] for a more detailed critique.

Antiprofiling solutions more practical than PIR are directed at obfuscating the user profile. A brief, partial review follows. Simple strategies in the context of Web search merely rely on the Web browser (selectively) rejecting cookies or on using dynamic IP addresses on the user's side. However, rejecting cookies may entail an unacceptable loss of functionality and using dynamic or static addresses is often beyond the user's control.

One of the conceptually simplest approaches to query profile obfuscation consists in including a trusted third party (TTP) acting as an intermediary between the user and the IP, which effectively hides the identity of the user. An

example is GoogleSharing [27], in which a proxy strips a user query of identifying information and submits it to Google under a pseudonym. Then, the proxy returns the query results to the user. Furthermore, GoogleSharing does not need to see the actual queries, which can travel encrypted from the user's browser up to Google. The negative point is that a collusion of GoogleSharing and Google is sufficient to reconstruct the query profile of any user. An appealing twist that does not require that the TTP be online is that of pseudonymizing digital credentials [3], [4], [9]. Recent surveys on anonymous Internet search include [23], [43].

Additional solutions have been proposed, especially in the special case of LBSs, many of them based on an intelligent perturbation of the user coordinates submitted to the provider [22], which, naturally, may lead to an inaccurate answer. Essentially, users may contact an untrusted LBS provider directly, perturbing their location information in order to hinder providers in their efforts to compromise user privacy in terms of location, although clearly not in terms of query contents and activity. This approach, sometimes referred to as obfuscation, presents the inherent tradeoff between data utility and privacy common to any perturbative privacy method. In the context of recommendation systems and general information provision, a user could slightly distort her profile of preferences in order to enjoy a higher degree of privacy, at an acceptable cost in recommendation success or reply accuracy. This strategy can alternatively be regarded as a compromise in lieu of the extreme case when an individual refrains from using an information system for a particular type of queries.

Indeed, an interesting approach to provide a distorted version of a user's profile of interests consists in query forgery. The underlying principle is to accompany original queries or query keywords with bogus ones, in order to preserve user privacy to a certain extent, at the cost of traffic and processing overhead, without the requirement that the user trust the service provider nor the network. Building on this simple principle, several protocols, mainly heuristic, have been proposed and implemented with various degrees of sophistication [23], [33], [49]. A theoretical study of how to optimize the introduction of bogus queries from an information-theoretic perspective, for a fixed constraint on the traffic overhead, appears in [44]. Naturally, the perturbation of user profiles for privacy preservation may be carried out not only by means of insertion of bogus activity, but also by suppression. A dual formulation of the problem on forgery in [44] appears in [42], as suppression. More precisely, the latter work analyzes privacy-preserving tag suppression in the Semantic Web, within a formulation that strives to optimize privacy in an information-theoretic sense, under a constraint on the suppression rate that captures the loss in semantic functionality incurred.

We would like to mention a couple of implementations of query forgery. GooPIR [21] is a standalone system installed on the user's computer which adds bogus keywords to a query before it is sent to a Web search engine; in this way, the search engine does not know the exact interests of the user, but knows that the interests are included in the query. An approach with the same spirit of GooPIR was independently and subsequently proposed in [36]. This approach is based on the so called privacy through plausibly deniable search. When a user wants to submit a target query, the system uses a latent semantic indexing-based approach to generate $k-1$ cover queries on different topics which are equally plausible. TrackMeNot [28] is another standalone system which automatically issues ghost queries to several search engines, in such a way that these do not know whether a query is really being submitted by the user or is rather a ghost query. A similar add-on for a popular browser is [54].

These standalone solutions are attractive because they can be easily bootstrapped. Yet, they are also saddled with shortcomings: for example, GooPIR requires a frequency-indexed thesaurus to generate bogus keywords with frequency similar to the one of the real keywords; privacy through plausibly deniable search requires semantic query processing; TrackMeNot causes bandwidth and computation to be consistently wasted due to the ghost queries and it is vulnerable to identification of real target queries by timing observation, as noted in [36].

An illustrative example of query forgery applied to LBSs is [32]. Query forgery appears also as a component of other privacy protocols, such as the location-based query obfuscation protocol via user collaboration in [45] and [46]. In addition to legal implications, there are a number of technical considerations regarding traffic forgery for privacy [50], as attackers may analyze not only contents but also activity, timing, routing, or any transmission protocol parameters, jointly across several queries or even across diverse information services. In addition, we insist that automated query generation is naturally bound to be frowned upon by network and information providers, thus any practical framework must take into account query overhead.

Another class of solutions against user profiling exploits user collaboration. P2P systems rely on the principle of having the queries of one user submitted by other peer users, in such a way that the search engine does not really know who is interested in a query it receives. Examples of P2P systems are Tor [52] with its Torbutton browser plug-in [53], the Crowds system [47] and proposals like [6], [19], [55]. Within P2P systems, one should distinguish single-hop protocols and multihop protocols. In a single-hop protocol (like [6], [19]), the query originator, say $A$, sends her query to one of her peers, $B$; $B$ may either submit $A$'s query to the information system and return the answer to $A$, or just discard $A$'s query. In a multihop protocol, such as Crowds, when $B$ receives a query from $A$, $B$ may either submit $A$'s query or forward it to another peer $C$, and so on, until the query reaches a peer who decides to submit it. Single-hop systems have the advantage over multihop that less peers see the queries that are submitted; however, they have the problem that peer $B$ knows that the query corresponds to $A$'s interests. Multihop systems have the symmetrical advantage and shortcoming: peer $B$ does not really know whether the received query corresponds to $A$'s or someone else's interests, but more peers see the generated queries (so more people know that a certain topic interests someone, even if they do not know who is interested).

The notion of mutual privacy gain we explore in this work bears certain similarity with the concept of *coprivacy*,

introduced for a P2P community in [17] and [18]. Coprivacy was defined as a situation where the best strategy for a peer to preserve her privacy is to help another peer in preserving his privacy. The advantage of coprivate protocols is that they make privacy preservation of each specific individual a goal that interests other individuals: therefore, privacy preservation becomes more attractive and hence easier to achieve and more sustainable. The concept was formalized in a game-theoretic fashion by defining privacy utilities for each peer and defining coprivacy as a Nash equilibrium [37], [39] between two players (peers). Prior to the introduction of the term coprivacy, [25] resorted to game theory to analyze noncooperative strategies of mobile nodes changing pseudonyms, at a cost, to preserve their location privacy. More recently, coprivacy was illustrated in [18] and [20] in the setting of anonymous keyword search, that is, information retrieval without profiling.

Unlike [17], which uses privacy metrics based on euclidean distances, and similarly to [18], [20], we use Shannon's entropies to measure privacy. There exists a point of coincidence between the work in [20] and the one developed in this paper. Namely, the on-the-fly strategy for query exchange that we shall describe in Section 5, as one among several alternatives. In that regard, it is important to stress that such strategy is formulated here for two users only, and thus the sophisticated peer selection method of [20], designed for the more general case of an arbitrary number of users, does not apply. On the other hand, the methodology, theoretical analysis, and experimental results presented in this manuscript also contemplate a more general framework that extends beyond the aforementioned strategy.

From a more conceptual perspective, an important point of divergence stems from the fact that the game-theoretic approach favors Nash equilibria, while we adhere to the approach of theoretical and numerical joint optimization in most cases, occasionally under traffic constraints. Our theoretical results, for instance, characterize the region of entropies derived from all possible query exchange strategies, and provide a closed-form solution to the maximization of the minimum privacy, with and without respect to the level of privacy of the original profiles. Still, it will be interesting to capture a small but insightful portion of the work in [20] within our analysis, for the purposes of strategy comparison, which we shall report in Section 6.

## 3 Problem Formulation

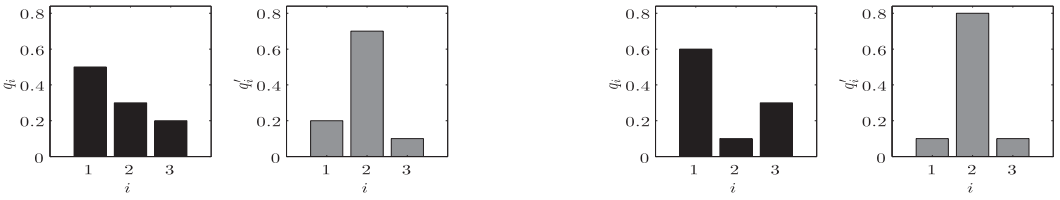
As stated in Section 1.1 and conceptually depicted in Fig. 1, our work addresses the problem of privacy protection against user profiling, by means of altering user profiles from the perspective of an external attacker, exploiting the possibility of query exchange among users. In the following sections, we present our assumptions and propose a measure of privacy. All of which will finally enable us to formulate the problem of query exchange mathematically, as an optimization problem whose objective is our measure of privacy, and the variables represent a particular choice of exchange policy.

