Optimized Query Forgery for Private Information Retrieval

David Rebollo-Monedero and Jordi Forné

***Abstract*—We present a mathematical formulation for the optimization of query forgery for private information retrieval, in the sense that the privacy risk is minimized for a given traffic and processing overhead. The privacy risk is measured as an information-theoretic divergence between the user's query distribution and the population's, which includes the entropy of the user's distribution as a special case. We carefully justify and interpret our privacy criterion from diverse perspectives. Our formulation poses a mathematically tractable problem that bears substantial resemblance with rate-distortion theory.**

***Index Terms*—Entropy, Kullback–Leibler divergence, privacy risk, private information retrieval, query forgery.**

I. Introduction

In August of 2006, AOL Research released a text file intended for research purposes containing twenty million search keywords from more than 650 000 users over a three-month period. Occasionally, the queries submitted by those users contained personally identifiable information written by themselves such as name, address or social security number. In addition, records corresponding to a common user were linked to a unique sequential key in the released file, which made the risk of cross referencing even higher, thereby seriously compromising the privacy of those users. In September of the same year, the scandal lead to a class action lawsuit filed against AOL in the U.S. District Court for the Northern District of California.

The relevance of user privacy is stressed in numerous examples in the literature of information retrieval. These examples include not only the risk of user profiling by an Internet search engine, but also by location-based service (LBS) providers, or even corporate profiling by patent and stock market database providers.

A. State of the Art in Private Information Retrieval

The literature on information retrieval also provides numerous of solutions to user privacy [1]. We would like to touch upon some of these solutions, often extensible to scenarios other than the ones they were intended for. Please bear in mind that, throughout the paper, we shall use the term private information retrieval in its widest sense, *not* only to refer to the particular class of cryptographically-based methods usually connected with the acronym PIR. In any case, such particular class of methods will be briefly discussed later in this section. While keeping a general perspective on the main large classes of existing solutions for private information retrieval, in order to make our exposition more concrete, we occasionally relate these solutions to the specific application scenario of LBSs, even though most ideas are immediately extensible to Internet search. Recent surveys with a greater focus on anonymous Internet search include [2], [3].

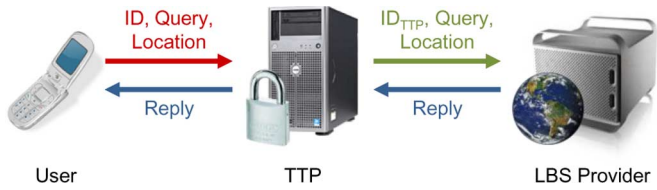


Fig. 1. Anonymous access to an LBS provider through a TTP.

One of the conceptually simplest approaches to anonymous information retrieval consists in including a trusted third party (TTP) acting as an intermediary between the user and the information service provider, which effectively hides the identity of the user. In the particularly rich, important example of LBSs, the simplest form of interaction between a user and an information provider involves a direct message from the former to the latter including a query and the location to which the query refers. An example would be the query "Where is the nearest bank?", accompanied by the geographic coordinates of the user's current location. In this case, the TTP-based solution, depicted in Fig. 1, would preserve user privacy in terms of both queries and locations. An appealing twist that does not require that the TTP be online is that of pseudonymizing digital credentials [4]–[6].

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 [7], which, naturally, may lead to an inaccurate answer. The principle behind TTP-free perturbative methods for privacy in LBSs is represented in Fig. 2. 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 trade-off between data

Manuscript received December 21, 2009; revised December 21, 2009. Date of current version August 18, 2010. This work was supported in part by the Spanish Government under Projects CONSOLIDER INGENIO 2010 CSD2007-00004 "ARES" and TSI2007-65393-C02-02 "ITACA", and in part by the Catalan Government under Grant 2009 SGR 1362.

The authors are with the Department of Telematics Engineering, Technical University of Catalonia (UPC), E-08034 Barcelona, Spain (e-mail: david.rebollo@entel.upc.edu; jforne@entel.upc.edu).

Communicated by K. M. Martin, Associate Editor for Complexity and Cryptography.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2054471

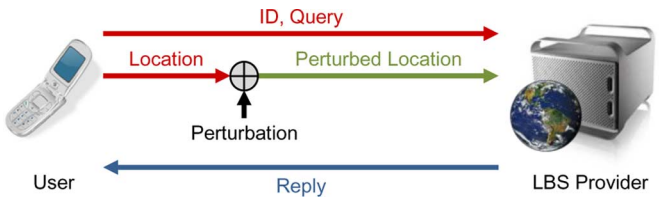


Fig. 2. Users may contact an untrusted LBS provider directly, perturbing their location information to help protect their privacy.