### 3.1 Preliminary Assumptions and Attacker Model

A tractable model of a user's activity with regard to profiling used, for example, in [17], [18], [20], [42], and [44], consists in representing profiles by histograms of relative frequencies of queries within a predefined set of categories of interest. In practice, this theoretical model would involve the establishment of such categories and a categorization procedure, which will impact the quantification of privacy [24], [33], [42]. The histogram that a user would originate by submitting queries directly to the IP, if privacy were not a concern, will be called *actual profile*. As this user decides to cooperate with another to partly exchange queries, an external observer, such as the IP itself, could only retrieve a perturbed version of this histogram, containing some queries originated by the user in question, and some submitted on behalf of the other. We shall call the resulting perturbed histogram *apparent profile*. This was informally depicted in Fig. 1.

In this paper, we formalize the problem of mutual privacy gain from a novel perspective that will enable us to tackle it systematically, by establishing privacy metrics, casting it as an optimization problem and analyzing its solution. Not to be overly ambitious, because the scope of our contribution must be necessarily limited, we now concordantly present a series of security and privacy assumptions, sufficiently adequate for a preliminary study of the problem of profile obfuscation by means of query exchange between users. A more detailed discussion on such assumptions is provided later on in Section 7, where we skirt around the edges of the scope of our work to glance at its applicability to more intricate scenarios.

1. Clearly, a protocol is required in order for users to agree to partly exchange their queries. Such group formation protocols, briefly discussed in Section 7, abound in the literature and fall outside the intended focus of our work.
2. We shall assume that users follow one of several variations of certain query exchange protocols, detailed in Section 5. Some of these will address the case of cooperating albeit curious users who also pose a privacy risk, as peers may not wish to entrust them with a partial view of their actual profile, consisting in the exchanged queries. We shall postpone, until Section 7, the consideration of denial-of-service (DoS) attacks and of dishonest users colluding with the IP to jointly infer the actual profile of a user they falsely claim to cooperate with.
3. Our privacy attacker model contemplates an observer of the apparent profiles after query exchange, particularly the IP itself. For each user submitting queries, we assume that the attacker does not know whether that particular user is simply submitting queries individually, or using a query exchange protocol, and, in the latter case, who is the cooperating peer, and which queries are being exchanged. In practice, this assumption may require certain network traffic analysis countermeasures put in place, as we shall discuss in Section 7. Finally, suppose further that given this lack of knowledge (especially if the probability of users cooperating to exchange queries is small) the privacy attacker decides to regard the apparent profile of the user in question as if it were the actual one.




(a) Case A. $q = (0.5, 0.3, 0.2)$, $q' = (0.2, 0.7, 0.1)$, $\mathrm{H}(q) \simeq 1.49$, $\mathrm{H}(q') \simeq 1.16$.

(b) Case B. $q = (0.6, 0.1, 0.3)$, $q' = (0.1, 0.8, 0.1)$, $\mathrm{H}(q) \simeq 1.30$ and $\mathrm{H}(q') \simeq 0.92$.

Fig. 2. Examples of pairs of users represented by their relative histograms $q$, $q'$ of interests across $n = 3$ categories, with activity rates $\alpha = 1 = \alpha'$. The maximum entropy attainable is $\log_2 3 \simeq 1.59$.

### 3.2 User Query Exchange as a Privacy Mechanism

In more precise, mathematical terms, two users submit queries to an IP, which we abstractly represent by samples in a predefined set of $n$ *categories* indexed by $i = 1, \ldots, n$. Accordingly, the behavior of the first user, recorded over a given period of time, is modeled by a frequency histogram, specifically by a relative frequency histogram or probability mass function (PMF) $q$, and an *activity* parameter $\alpha > 0$, such that $\alpha q$ gives the absolute frequency histogram across the $n$ categories. Similarly, the *profile* of the second user is given by an entirely analogous activity parameter $\alpha'$ and a PMF $q'$.

Throughout the paper, we shall resort to a couple of simple albeit insightful running examples, corresponding to a first pair of users A represented by

$$q = (0.5, 0.3, 0.2) \text{ and } q' = (0.2, 0.7, 0.1), \tag{1}$$

and a second pair of users B modeled by

$$q = (0.6, 0.1, 0.3) \text{ and } q' = (0.1, 0.8, 0.1). \tag{2}$$

Although the examples are synthetic, the $n = 3$ categories could very well reflect interests across categories such as business, technology, and health. The relative histograms are shown in Fig. 2. For the first pair of users A, which we may view as *"unbalanced,"* the particular distribution of interests appears to make it difficult to compensate a shared lack of interest in category 3. However, the second case B, more *"balanced,"* we intentionally compensated the lack of interest of user one in category 2 with a strong interest of user two, so that a fairly symmetrical exchange should lead to fairly uniform apparent profiles and allegedly to a significant mutual privacy gain. In the numerical examples, we shall use identical activity rates $\alpha = 1 = \alpha'$, as unequal values lead to fairly similar results.

Define $r$ and $r'$ as the vectors containing the relative histograms of queries exchanged from the first user to the second and vice versa, respectively. Let $s$ and $s'$ represent the *apparent* relative profiles from the point of view of an external observer, also PMFs. We have

$$s = \frac{\alpha(q - r) + \alpha' r'}{\alpha(1 - \sum r_i) + \alpha' \sum r'_i}, \quad s' = \frac{\alpha'(q' - r') + \alpha r}{\alpha'(1 - \sum r'_i) + \alpha \sum r_i}.$$

Clearly, the *exchange policies* $r$ and $r'$ must satisfy the constraints $0 \le r_i \le q_i$ and $0 \le r'_i \le q'_i$ for all $i$, which we write more compactly as $0 \le r \le q$ and $0 \le r' \le q'$. Numerical examples of apparent profiles $s$ and $s'$ and exchange policies $r$ and $r'$ will be supplied in Section 6, such as those in Figs. 6 and 7.

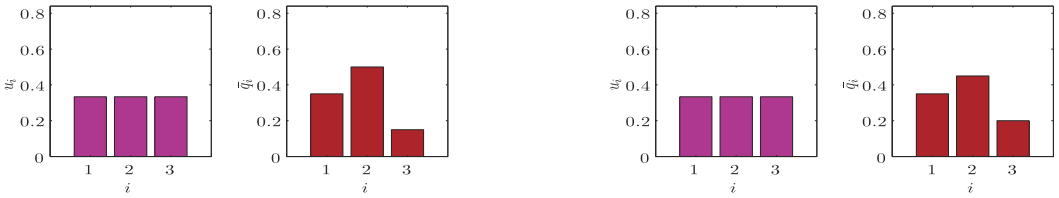
### 3.3 Measuring Privacy

We mentioned in Section 2 that alternative mechanisms to query exchange included query forgery [44], e.g., applied to Internet search, and tag suppression [42] for the Semantic Web. Whether using query forgery, suppression or exchange, one must strive to optimize privacy through data perturbation, under a constraint on traffic overhead, data utility, or any loss of system functionality. But in order to select a specific query exchange, forgery or suppression strategy that numerically optimizes privacy, we must necessarily equip the model assumed with a quantitative measure of privacy. In this work, just as in [42] and [44], we propose to quantify the privacy of the apparent user profiles $s$ and $s'$ observed by the attacker by an information-theoretic quantity, namely, by the Shannon entropies, $\mathrm{H}(s)$ and $\mathrm{H}(s')$, respectively.

Recall [12] that the Shannon entropy $\mathrm{H}(s)$ of a PMF $s$ is defined as $\mathrm{H}(s) = -\sum_{i=1}^{n} s_i \log s_i$ (and similarly for $s'$), that it is a measure of the uncertainty of the outcome of a random variable distributed according to such PMF, and that it is maximized, among all distributions on $\{1, \ldots, n\}$, by the uniform distribution $u_i = 1/n$ for all $i$, for which $\mathrm{H}(u) = \log n$. Commonly, the basis of the logarithm is chosen to be 2 and concordantly the entropy units are bits. In the running examples of Fig. 2, the starting uncertainties corresponding to the original profiles are $\mathrm{H}(q) \simeq 1.49$ and $\mathrm{H}(q') \simeq 1.16$ bit in case A, and $\mathrm{H}(q) \simeq 1.30$ and $\mathrm{H}(q') \simeq 0.92$ bit in case B. The maximum entropy attainable is $\log_2 3 \simeq 1.59$ bit.

Having established a measure of privacy, we are ready to tackle, both mathematically in Section 4 and numerically in Section 6, the problem of finding a query exchange policy. In more formal terms, we are faced with the selection of such exchange policy, represented by the choice of variables $r$ and $r'$, in order to maximize the interdependent privacy measures $\mathrm{H}(s)$, $\mathrm{H}(s')$ in a way that will be made precise shortly, given the user profile models represented by the activities $\alpha$, $\alpha'$, and the relative histograms $q$, $q'$, across the predefined query categories $i = 1, \ldots, n$.

Before proceeding any further, we would like to justify our choice of Shannon's entropy as a measure of the privacy of a user profile. The use of entropy as a measure of privacy, in the widest sense of the term, is by no means new. Shannon's work in the fifties introduced the concept of *equivocation* as the conditional entropy of a private message given an observed cryptogram [48] as a measure of confidentiality. More recent studies [15] rescue the suitable applicability of the concept of entropy as a measure of




(a) Case A. $\bar{q} = (0.35, 0.5, 0.15)$, $\mathrm{H}(\bar{q}) \simeq 1.44$.

(b) Case B. $\bar{q} = (0.35, 0.45, 0.2)$, $\mathrm{H}(\bar{q}) \simeq 1.51$.

Fig. 3. Uniform profile $u$ and group profile $\bar{q}$.

privacy, by proposing to measure the degree of anonymity observable by an attacker as the entropy of the probability distribution of possible senders of a given message.

In the context of this paper, an intuitive justification in favor of entropy maximization is that it boils down to making the apparent user profile as uniform as possible, thereby hiding a user's particular bias toward certain categories of interest. A richer argumentation stems from Jaynes' rationale behind entropy maximization methods [30], [31], partly motivated by the celebrated spectral estimation method postulated by Burg, and more generally understood under the beautiful perspective of the method of types and large deviation theory [12, Section 11].

Under Jaynes' rationale on entropy maximization methods, the entropy of an apparent user profile, modeled by a relative frequency histogram of categorized queries, may be regarded as a measure of privacy, or perhaps more accurately, anonymity. The leading idea is that the method of types from information theory establishes an approximate monotonic relationship between the likelihood of a PMF in a stochastic system and its entropy. Loosely speaking and in our context, the higher the entropy of a profile, the more likely it is, the more users behave similarly. This is in absence of a probability distribution model for the PMFs, viewed abstractly as random variables themselves. Under this interpretation, entropy is a measure of anonymity *not* in the sense that the user's identity remains unknown but only in the sense that higher likelihood of an apparent profile believed by an external observer to be the actual profile makes that profile more common, hopefully helping the user go unnoticed, less interesting to an attacker assumed to strive to target peculiar users.

If an aggregated histogram of the population were available as a reference profile, the extension of Jaynes' argument to relative entropy, that is, to the Kullback-Leibler (KL) divergence, would also give an acceptable measure of privacy (or anonymity). Recall [12] that KL divergence is a measure of discrepancy between probability distributions, which includes Shannon's entropy as the special case when the reference distribution is uniform. In fact, KL divergence was used in [42] and [44] as a generalization of entropy to measure privacy.

## 4 Theoretical Analysis

We devote this section to a partial characterization of the privacy region, that is, the region of pairs of entropy values associated with pairs of possible apparent profiles. While partial, this characterization will suffice to identify the optimal exchange policies when our objective is to maximize, either the minimum privacy or the minimum privacy gain with respect to the original values.

We shall assume that for each category $i = 1, \ldots, n$, at least $q_i$ or $q'_i$ is positive. In other words, we shall assume that $n$ is the number of active categories. In practice, no query exchange policy can modify the complete absence of activity within a category in which none of the users involved shows any interest. On a different note, full traffic exchange will not be contemplated, as one of the apparent profiles would have zero absolute activity, and its entropy would become undefined. Instead, later in this section, we shall consider the case of arbitrarily low activity on a uniformly distributed profile, which formally maximizes the entropy. For practical purposes, one may regard these two situations not too dissimilarly.

Let $\mathcal{S}$ denote the region of possible pairs of apparent profiles $(s, s')$, as defined in Section 3.2. That is,

$$\mathcal{S} = \{(s, s') \mid 0 \leq r \leq q,\ 0 \leq r' \leq q'\}.$$

A pair of exchange policy vectors $r$ and $r'$ will be called *feasible* when the defining constraints are met. Define the *privacy region* as the region of possible entropy pairs

$$\mathcal{H} = \{(\mathrm{H}(s), \mathrm{H}(s')) \mid (s, s') \in \mathcal{S}\}.$$

The term *mutual privacy gain* is used here to refer to a nonnegative entropy gain of both users. Accordingly, define the following region:

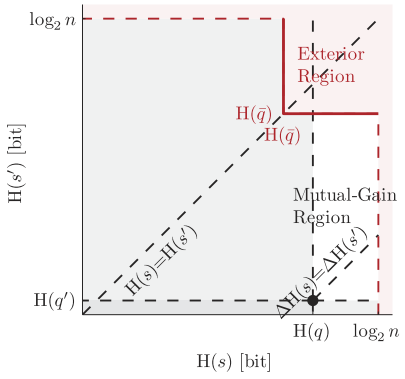
$$\mathcal{C} = \{(H, H') \in \mathcal{H} \mid H \geq \mathrm{H}(q),\ H' \geq \mathrm{H}(q')\}. \tag{3}$$

We shall denote the *uniform distribution* by $u$. Define the *group profile*

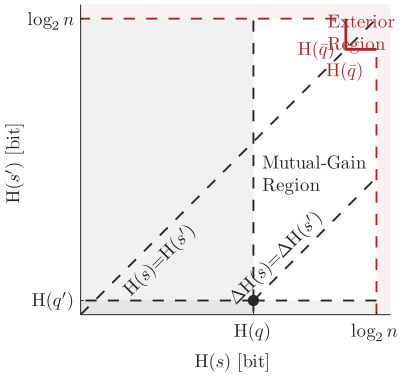
$$\bar{q} = \frac{\alpha}{\alpha + \alpha'}\, q + \frac{\alpha'}{\alpha + \alpha'}\, q',$$

that is, a convex combination of the individual user profiles weighted according to their activity. The uniform profile $u$ and the group profile $\bar{q}$ are illustrated in Fig. 3 with our running example of user profiles $q$ and $q'$, started in Section 3.2.

Our first proposition asserts that an equivalent, slightly simpler characterization of the region $\mathcal{S}$ can be obtained by merging the unidirectional exchange policy vectors $r$, $r'$ into a single bidirectional exchange policy $d = -\alpha r + \alpha' r'$, under the constraint $-\alpha q \leq d \leq \alpha' q'$. The components $d_i$ of the bidirectional exchange represent net exchanges, canceling out opposite exchanges of equal magnitude in $r_i$ and $r'_i$. This simpler characterization not only will occasionally facilitate the proofs, but will also reduce the number of variables in some of the associated optimization problems, proposed in Section 5.2, and numerically solved in Section 6.




(a) Case A.




(b) Case B.

Fig. 4. Privacy region.

**Proposition 1 (Bidirectional Exchange).**

$$\mathcal{S} = \{(s, s') \mid -\alpha q \le d \le \alpha' q'\},$$

*with*

$$s = \frac{\alpha q + d}{\alpha + \sum d_i}, \quad s' = \frac{\alpha' q' - d}{\alpha' - \sum d_i}.$$

**Proof.** It follows directly from the definition of $d$ that the constraints $0 \le r \le q$ and $0 \le r' \le q'$ imply the constraint $-\alpha q \le d \le \alpha' q'$. Conversely, for any feasible $d$, set $r_i = -d_i/\alpha$ and $r'_i = 0$ whenever $d_i < 0$, and $r_i = 0$ and $r'_i = d_i/\alpha'$ otherwise. □

Our next proposition confirms a number of fairly intuitive symmetrical properties satisfied by the possible apparent profiles.

**Proposition 2 (Symmetry).** *For any $(s, s') \in \mathcal{S}$ induced by (feasible) exchange policies $r$, $r'$,*

1. $\frac{s+s'}{2} = \bar{q}$,
2. $(s', s) \in \mathcal{S}$, *achieved with exchange policies* $q - r$, $q' - r'$, *and*
3. $(\bar{q}, \bar{q}) \in \mathcal{S}$, *achieved with exchange policies* $\frac{1}{2}q$, $\frac{1}{2}q'$.

**Proof.** In terms of the simplified characterization of Proposition 1,

$$\frac{s + s'}{2} = \frac{\alpha q + d + \alpha' q' - d}{\alpha + \sum_i d_i + \alpha' - \sum_i d_i} = \bar{q}.$$

The results concerning achievability follow immediately from the formulas for $s$ and $s'$ in terms of $r$ and $r'$, or the equivalent one in terms of $d$. □

Observe that $q = q'$ would imply $q = q' = \bar{q}$. For any $q$ and $q'$, the following proposition states that the group profile $\bar{q}$ is in fact the only point where the apparent profiles $s$ and $s'$ may coincide.

**Proposition 3 (Equal Profiles).** *$s = s'$ if and only if $s = \bar{q} = s'$. Either case is equivalent to*

$$(\alpha + \alpha')\, d = -\Big(\alpha' - \sum d_i\Big)\alpha q + \Big(\alpha + \sum d_i\Big)\alpha' q'.$$

**Proof.** On account of Proposition 2, provided that $s = s'$, clearly $s = \frac{s+s'}{2} = \bar{q}$. The second part of the proposition follows from routine algebraic manipulation, after writing $s = s'$ in terms of the equivalent formulation with a bidirectional exchange $d$. □

In the following, we shall consider two specific optimization criteria. On the one hand, the *maximin* criterion, which we define to correspond to maximizing the smallest privacy. In mathematical terms, $\max \min\{\mathrm{H}(s), \mathrm{H}(s')\}$. On the other hand, we consider the *maximin-gain* criterion, which entails the maximization of the smallest privacy gain,

$$\max \min\{\mathrm{H}(s) - \mathrm{H}(q), \mathrm{H}(s') - \mathrm{H}(q')\}.$$

We are now equipped to partly characterize the privacy region, to an extent that will enable us to find the maximum of the minimum among the entropies, and the entropy gains. The corresponding results are gathered in the next three theorems, and depicted in Fig. 4.