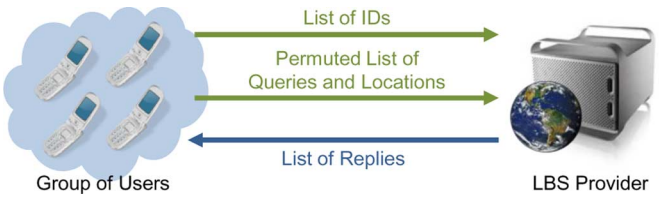


Fig. 3. Communication between a set of users and an untrusted provider without using a TTP.

utility and privacy common to any perturbative privacy method.

Fig. 3 is a conceptual depiction of TTP-free methods relying on the collaboration between multiple users, in the special case of LBSs. A proposal based on this collaborative principle considers groups of users that know each other's locations but trust each other, who essentially achieve anonymity by sending to the LBS provider a spatial cloaking region covering the entire group [8]. An effort towards k-anonymous [9], [10] LBSs, this time not assuming that collaborating users necessarily trust each other, is [11], [12]. Fundamentally, k users add zero-mean random noise to their locations and share the result to compute the average, which constitutes a shared perturbed location sent to the LBS provider.

Alternatively, 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 [13], [14]. Unfortunately, these methods require the provider's cooperation in the privacy protocol, are limited to a certain extent to query-response functions in the form of a finite lookup table of precomputed answers, and are burdened with a significant computational overhead.

An approach to preserve user privacy to a certain extent, at the cost of traffic and processing overhead, which does not require that the user trust the service provider nor the network, consists in accompanying original queries with bogus queries. Building on this simple principle, several PIR protocols, mainly heuristic, have been proposed and implemented, with various degrees of sophistication [2], [15], [16]. An illustrative example for LBSs is [17]. Query forgery appears also as a component of other privacy protocols, such as the private location-based information retrieval protocol via user collaboration in [18], [19]. Simple, heuristic implementations in the form of add-ons for popular browsers have begun to appear recently [20], [21]. In addition to legal implications, there are a number of technical considerations regarding bogus traffic generation for privacy [22], 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, 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.

B. Contribution and Organization of the Paper

A patent issue regarding query forgery is the trade-off between user privacy on the one hand, and cost in terms of traffic and processing overhead on the other. The object of this paper is to investigate this trade-off in a mathematically systematic fashion. More specifically, we present a mathematical formulation of optimal query forgery for private information retrieval. We propose an information-theoretic criterion to measure the privacy risk, namely a divergence between the user's and the population's query distributions, which includes the entropy of the user's distribution as a special case, and which we carefully interpret and justify from diverse perspectives. Our formulation poses a mathematically tractable problem that bears substantial resemblance with rate-distortion theory.

Section II presents an information-theoretic formulation of the compromise between privacy and redundancy in query forgery for private information retrieval. The privacy criterion proposed is justified and interpreted essentially in Section III. Section IV contains a detailed theoretical analysis of the optimization problem characterizing the privacy-redundancy trade-off, illustrated by means of simple, conceptual examples in Section V. Conclusions are drawn in Section VI.

II. Formal Problem Statement

We model user *queries* as random variables on a common measurable space. In practice, rather than specific, complete queries, these random variables may actually represent query categories or topics, individual keywords in a small indexable set, or parts of queries such as coordinates sent to an LBS provider. A sequence of related queries may be modeled as a single multivariate random variable, in order to capture any existing statistical dependence and, in this way, hinder privacy attackers in their efforts to exploit spacial and temporal correlations. Alternatively, conditional probability distributions given previous values of statistically dependent queries may be contemplated.

To avoid certain mathematical technicalities, we shall assume that the query *alphabet* is finite, for example given by finite discretizations of continuous probability distributions, suitable for numerical computation. Having assumed that the alphabet is finite, we may suppose further, although this time without loss of generality, that queries take on values in the set $\{1,\ldots,n\}$ for some $n \in \mathbb{Z}^+$, even in the case of multivariate queries.