**Theorem 4 (Privacy Region, Maximin).**

1. $\mathrm{H}(s), \mathrm{H}(s') \le \mathrm{H}(u) = \log n$.
2. $s = s' = \bar{q}$ *maximize* $\min\{\mathrm{H}(s), \mathrm{H}(s')\}$.
3. *The closure of $\mathcal{H}$ contains the two segments connecting $(\mathrm{H}(\bar{q}), \mathrm{H}(\bar{q}))$ with $(\mathrm{H}(u), \mathrm{H}(\bar{q}))$ and with $(\mathrm{H}(\bar{q}), \mathrm{H}(u))$. Any of the points along these segments maximizes the minimum entropy over the closure.*

**Proof.** The first statement of the theorem is an immediate consequence of the fact that the uniform distribution maximizes the entropy. To prove the second statement, observe that Proposition 2 implies that $\mathcal{H}$ is symmetrical around the bisector $\mathrm{H}(s) = \mathrm{H}(s')$. Therefore, without loss of generality, one may assume that the solution to the maximization of $\min\{\mathrm{H}(s), \mathrm{H}(s')\}$ satisfies the constraint $\mathrm{H}(s') \le \mathrm{H}(s)$. In other words, the maximization of the minimum is equivalent to the maximization of $\mathrm{H}(s')$ subject to that constraint. Let $(s, s')$ be a solution to the latter equivalent maximization problem. Recall from Proposition 2 that $\bar{q} = \frac{s+s'}{2}$, and that $(\bar{q}, \bar{q}) \in \mathcal{S}$. We claim that $(\bar{q}, \bar{q})$ is also a solution to the latter maximization problem. First, this solution trivially satisfies the constraint. Second, the concavity of the entropy and the constraint of the maximization problem guarantee

that $\mathrm{H}(\bar{q}) \geq \frac{1}{2}\mathrm{H}(s) + \frac{1}{2}\mathrm{H}(s') \geq \mathrm{H}(s')$. Because $(s, s')$ was assumed to maximize $\mathrm{H}(s')$, this means that $(\bar{q}, \bar{q})$ must also yield a maximum, as claimed.

Regarding the third statement of the theorem, because the role of the two users is interchangeable, it suffices to prove only the inclusion of one of the segments, for example, that connecting $(\mathrm{H}(\bar{q}), \mathrm{H}(\bar{q})$ with $(\mathrm{H}(u), \mathrm{H}(\bar{q}))$. We assumed at the beginning of this section that for all $i$, either $q_i$ or $q'_i$ was positive. We show that for any distribution $t$, the entropy pair $(\mathrm{H}(t), \mathrm{H}(\bar{q}))$ belongs to the closure of the entropy region $\mathcal{H}$. To see this, consider that all queries from user one are sent to user two, except for an arbitrarily small residual that leaves user one with a histogram of relative frequencies $t$, with arbitrarily small absolute frequencies, and user two with a profile arbitrarily close to $\frac{s+s'}{2} = \bar{q}$, responsible for nearly the entirety of the submitted queries.

Lastly, any entropy pair on the two segments mentioned in the theorem maximizes the minimum, because any of these points have both components greater than or equal to $\bar{q}$, and we have shown $(\bar{q}, \bar{q})$ to be a solution. A graphical interpretation of this last fact is that the set of points in the entropy plane sharing a common minimum form a right angle with vertex along the bisector of equal entropies, and that maximization corresponds to sliding the angle along the bisector. □

Theorem 4(2), interpreted under Proposition 2, asserts that $r = \frac{1}{2}q$ and $r' = \frac{1}{2}q'$ are a solution to the maximization of the minimum of the entropies. Put it simply, the theorem states, somewhat surprisingly, that a symmetrical exchange at half rate, regardless of the specific category a query belongs to, offers the "best worst" privacy, a strategy implementable even if queries are encrypted for the IP. This result is confirmed numerically later on in Section 6. In light of Proposition 3, we may regard the solution $(\bar{q}, \bar{q})$ as the sole point of coincidence between the apparent profiles. Further, careful inspection of the proof of Theorem 4 leads us to conclude that the result holds for any concave measure of privacy, or for a convex measure to be minimized rather than maximized, such as the relative entropy or KL divergence used in [42] and [44]. Finally, item (3) states that the solution is not unique. In this regard, observe that we may feel inclined toward one or another according to their exchange traffic demands. We will return to the issue of traffic constraints later on in the numerical examples of Section 6.

Unfortunately, the maximum minimum entropy need not lie in the region $\mathcal{C}$ of mutual privacy gain. The last result of the section concerns the maximization of the minimum privacy gain with respect to the original profiles. While Theorem 5 seems to provide the be-all, end-all solution, bear in mind that that solution will require nearly full traffic exchange.

**Theorem 5 (Maximin-Gain).** *Suppose without loss of generality that* $\mathrm{H}(q') \leq \mathrm{H}(q)$. *Then* $s = u$ *and* $s' = \bar{q}$ *belong to the closure of the attainable region* $\mathcal{S}$, *maximize* $\min\{\mathrm{H}(s) - \mathrm{H}(q), \mathrm{H}(s') - \mathrm{H}(q')\}$, *and also* $\min\{\mathrm{H}(s), \mathrm{H}(s')\}$, *over the closure.*

**Proof.** The last claim, namely that $s = u$ and $s' = \bar{q}$ are a solution to the maximization of the absolute minimum over the closure, not the minimum gain, is a particularization of Theorem 4(3).

Two proofs are provided for the claim regarding the maximization of the minimum gain. The simplest argument is a graphical one, similar to the interpretation at the end of the proof of Theorem 4. The key idea consists in realizing that the set of points in the entropy plane that yield a common minimum gain are shaped as a right angle with vertex along the bisector of equal gains, and that maximization corresponds to sliding this angle along the bisector, within the region established in Theorem 4, and depicted in Fig. 4.

A more formal argument follows. For compactness, here we denote $\mathrm{H}(a) \leq \mathrm{H}(b)$ as $a \preceq b$, redefine $\min$ under this order relation, and write $\mathrm{H}(a) - \mathrm{H}(b)$ simply as $a - b$. Now in terms this notation, we assumed, without loss of generality, that $q' \preceq q$, and Theorem 4 entails that $s, s' \preceq u$, and that $\min\{s, s'\} \preceq \bar{q}$. Because we know $s = u$ and $s' = \bar{q}$ to be achievable, it suffices to show that

$$\min\{s - q,\ s' - q'\} \preceq \min\{u - q,\ \bar{q} - q'\}.$$

To this end, observe that $\min\{s, s'\} \preceq \bar{q}$ is equivalent to stating the disjunction $s \preceq \bar{q}$ or $s' \preceq \bar{q}$. What we need to prove is that $s - q$ or $s' - q' \preceq u - q$, and that $s - q$ or $s' - q' \preceq \bar{q} - q'$. The first part of this statement is a direct consequence of the fact that $s \preceq u$. For the second part, consider the two cases $s$ or $s' \preceq \bar{q}$. In the first case, $s \preceq \bar{q}$ and $q' \preceq q$ imply $s - q \preceq \bar{q} - q'$. In the second case, $s' \preceq \bar{q}$ implies $s' - q' \preceq \bar{q} - q'$. □

## 5 Query Exchange Strategies

In this section, we define three classification criteria, whose combinations lead us to propose various query exchange strategies. These classification criteria are formulated in a practical manner, taking into account privacy, user trust and traffic overhead constraints. The suitability of these criteria should be easily assessed under the specific requirements of a particular field of application, which should in turn suggest feasible choices of query exchange strategies.

### 5.1 Classification Criteria

The three aforementioned classification criteria are as follows:

1. We shall first consider the two *privacy criteria* already introduced in the theoretical analysis of Section 4 as optimization objectives, namely, (absolute) *maximin* entropy on the one hand, and *maximin-gain* with respect to the original entropies, on the other. The choice of one or the other may be a matter of user agreement or system design.
2. Next, we contemplate the level of *user trust* with regard to the disclosure of their mutual interests. That is, users may wish to keep their profile of interests hidden not only from the IP, but also from each other. In this case, they may wish to encrypt their queries for the IP prior to exchanging them, making it impossible for a user to see the category an

exchanged query belongs to. The following (simplification of a) cryptographic protocol shows how user two can submit a query to the IP, on behalf of 1:

$$1 \rightarrow 2 \rightarrow \text{IP} : \text{E}_{U_{\text{IP}}}(\text{query},\ K_1),$$
$$\text{IP} \rightarrow 2 \rightarrow 1 : \text{E}_{K_1}(\text{SIGN}_{R_{\text{IP}}}(\text{query},\ \text{reply})),$$

where $U_{\text{IP}}$ and $R_{\text{IP}}$ represent the public and private key of the IP, respectively, $K_1$ a symmetric session key proposed by user one, used once to prevent pseudoidentification, and E and SIGN denote encryption and digital signature, respectively.

3. Finally, we must give due regard to the possibility of *traffic constraints*, that is, constraints on the amount of traffic overhead resulting from the exchange of queries. In terms of our mathematical formulation, we can impose constraints of the form

$$\alpha \sum_i r_i \leq D, \quad \alpha' \sum_i r'_i \leq D', \tag{4}$$

for given target traffic maxima $D$ and $D'$ in queries per time unit in each direction.

### 5.2 Proposed Strategies

We propose the following query exchange strategies or protocols, which are motivated by the theoretical analysis of Section 4, and described in terms of the classification criteria just established in Section 5.1.

- *Symmetrical exchange*. A fraction of encrypted queries is exchanged regardless of category, in equal absolute number in each direction, according to the traffic constraint. As the category needs not be seen, users do not need to trust each other.
- *Maximin exchange*. Provided that the traffic constraints determined by $D$ and $D'$ are sufficiently lax, precisely, $D \geq \frac{1}{2}\alpha$ and $D' \geq \frac{1}{2}\alpha'$, recall from our theoretical analysis that the exchange policy $r = \frac{1}{2}q$, $r' = \frac{1}{2}q'$ produces $\text{H}(s) = \text{H}(s') = \text{H}(\bar{q})$, and it is optimal according to the maximin criterion. In practice, this would correspond to each user exchanging half of the queries they generate, regardless of category, encrypted in the case of untrusted users. When $\alpha = \alpha'$, this coincides with the symmetrical exchange policy at half rate. With more severe traffic constraints, we would strive to maximize the minimum entropy subject to these constraints. To solve this constrained optimization problem, the original profiles of interest $q$ and $q'$ must be known by both trusted users, or by a trusted third party communicating the exchange policy.
- *Maximin-gain exchange*. Suppose that $\text{H}(q) \leq \text{H}(q')$. The opposite case is entirely analogous. In the absence of traffic constraints, recall from the theoretical analysis that nearly full exchange from the first user to the second arbitrarily approximates $\text{H}(s) = \text{H}(u)$ and $\text{H}(s') = \text{H}(\bar{q})$, which is optimal according to the maximin-gain criterion. The second user does not need to know the category of the encrypted queries to be submitted on behalf of the first user. Recall that the endpoint of the critical boundary, with components $\text{H}(u)$ and $\text{H}(\bar{q})$, is optimal both according to the maximin and the maximin-gain criterion, but the traffic requirement is fairly steep. With traffic constraints, however, we seek to maximize the minimum gain, subject to these constraints. To solve the constrained optimization problem, the original profiles of interest $q$ and $q'$ must be known by both trusted users, or by a trusted third party communicating the exchange policy.
- *On-the-fly*. Users trust each other with the category their queries belong to. A user generates a query, and maintains a histogram of sent queries, whether they were generated by that user or sent on behalf of the other. If, by sending the current query, the entropy would decrease, then the other user is asked to submit it instead. The other user accepts only if her own entropy would increase by doing so. Traffic constraints can optionally be implemented. We shall see in our experiments that this strategy approaches the privacy performance of the maximin-gain exchange. With regard to its application in the case of traffic constraints, note that it requires a lower level of trust than that of the maximin-gain strategy, because only the categories of the queries requested to be exchanged are disclosed, not the entire profile.

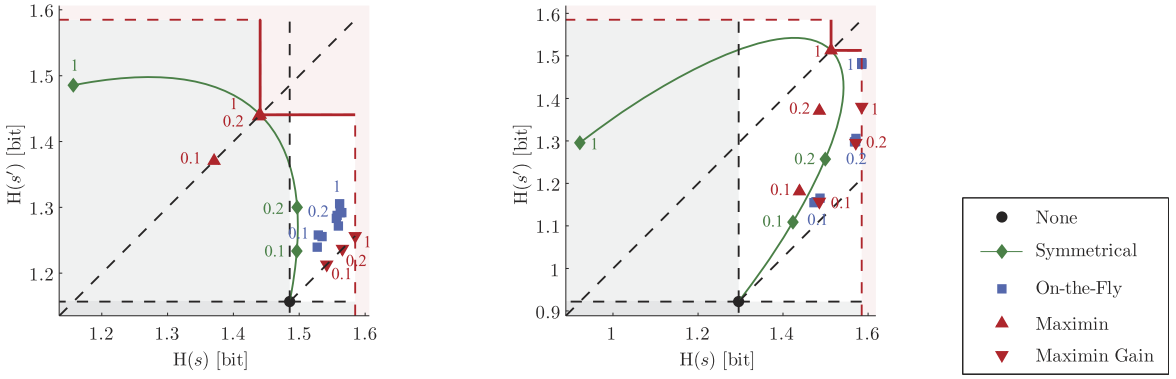
We mentioned at the end of Section 3.1 that a practical implementation of these protocols may require certain added measures aimed to enforce them. Consider, for example, the special case of symmetrical strategy with encrypted queries, and suppose that we wish to avoid the structural cost of a trusted intermediary enforcing the protocol. In order to ensure that the selection of queries to be forwarded is truly independent of their category, one user may wish to request the other to have all queries sent, in encrypted form, so that the receiver not the sender is the user who shall choose which queries to actually forward.

## 6 Numerical Examples

In this section, we apply the strategies proposed in Section 5 to the intuitive pair of running examples of user profiles defined by (1) and (2) in Section 3.2. The activity rates were set to $\alpha = 1 = \alpha'$ queries per time unit. All strategies are computed for three cases, namely the case without traffic constraints, and two cases with traffic constraints (4) determined by $D = D' = 0.1$ and 0.2.

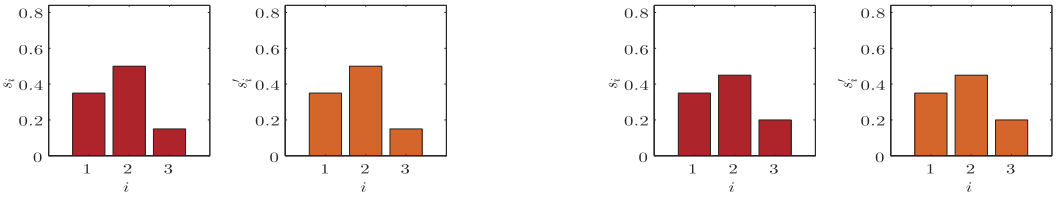
Finding the numerical solution to the constrained optimization problems is a fairly straightforward application of the function `fmincon` implemented in Matlab, based on an interior-point method [5], hardly requiring a split second on a modern computer. The on-the-fly strategy was simulated three times for each of the three constraints, each time using an independent generation of 10,000 queries for each user, with random categories distributed according to their profiles, at times chosen uniformly on an interval of 10,000 time units.

Fig. 5 plots the performance in the privacy plane of all of these strategies. In order to verify the theoretical analysis of Section 4, the characterization of the entropy region of Fig. 4




(a) Case A. (b) Case B.

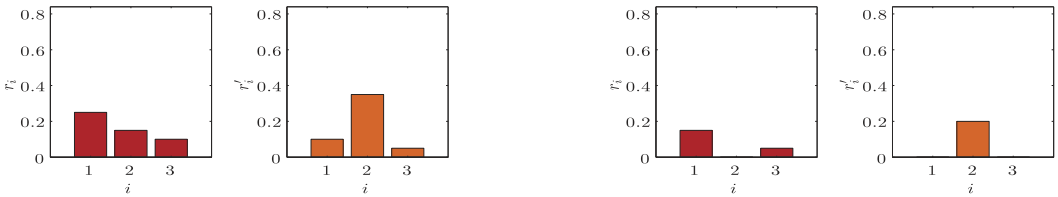
Fig. 5. Numerical computation of all the strategies for traffic constraints $D = D' = 0.1, 0.2, 1$.



(a) Case A. $\bar{q} = (0.35, 0.5, 0.15)$, $\mathrm{H}(\bar{q}) \simeq 1.44$, $\mathrm{H}(\bar{q}) - \mathrm{H}(q) \simeq -0.045$, $\mathrm{H}(\bar{q}) - \mathrm{H}(q') \simeq 0.28$.

(b) Case B. $\bar{q} = (0.35, 0.45, 0.2)$, $\mathrm{H}(\bar{q}) \simeq 1.51$, $\mathrm{H}(\bar{q}) - \mathrm{H}(q) \simeq 0.22$, $\mathrm{H}(\bar{q}) - \mathrm{H}(q') \simeq 0.59$.

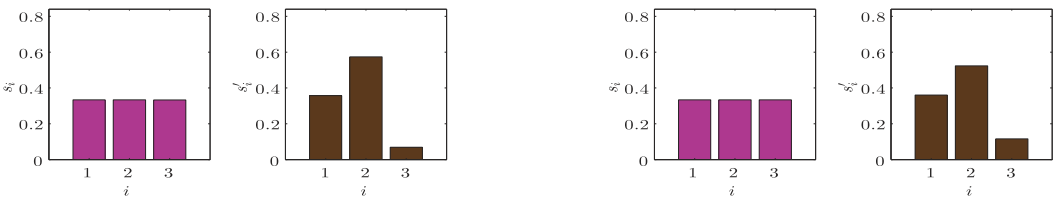
Fig. 6. Apparent profiles $s$, $s'$ of the maximin strategy, without traffic constraints. $s = s' = \bar{q}, \max\{\mathrm{H}(s), \mathrm{H}(s')\} = \mathrm{H}(\bar{q})$.