Accordingly, define p as the probability distribution of the *population*'s queries, q as the distribution of the authentic queries of a particular *user*, and r as the distribution of the user's *forged* queries, all on the same query alphabet. Whenever the user's distribution differs from the population's, a privacy attacker will have actually gained some information about the user, in contrast to the statistics of the general population. Inspired by the measures of privacy proposed in [23]–[26], we define the initial privacy risk $\mathcal{R}_0$ as the Kullback–Leibler (KL)

divergence [27] D between the user's and the population's distributions, that is

$$\mathcal{R}_0 = \mathrm{D}(q||p).$$

Section III below is devoted to the interpretation and justification of this privacy criterion.

Define $0 \leqslant \rho \leqslant 1$ as the query *redundancy*, measured as the ratio of forged queries to total queries. The user's *apparent* query distribution is the convex combination $(1-\rho)q + \rho r$, which, for brevity, we shall occasionally denote by s. Accordingly, we define the (final) privacy risk $\mathcal{R}$ as the divergence between the apparent distribution and the population's

$$\mathcal{R} = \mathrm{D}(s||p) = \mathrm{D}((1-\rho)q + \rho r||p).$$

Suppose that the population is large enough to neglect the impact of the choice of r on p. Consistently, we define the *privacy-redundancy* function

$$\mathcal{R}(\rho) = \min_r \mathrm{D}((1-\rho)q + \rho r||p) \tag{1}$$

which characterizes the optimal trade-off between query privacy (risk) and redundancy. We would like to remark that the optimization problem inherent in this definition involves a lower bounded, lower semicontinuous function over a compact set, namely the probability simplex to which r belongs. Hence, we are justified in using the term minimum rather than infimum. On a side note, analogous theoretical results can be developed for an alternative definition of privacy risk, given by the inversion of the two arguments of the KL divergence.

From a practical perspective, while a user wishing to solve the optimization problem formulated may be able to easily keep track of its own query distribution q, estimating the population's query distribution p may be trickier, unless the information provider is willing to collect and share reliable aggregated data, as Google Insights [28], for instance, intends. Section III-A will elaborate on the alternative of measuring privacy risk as an entropy, for which only the knowledge of q is required, and will argue that this is formally a special case of our general divergence measure; precisely, when p is (assumed to be) the uniform distribution.

For simplicity, we use natural logarithms throughout the paper, particularly because all bases produce equivalent optimization objectives.

III. KL Divergence as a Measure of Privacy Risk

Before analyzing the theoretical properties of the privacy-redundancy function (1), defined as an optimization problem in the previous section, we would like to interpret and justify our choice of KL divergence as a privacy criterion, mainly inspired by [23]. In this section, we shall emphasize what we find to be the essential interpretations of our privacy criterion.

In spite of its information-theoretic appeal and mathematical tractability, we must acknowledge that the adequacy of our formulation relies on the appropriateness of the criteria optimized, which in turn depends on the specific application, on the query statistics of the users, on the actual network and processing overhead incurred by introducing forged queries, and last but not least, on the adversarial model and the mechanisms against privacy contemplated. The interpretations and justifications that follow are merely intended to aid users and system designers in their assessment of the suitability of our proposal to a specific information-retrieval application.

We would like to stress as well that the use of an information-theoretic quantity for privacy assessment is by no means new, as the work by Shannon in 1949 [29] already introduced the concept of *equivocation* as the conditional entropy of a private message given an observed cryptogram, later used in the formulation of the problem of the wiretap channel [30], [31] as a measure of confidentiality. We can also trace back to the fifties the information-theoretic interpretation of the divergence between a prior and a posterior distribution, named (average) information gain in some statistical fields [32], [33]. More recent work reaffirms the applicability of the concept of entropy as a measure of privacy. For example, [26] (see also [34]) is one of the earliest proposals for measuring the degree of anonymity observable by an attacker as the entropy of the probability distribution of possible senders of a given message.

A. Entropy Maximization

Our first interpretation arises from the fact that Shannon's entropy may be regarded as a special case of KL divergence. Precisely, let u denote the uniform distribution on $\{1, \ldots, n\}$, that is, $u_i = 1/n$. In the special case when $p = u$, the privacy risk becomes

$$\mathrm{D}((1-\rho)q + \rho r||u) = \ln n - \mathrm{H}((1-\rho)q + \rho r).$$

In other words, minimizing the KL divergence is equivalent to maximizing the entropy of the user's apparent query distribution

$$\mathcal{R}(\rho) = \ln n - \max_r \mathrm{H}((1-\rho)q + \rho r).$$

Accordingly, rather than using the measure of privacy risk represented by the KL divergence, relative to the population's distribution, we would use the entropy $\mathrm{H}((1-\rho)q + \rho r)$ as an absolute measure of privacy gain.