(a) Symmetrical exchange at half rate, $r = \frac{1}{2}q$, $r' = \frac{1}{2}q'$.

(b) Equivalent exchange policy with same net effect $d$.

Fig. 7. Two pairs of exchange policies $r$, $r'$, corresponding to a single bidirectional exchange $d = \frac{1}{2}(-q + q')$, optimal for the maximin criteria, in case A.



(a) Case A. $\mathrm{H}(s) = \mathrm{H}(u) \simeq 1.59$, $\mathrm{H}(s') \simeq 1.26 < \mathrm{H}(\bar{q}) \simeq 1.44$, $\mathrm{H}(s) - \mathrm{H}(q) \simeq 0.099 \simeq \mathrm{H}(s') - \mathrm{H}(q')$.

(b) Case B. $\mathrm{H}(s) = \mathrm{H}(u) \simeq 1.59$, $\mathrm{H}(s') \simeq 1.38 < \mathrm{H}(\bar{q}) \simeq 1.44$, $\mathrm{H}(s) - \mathrm{H}(q) \simeq 0.29$, $\mathrm{H}(s') - \mathrm{H}(q') \simeq 0.46$.

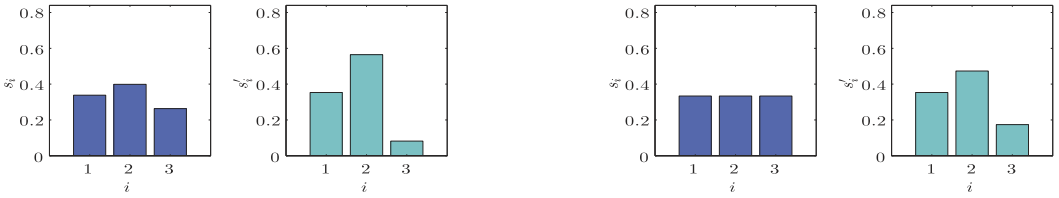
Fig. 8. Apparent profiles $s$, $s'$ of the maximin-gain strategy, without traffic constraints.

has been superimposed. The exterior of the region $\mathcal{C}$, defined by (3), is grayed out. It is interesting to note that the 0.2 traffic restriction in the maximin strategy for case A does not seem to affect its performance with respect to the unconstrained case, and for this reason, it is hard to distinguish a third point in the plot. The resulting profiles for the points without traffic constraints are shown in Fig. 6 for the maximin strategy, Fig. 8 for the maximin-gain strategy, and Fig. 9 for the on-the-fly strategy.

Under the maximin strategy, with apparent profiles shown in Fig. 6, and for case A, the traffic required to merge the apparent profiles into the group profile $\bar{q}$ corresponds to the exchange policy given by

$$d = -r + r' = \frac{1}{2}(-q + q') = (-0.15, 0.2, -0.05).$$

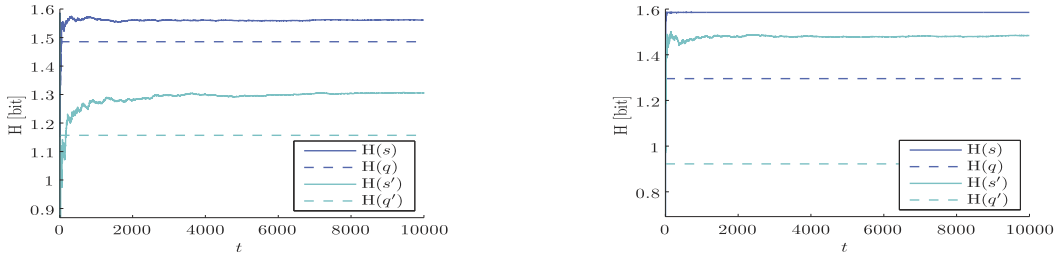
In the untrusted case, the optimal apparent profiles $(\bar{q}, \bar{q})$ may be achieved with a symmetrical exchange with $r = \frac{1}{2}q$ and $r' = \frac{1}{2}q'$, that is, with the exchange of half of the queries regardless of category. If users trust each other or a third party with the knowledge of $q$ and $q'$, they can compute unidirectional exchanges $r$ and $r'$ less demanding in terms of traffic, keeping the same exact bidirectional exchange $d$, namely $r = (0.15, 0, 0.05)$ and $r' = (0, 0.2, 0)$, which tightly meet the 0.2 traffic constraint. The two pairs of exchange policies mentioned are plotted in Fig. 7. Lastly, as expected,




(a) Case A. $\mathrm{H}(s) \simeq 1.56$, $\mathrm{H}(s') \simeq 1.31$, $\mathrm{H}(s) - \mathrm{H}(q) \simeq 0.076$, $\mathrm{H}(s') - \mathrm{H}(q') \simeq 0.15$.

(b) Case B. $\mathrm{H}(s) \simeq 1.59$, $\mathrm{H}(s') \simeq 1.48$, $\mathrm{H}(s) - \mathrm{H}(q) \simeq 0.29$, $\mathrm{H}(s') - \mathrm{H}(q') \simeq 0.56$.

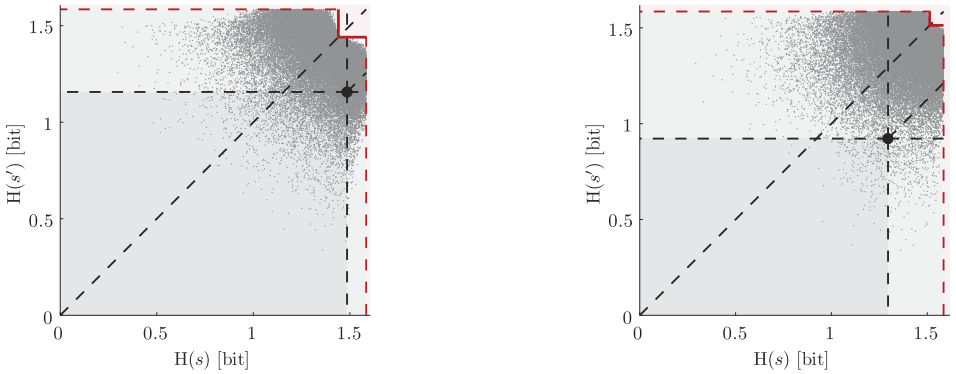
Fig. 9. Apparent profiles $s$, $s'$ of the on-the-fly strategy, without traffic constraints.




(a) Case A.

(b) Case B.

Fig. 10. Privacy versus time for the on-the-fly strategy, without traffic constraints.




(a) Case A.

(b) Case B.

Fig. 11. Random strategies.

case B yields a higher minimum privacy. In case A, one of the gains is slightly negative.

Regarding the maximin-gain strategy in Fig. 8, unsurprisingly, the gain of case B is higher than that of A. The profiles in Fig. 9 confirm the intuition that the on-the-fly strategy should produce results similar to the maximin-gain strategy. One of the three traces of the simulation of the on-the-fly strategy is plotted in Fig. 10. According to the plots, the apparent entropy of the users converges toward a fairly stable value after roughly 1,000 time units.

Finally, recall that our theoretical characterization of the privacy region is partial, albeit sufficient for our purposes, namely the analysis of the maximin and maximin-gain strategies. In order to help us visualize the actual privacy region, and to assess the risk of choosing exchange policies at random, we proceed to draw $5 \cdot 10^4$ random exchange policies $r$ and $r'$, with components $r_i$ and $r'_i$ uniformly distributed in the intervals $(0, q_i)$ and $(0, q'_i)$, respectively. The resulting entropy pairs give the points in the privacy plane depicted in Fig. 11. Although there is a high density of solutions near the optimal maximin and maximin-gain policies, we observe a clear risk of falling far out of the region of mutual privacy gain.

## 7 Further Security Considerations

Because the scope of our contribution must be necessarily limited, so must be the range of applicability of the model assumed in Section 3.1. Having said that, it remains appropriate to discuss the feasibility of this model and its integration with additional privacy and security measures, aside from pseudonymization.

With regard to the first assumption of Section 3.1, on the formation of groups of users agreeing to follow a certain cooperative protocol, two users exchanging queries in our case, the literature abounds with mechanisms, especially for decentralized P2P architectures, which become particularly challenging in mobile ad hoc networks [7], [13]. A particularly interesting example is the collaborative structure, also for privacy purposes of [45], which devotes an entire section to the creation and maintenance of such structure.

The second assumption in Section 3.1 referred to DoS attacks and peers curious about each other's profile. The former issue acknowledges the fact that users may refrain from carrying out parts of a particular protocol in order to save traffic, or simply act maliciously. The latter issue is relevant because some of the strategies of Section 5 assume

that users are not curious and restrict the privacy risk to the IP or an observer external to the group. For that reason, in the group formation protocols aforementioned, users may wish to choose to cooperate with trustworthy peers rather than randomly assigned users. When simply choosing a known friend in real life is not an option, an online form of friendship can be established for our purposes by means of reputation systems, which can at the same time deal with the issue of DoS attacks. More precisely, reputation systems can frustrate the intentions of selfish users, acting against observable misbehavior to enforce cooperation. In this way, if a node does not behave cooperatively, the affected nodes may decide to deny reciprocal cooperation, through the implementation of measures to detect and expel them from the collaborative structure [34], [51]. In our context, detection might entail users occasionally checking the veracity of replies to exchanged queries, by occasionally repeating their submission directly to the IP. Open questions are the detection/quantification of the information leaked by/to untrusted peers.

Even if choosing a known friend were a possibility, Sybil attacks remain a potential risk. Sybil attacks [38] are those in which an attacker forges an identity. In our setting, the privacy of honest users could be in jeopardy since a single user can collect queries intended for a trusted peer and the IP. Solutions to this issue include mutual authentication between parties by means of secure communication protocols, which would also permit hiding the content of the queries exchanged from unintended observers, and even from the peer involved in the exchange, as some of the strategies in Section 5 require.

Last but not least, extending query exchange beyond two users may be an effective means to hinder both peers and external observers in their efforts to profile a particular user, even if they collude toward that purpose, simply because of the practical difficulty of collecting queries submitted to a large population. The idea of privacy through multiple user collaboration has been explored in the literature, in the form of a number of protocols, such as [45], [47]. The privacy gained by enlarging the number of cooperating users may come at the cost of a more complex group formation, higher traffic overhead, and more intricate trust requirements with regards to privacy or DoS.

As for our third assumption in Section 3.1, on the inability of the IP to discern the query exchange policy to better infer the actual profiles from the apparent ones, cannot be fully guaranteed without traffic analysis countermeasures. Indeed, even if queries exchanged are encrypted and their content made confidential, size, timing, and bitwise packet comparisons may unveil which queries are being exchanged between whom. For example, in order to strengthen the simplified cryptographic protocol of Section 5.1, both channels between one and two and between two and the IP could be encrypted with session keys, but packets should still be padded and delayed to prevent size and timing analysis. The issue of anonymous communication has been extensive studied, in many cases building upon the principles underlying Chaum's mixes [8], [14], [29], [35]. Finally, an attacker could have access to certain observations, such as the transient regime of the on-the-fly strategy, that would help refine the estimates of the actual profiles beyond what is merely contained in the apparent profiles.

## 8 Conclusion

This work tackles the problem of query profile obfuscation by means of partial query exchanges between two users, in order for their profiles of interest to appear distorted from an external observer's perspective. Our approach starts with a mathematical formulation, involving the modeling of their apparent user profiles as PMFs over categories of interest, measuring their privacy as the corresponding Shannon entropy. The question of which query categories to exchange translates into optimization variables, for various optimization objectives based on those entropies, possibly under exchange traffic constraints. The formulation is then investigated mainly theoretically, but also numerically for a couple of simple albeit insightful pairs of user profiles. In a way, this is a continuation of our work on optimal query forgery and tag suppression for the Semantic Web [42], [44].

Our main objective is *not* the specific metric of privacy adopted, or the precise mathematical characterizations and optimizations presented, but an illustration of a methodology of systematic, formal analysis of a privacy problem, oriented toward optimized engineering of practical, privacy-protective, information systems.

The fact that our main contribution is intended to be methodological does not prevent us from indulging in a number of results with both theoretical and practical interest. Somewhat surprisingly, we find that symmetrical exchange of half of the queries regardless of their category maximizes the minimum of the entropies. Simply put, it gives the "best worst" privacy. However, the maximum minimum entropy need not lie in the region of mutual privacy gain. In the event that we wish to restrict ourselves to mutual gain, we contemplate, as an alternative optimization criterion, the maximization of the minimum entropy gain. Under this criterion, an optimal solution consists of nearly full exchange of queries from the user with higher initial entropy to the other, so that the first keeps a residual with uniform distribution, and the second appears to behave according to a mixture of the original profiles of interest.

Bearing in mind these theoretical results, we propose a number of query exchange strategies, in terms of the privacy criteria, the level of trust between users, and the presence of constraints in the amount of traffic exchanged. These strategies are then evaluated numerically, by verifying the theoretical results, and assessing the influence of traffic constraints. The examples considered confirm a substantial degree of influence of the original profiles of interest into the privacy benefits obtained, illustrate the applicability of constrained optimization techniques, such as interior-point methods, to the selection of optimal exchange policies, and, last but not least, wrap up our illustration of methodology for formally approaching private information systems.

## Acknowledgments

The authors are grateful to Javier Parra-Arnau and the anonymous reviewers for their helpful comments. This work was supported in part by the Spanish Government through Projects CONSOLIDER INGENIO 2010 CSD2007-00004

"ARES," TEC2010-20572-C02-02 "CONSEQUENCE," TSI2007-65406-C03-01 "E-AEGIS," and TIN2011-27076-C03-01 "CO-PRIVACY," by the Catalan Government under Grants 2009 SGR 1362 and 2009 SGR 1135, and by the European Commission under FP7 project "DwB." The third author is partly supported as an ICREA-Acadèmia researcher by the Government of Catalonia. Also, he holds the UNESCO Chair in Data Privacy, but this paper does not necessarily reflect the position of UNESCO nor commits that organization.

## REFERENCES

[1] C. Aguilar-Melchor and Y. Deswarte, "Trustable Relays for Anonymous Communication," *Trans. Data Privacy,* vol. 2, no. 2, pp. 101-130, 2009.

[2] "AOL Search Data Scandal," http://en.wikipedia.org/wiki/AOL_search_data_scandal, Aug. 2006.

[3] V. Benjumea, J. López, and J.M.T. Linero, "Specification of a Framework for the Anonymous Use of Privileges," *Telematics and Informatics,* vol. 23, no. 3, pp. 179-195, Aug. 2006.

[4] G. Bianchi, M. Bonola, V. Falletta, F.S. Proto, and S. Teofili, "The SPARTA Pseudonym and Authorization System," *Science of Computer Programming,* vol. 74, nos. 1/2, pp. 23-33, 2008.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge Univ. Press, 2004.

[6] J. Castellà-Roca, A. Viejo, and J. Herrera-Joancomartí, "Preserving User's Privacy in Web Search Engines," *Computer Comm.,* vol. 32, nos. 13/14, pp. 1541-1551, 2009.

[7] N. Chatterjee, A. Potluri, and A. Negi, "A Scalable and Adaptive Clustering Scheme for MANETs," *Proc. Fourth Int'l Conf. Distributed Computing Internet Technology (ICDCIT '07),* pp. 73-78, Dec. 2007.

[8] D. Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms," *Comm. ACM,* vol. 24, no. 2, pp. 84-88, 1981.

[9] D. Chaum, "Security without Identification: Transaction Systems to Make Big Brother Obsolete," *Comm. ACM,* vol. 28, no. 10, pp. 1030-1044, Oct. 1985.

[10] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private Information Retrieval," *Proc. 36th Ann. Symp. Foundations of Computer Science (FOCS '95),* pp. 41-50, 1995.

[11] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private Information Retrieval," *J. ACM,* vol. 45, no. 6, pp. 965-981, 1998.

[12] T.M. Cover and J.A. Thomas, *Elements of Information Theory,* second ed. Wiley, 2006.

[13] C. Cramer, O. Stanze, K. Weniger, and M. Zitterbart, "Demand-Driven Clustering in MANETs," *Proc. Int'l Conf. Wireless Networking (ICWN),* vol. 1, pp. 81-87, June 2004.

[14] G. Danezis, R. Dingledine, and N. Mathewson, "Mixminion: Design of a Type III Anonymous Remailer Protocol," *Proc. Symp. Security Privacy (SP),* pp. 2-15, May 2003.

[15] C. Díaz, "Anonymity and Privacy in Electronic Services," PhD dissertation, Katholieke Univ. Leuven, Dec. 2005.

[16] R. Dingledine, "Free Haven's Anonymity Bibliography," www.freehaven.net/anonbib/, 2009.

[17] J. Domingo-Ferrer, "Coprivacy: Towards a Theory of Sustainable Privacy," *Proc. Int'l Conf. Privacy in Statistical Databases (PSD '10),* pp. 258-268, Sept. 2010.

[18] J. Domingo-Ferrer, "Coprivacy: An Introduction to the Theory and Applications of Co-Operative Privacy," *Special Issue: Privacy in Statistical Databases,* vol. 35, pp. 25-40, 2011.

[19] J. Domingo-Ferrer, M. Bras-Amorós, Q. Wu, and J. Manjón, "User-Private Information Retrieval Based on a Peer-to-Peer Community," *Data Knowledge Eng.,* vol. 68, no. 11, pp. 1237-1252, 2009.

[20] J. Domingo-Ferrer and Ú. González-Nicolás, "Rational Behavior in Peer-to-Peer Profile Obfuscation for Anonymous Keyword Search," *Information Science: An Int'l J.,* vol. 185, no. 1, pp. 191-204, 2012.