This observation enables us to connect, at least partly, our privacy criterion with the rationale behind *maximum-entropy* methods, an involved topic not without controversy, which arose in statistical mechanics [35], [36], and has been extensively addressed by abundant literature [37] over the past half century. Some of the arguments advocating maximum-entropy methods deal with the highest number of permutations with repeated elements associated with an empirical distribution [38], or more generally, the method of types and large deviation theory [27, § 11]. Some others deal with a consistent axiomatization of entropy [37], [39]–[41], further to the original one already established by Shannon [42], slightly reformulated in [43], [44], and generalized by Rényi in [45], and some relate Bayesian inference to divergence minimization [46].

B. Hypothesis Testing

We turn back to the more general case of KL divergence as a measure of privacy, that is, when the reference distribution p is not necessarily uniform. The above-mentioned arguments

concerning a consistent axiomatization of the Shannon entropy have been extended to the KL divergence [37], [39]–[41], which may in fact be regarded as an entropy relative to a reference probability measure.

We believe, however, that one of the most interesting justifications for measuring privacy risk as a KL divergence stem from the arguments based on the method of types and large deviation theory. More precisely, through *Stein's lemma* [27, § 11] we shall interpret KL divergences as false positives and negatives when an attacker applies *hypothesis testing* to ascertain whether a sequence of observed queries belongs to a predetermined user or not. We explain this justification here.

Our interpretation contemplates the scenario where an attacker knows, or is able to estimate, the apparent query distribution s of a given user. Further, the attacker observes a sequence of k i.i.d. queries, and attempts to decide whether they belong to that particular user or not. More precisely, the attacker considers the hypothesis testing between two alternatives, namely whether the queries have been drawn according to s, the user's apparent distribution (first hypothesis), or p, the general population's distribution (second hypothesis). Define the *acceptance region* $\mathcal{A}_k$ as the set of sequences of observed queries over which the attacker decides to accept the first hypothesis. Accordingly, define two probabilities of decision error:

(a) The error of the first kind $\alpha_k = s(\bar{\mathcal{A}}_k)$, which is the probability of a false negative.
(b) The error of the second kind $\beta_k = p(\mathcal{A}_k)$, which is the probability of a false positive.

Above, $\bar{\mathcal{A}}_k$ denotes the complement of $\mathcal{A}_k$. $p(\mathcal{A}_k)$, for example, represents the probability of all query sequences in $\mathcal{A}_k$, i.i.d. according to p, and similarly for $s(\bar{\mathcal{A}}_k)$.

For any $0 < \epsilon < 1/2$, define

$$\beta_k^\epsilon = \min_{\substack{\mathcal{A}_k \\ \alpha_k \leqslant \epsilon}} \beta_k$$

in other words, we choose the acceptance region with least false positive rate among those with a bounded false negative rate. Stein's lemma asserts that

$$\lim_{\epsilon \to 0} \lim_{k \to \infty} \frac{1}{k} \ln \beta_k^\epsilon = -\mathrm{D}(s||p).$$

Less formally, $\beta_k^\epsilon \simeq e^{-k\mathrm{D}(s||p)}$ for large k. The minimization of $\mathrm{D}(s||p)$ in the definition of the privacy-redundancy function (1) maximizes the exponent in the error rate of false positives, for an optimal choice of acceptance region with a false negative rate constraint. Simply put, the optimal forgery strategy r^* makes the attacker's job more difficult.

Clearly, we might very well exchange the roles of s and p in this interpretation, and concordantly define privacy risk as $\mathrm{D}(p||s)$ in lieu of $\mathrm{D}(s||p)$. It turns out that most of the results obtained in this work can be easily adapted to this alternative formulation, but some of the additional interpretations we have presented, and particularly the important entropy case of the previous subsection, would not apply.

IV. Optimal Query Forgery

This section investigates the fundamental properties of the privacy-redundancy function (1) defined in Section II, and presents a closed-form solution to the inherent minimization problem. In the interest of brevity, our theoretical analysis only considers the case when all given probabilities are strictly positive

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

The general case can easily be dealt with, occasionally via continuity arguments. We shall suppose further, without loss of generality, that

$$\frac{q_1}{p_1} \leqslant \cdots \leqslant \frac{q_n}{p_n}. \tag{3}$$

It is immediate from the definition of the privacy-redundancy function that its initial and final values are $\mathcal{R}(0) = \mathrm{D}(q||p)$ and $\mathcal{R}(1) = 0$. The behavior of $\mathcal{R}(\rho)$ at intermediate values of ρ is characterized by the theorems in this section.