[21] J. Domingo-Ferrer, A. Solanas, and J. Castellà-Roca, "$h(k)$-Private Information Retrieval from Privacy-uncooperative Queryable Databases," *Online Information Rev.,* vol. 33, no. 4, pp. 720-744, 2009.

[22] M. Duckham, K. Mason, J. Stell, and M. Worboys, "A Formal Approach to Imperfection in Geographic Information," *Computers Environment and Urban Systems,* vol. 25, no. 1, pp. 89-103, 2001.

[23] Y. Elovici, C. Glezer, and B. Shapira, "Enhancing Customer Privacy while Searching for Products and Services on the World Wide Web," *Internet Research: Electronic Networking Applications and Policy,* vol. 15, no. 4, pp. 378-399, 2005.

[24] Y. Elovici, B. Shapira, and A. Maschiach, "A New Privacy Model for Hiding Group Interests while Accessing the Web," *Proc. Workshop Privacy in the Electronic Soc. (WPES '02),* pp. 63-70, 2002.

[25] J. Freudiger, M.H. Manshaei, J.-P. Hubaux, and D.C. Parkes, "On Non-Cooperative Location Privacy: A Game-theoretic Analysis," *Proc. 16th ACM Conf. Computer and Comm. Security (CCS '09),* Nov. 2009.

[26] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private Queries in Location Based Services: Anonymizers Are Not Necessary," *Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '08),* pp. 121-132, June 2008.

[27] "GoogleSharing," www.googlesharing.net, 2012.

[28] D.C. Howe and H. Nissenbaum, "TrackMeNot: Resisting Surveillance in Web Search," *Lessons from the Identity Trail: Privacy, Anonymity and Identity in a Networked Soc.,* Oxford Univ. Press, http://mrl.nyu.edu/~dhowe/trackmenot, 2006.

[29] "I2P Anonymous Network," www.i2p2.de, 2012.

[30] E.T. Jaynes, "Information Theory and Statistical Mechanics II," *Physical Rev.,* vol. 108, no. 2, pp. 171-190, 1957.

[31] E.T. Jaynes, "On the Rationale of Maximum-Entropy Methods," *Proc. IEEE,* vol. 70, no. 9, pp. 939-952, Sept. 1982.

[32] H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of Location Privacy Using Dummies for Location-Based Services," *Proc. IEEE Int'l Conf. Data Eng. (ICDE),* p. 1248, Oct. 2005.

[33] T. Kuflik, B. Shapira, Y. Elovici, and A. Maschiach, "Privacy Preservation Improvement by Learning Optimal Profile Generation Rate," *Proc. Ninth Int'l Conf. User Modeling (UM '03),* pp. 168-177, 2003.

[34] M. Mejía, N. Pena, J.L. Munoz, and O. Esparza, "A Review for Trust Modelling in Ad Hoc Networks," *Internet Research,* vol. 19, no. 1, pp. 88-104, 2009.

[35] U. Möller, L. Cottrell, P. Palfrader, and L. Sassaman, "Mixmaster Protocol—Version 2," Internet Eng. Task Force, Internet Draft, http://www.freehaven.net/anonbib/cache/mixmaster-spec.txt, July 2003.

[36] M. Murugesan and C. Clifton, "Providing Privacy through Plausibly Deniable Search," *Proc. SIAM Int'l Conf. Data Mining (SDM),* 2009.

[37] J. Nash, "Non-Cooperative Games," *Annals Math.,* vol. 54, pp. 289-295, 1951.

[38] J. Newsome, E. Shi, D. Song, and A. Perrig, "The Sybil Attack in Sensor Networks: Analysis and Defenses," *Proc. Third Int'l Symp. Information Processing in Sensor Networks (IPSN '04),* pp. 259-268, Apr. 2004.

[39] N. Nisan, T. Roughgarden, É. Tardos, and V.V. Vazirani, *Algorithmic Game Theory.* Cambridge Univ. Press, 2007.

[40] F. Olumofin and I. Goldberg, "Revisiting the Computational Practicality of Private Information Retrieval," *Proc. Financial Cryptography Data Security (FI),* Feb. 2011.

[41] R. Ostrovsky and W.E. Skeith III, "A Survey of Single-database PIR: Techniques and Applications," *Proc. Int'l Conf. Practice, Theory Public-Key Cryptography (PKC),* pp. 393-411, Sept. 2007.

[42] J. Parra-Arnau, D. Rebollo-Monedero, and J. Forné, "A Privacy-Preserving Architecture for the Semantic Web Based on Tag Suppression," *Proc. Seventh Int'l Conf. Trust, Privacy and Security in Digital Business (TrustBus '10),* Aug. 2010.

[43] R. Puzis, D. Yagil, Y. Elovici, and D. Braha, "Collaborative Attack on Internet Users Anonymity," *Internet Research,* vol. 19, no. 1, pp. 60-77, 2009.

[44] D. Rebollo-Monedero and J. Forné, "Optimal Query Forgery for Private Information Retrieval," *IEEE Trans. Information Theory,* vol. 56, no. 9, pp. 4631-4642, Sept. 2010.

[45] D. Rebollo-Monedero, J. Forné, A. Solanas, and T. Martínez-Ballesté, "Private Location-Based Information Retrieval through User Collaboration," *Computer Comm.,* vol. 33, no. 6, pp. 762-774, http://dx.doi.org/10.1016/j.comcom.2009.11.024, 2010.

[46] D. Rebollo-Monedero, J. Forné, L. Subirats, A. Solanas, and A. Martínez-Ballesté, "A Collaborative Protocol for Private Retrieval of Location-Based Information," *Proc. IADIS Int'l Conf. e-Soc.,* Feb. 2009.

[47] M.K. Reiter and A.D. Rubin, "Crowds: Anonymity for Web Transactions," *ACM Trans. Information System Security,* vol. 1, no. 1, pp. 66-92, 1998.

[48] C.E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical J.,* vol. 28, pp. 656-715, 1949.
[49] B. Shapira, Y. Elovici, A. Meshiach, and T. Kuflik, "PRAW—The Model for PRivAte Web," *J. Am. Assoc. Information Soc. Information Science and Technology,* vol. 56, no. 2, pp. 159-172, 2005.
[50] C. Soghoian, "The Problem of Anonymous Vanity Searches," *I/S: A J. Law and Policy for the Information Soc. (ISJLP),* vol. 3, Jan. 2007.
[51] A. Srinivasan, J. Teitelbaumy, H. Liangz, J. Wuyand, and M. Cardei, "Reputation and Trust-Based Systems for Ad Hoc and Sensor Networks," *Algorithms and Protocols for Wireless Ad Hoc and Sensor Networks,* John Wiley & Sons, 2007.
[52] "The Tor Project, Tor: Overview," http://torproject.org/overview.html.en, 2011.
[53] "Torbutton 1.2.5," https://addons.mozilla.org/ca/firefox/addon/2275, 2010.
[54] V. Toubiana, "SquiggleSR," www.squigglesr.com, 2007.
[55] A. Viejo and J. Castellà-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines," *Computer Networks,* vol. 54, no. 9, pp. 1343-1357, 2010.

**David Rebollo-Monedero** received the MS and PhD degrees in electrical engineering from Stanford University, in California, 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. He is currently working toward the PhD degree at Stanford, during the summer of 2003 he worked for Apple Computer with the QuickTime video codec team in California. He is currently a postdoctoral researcher with the Information Security Group of the Department of Telematics Engineering at 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 MS degree in telecommunications engineering from the Universitat Politècnica de Catalunya (UPC) in 1992, and the PhD degree in 1997. 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 the coordinator of the PhD program in Telematics Engineering and the director of the research MS program in telematics engineering. His research interests include 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 several international journals.

**Josep Domingo-Ferrer** (M'1988, SM'2002, F'2011) received the MS and PhD degrees in computer science from the Universitat Autònoma de Barcelona in 1988 and 1991 (Outstanding Graduation Award). He received the MS degree in mathematics. He is a full professor of computer science and an ICREA-Acadèmia Researcher at Universitat Rovira i Virgili, Tarragona, Catalonia, where he holds the UNESCO Chair in Data Privacy. His research interests include data privacy, data security, statistical disclosure control, and cryptographic protocols, with a focus on the conciliation of privacy, security, and functionality. He won the first edition of the ICREA Acadèmia Prize 2008 awarded by the Government of Catalonia, which distinguished him as one of the strongest 40 faculty members in public Catalan universities as far as research is concerned. Between 2007 and 2008, he was a corecipient of four entrepreneurship prizes. In 2004 and 2003, he received two research prizes. He has authored three patents and more than 270 publications. He has been the coordinator of EU FP5 project CO-ORTHOGONAL and of several Spanish funded and US funded research projects. He currently coordinates the CONSOLIDER "ARES" team on security and privacy, one of Spain's 34 strongest research teams. He has chaired 11 international conferences and has served in the program committee of more than 125 conferences on privacy and security. He is a coeditor-in-chief of "*Transactions on Data Privacy*," an area editor of "*Computer Communications*," and an associate editor of the "*Journal of Official Statistics*." He has held visiting appointments in Rome, Leuven, Princeton, Munich, and Milwaukee. He is a fellow of the IEEE.