A. Monotonicity and Convexity

Theorem 1: The privacy-redundancy function $\mathcal{R}(\rho)$ is nonincreasing and convex.

Proof: First, let $0 \leqslant \rho < \rho' \leqslant 1$. Based on the solution r to the minimization problem corresponding to $\mathcal{R}(\rho)$, construct the distribution $r' = (1 - \rho/\rho')\, q + \rho/\rho' r$, satisfying

$$(1 - \rho')q + \rho' r' = (1 - \rho)q + \rho r.$$

Because r' is not necessarily a minimizer of the problem corresponding to $\mathcal{R}(\rho')$, it follows that $\mathcal{R}(\rho') \leqslant \mathcal{R}(\rho)$, and consequently, that the privacy-redundancy function is nonincreasing.

Secondly, we prove convexity by verifying that

$$(1 - \lambda)\mathcal{R}(\rho) + \lambda \mathcal{R}(\rho') \geqslant \mathcal{R}((1 - \lambda)\rho + \lambda\rho')$$

for all $0 \leqslant \lambda, \rho, \rho' \leqslant 1$. The solutions corresponding to $\mathcal{R}(\rho)$ and $\mathcal{R}(\rho')$ are denoted by r and r', respectively. Define $\rho_\lambda = (1 - \lambda)\rho + \lambda\rho'$ and

$$r_\lambda = \frac{(1 - \lambda)\rho r + \lambda \rho' r'}{(1 - \lambda)\rho + \lambda\rho'}.$$

We have

$$\begin{aligned}
&(1 - \lambda)\mathcal{R}(\rho) + \lambda \mathcal{R}(\rho') \\
&\quad = (1 - \lambda)\mathrm{D}((1 - \rho)q + \rho r||p) \\
&\qquad + \lambda \mathrm{D}((1 - \rho')q + \rho' r'||p) \\
&\quad \overset{(a)}{\geqslant} \mathrm{D}\Big((1 - \lambda)((1 - \rho)q + \rho r) \\
&\qquad\qquad + \lambda((1 - \rho')q + \rho' r')||p \Big) \\
&\quad = \mathrm{D}((1 - \rho_\lambda)q + \rho_\lambda r_\lambda||p) \overset{(b)}{\geqslant} \mathcal{R}(\rho_\lambda)
\end{aligned}$$

where

(a) follows from the fact that the KL divergence is convex in pairs of probability distributions [27, § 2.7];
(b) reflects that r_λ is not necessarily the solution to the minimization problem corresponding to $\mathcal{R}(\rho_\lambda)$.

■

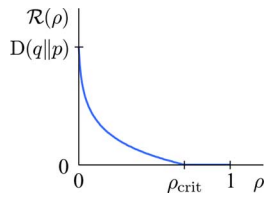


Fig. 4. Conceptual plot of the privacy-redundancy function.

The convexity of the privacy-redundancy 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 $\mathcal{R}(\rho)$ and under the positivity assumption (2), that continuity also holds at the interval endpoints, 0 and 1. Sections IV-D and IV-E analyze in greater detail the behavior of the function at extreme values of ρ.

B. Critical Redundancy

Our second theorem will confirm the intuition that there must exist a redundancy beyond which perfect privacy is attainable, in the sense that $\mathcal{R}(\rho) = 0$. Precisely, this *critical redundancy* is

$$\rho_{\text{crit}} = 1 - \min_i \frac{p_i}{q_i} = 1 - \frac{p_n}{q_n}$$

according to the labeling assumption (3). It is important to realize that $\rho_{\text{crit}} > 0$ unless $p = q$. The key idea is that $\frac{q_n}{p_n} \geqslant 1$, for otherwise $q_i < p_i$ and $1 = \sum q_i < \sum p_i = 1$, a contradiction. On the other hand, the positivity assumption (2) ensures that $\rho_{\text{crit}} < 1$. Unsurprisingly, ρ_{crit} becomes worse (closer to one) with worse (larger) ratio $\max_i \frac{q_i}{p_i} = \frac{q_n}{p_n}$. Fig. 4 is a conceptual depiction of the results stated by Theorems 1 and 2 . The same theorems imply that $\mathcal{R}(\rho) \leqslant (1-\rho/\rho_{\text{crit}})\mathrm{D}(q\|p)$ for $0 \leqslant \rho \leqslant \rho_{\text{crit}}$.

Theorem 2 (Critical Redundancy): Suppose $\rho \geqslant \rho_{\text{crit}}$. Then, $\mathcal{R}(\rho) = 0$. In addition, the optimal forged query distribution is $r^* = \frac{1}{\rho}p + \left(1 - \frac{1}{\rho}\right) q$, for which the user's apparent distribution and the population's match: $p = (1-\rho)\, q + \rho r^*$.

Proof: We consider only the nontrivial case when $p \neq q$, thus $\rho \geqslant \rho_{\text{crit}} > 0$. It is clear from the form of r^* that $\sum_i r_i^* = 1$, thus it suffices to verify that r^* is nonnegative to ascertain whether it is a probability distribution. Observe that requiring that $r_i^* = \frac{1}{\rho}p_i + \left(1 - \frac{1}{\rho}\right) q_i \geqslant 0$ for all i is equivalent to $p_i + (\rho - 1)q_i \geqslant 0$, to $\frac{p_i}{q_i} \geqslant 1 - \rho$, and finally to $\rho \geqslant 1 - \frac{p_i}{q_i}$. But this is also equivalent to requiring that

$$\rho \geqslant \max_i 1 - \frac{p_i}{q_i} = 1 - \min_i \frac{p_i}{q_i}$$

as assumed in the theorem. To complete the proof, it is routine to check that the proposed r^* satisfies $p = (1-\rho)\, q + \rho r^*$, thereby vanishing the privacy risk. ∎

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

$$r_i^* = \frac{p_i q_n - p_n q_i}{q_n - p_n}$$

equal to zero if, and only if, $\frac{q_i}{p_i} = \frac{q_n}{p_n}$. Owing to the fact that we are dealing with relative rather than absolute frequencies, it is not surprising that $r_n^* = 0$ at $\rho = \rho_{\text{crit}}$. More generally, in accordance with the labeling assumption (3), observe that only the last components of r^* may vanish.

C. Closed-Form Solution

Our last theorem, Theorem 4, will provide a closed-form solution to the minimization problem defining the privacy-redundancy function (1). Our solution will be based on a resource allocation lemma, namely Lemma 3, which addresses an extension of the usual water filling problem. Even though Lemma 3 provides a parametric-form solution, fortunately, we will be able to proceed towards an explicit closed-form solution (albeit piecewise), trivially derivable from the implicit form presented in the theorem for elegance.

The lemma in question 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, and to amount to a normalized total of $\sum_i x_i = 1$, thus x is a probability distribution. The well-known water-filling problem [47, § 5.5] may be regarded as the special case when $f_i(x_i) = -\ln(\alpha_i + x_i)$, for $\alpha_i > 0$.

Lemma 3 (Resource Allocation): For all $i = 1, \ldots, n$, let $f_i : \mathbb{R} \to \mathbb{R}$ be twice differentiable on $[0, 1]$, with $f_i'' > 0$ and, therefore, strictly convex. Thus, f_i' is strictly increasing, and, interpreted as a function from $[0, 1]$ to $f_i'([0, 1])$, invertible. Denote the inverse by $f_i'^{-1}$. Consider the following optimization problem in the variables $x_1, \ldots, x_n$

$$\text{minimize } \sum_{i=1}^{n} f_i(x_i)$$

$$\text{subject to } x_i \geqslant 0 \text{ for all } i, \text{ and } \sum_{i=1}^{n} x_i = 1.$$

(i) The solution to the problem exists, is unique and of the form $x_i^* = \max\left\{0, f_i'^{-1}(\mu)\right\}$, for some $\mu \in \mathbb{R}$ such that $\sum_i x_i^* = 1$.

(ii) Suppose further, albeit without loss of generality, that $f_1'(0) \leqslant \cdots \leqslant f_n'(0)$. Then, either $f_i'(0) < \mu \leqslant f_{i+1}'(0)$ for $i = 1, \ldots, n-1$, or $f_i'(0) < \mu$ for $i = n$, and for the corresponding index i

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

and

$$\sum_{j=1}^{i} x_j^* = \sum_{j=1}^{i} f_j'^{-1}(\mu) = 1.$$

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 [47] leads to the Lagrangian cost

$$\mathcal{J} = \sum f_i(x_i) - \sum \lambda_i x_i + \mu(1 - \sum x_i)$$

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

$$\begin{aligned} x_i \geqslant 0, \sum x_i = 1 &\quad \text{(primal feasibility)} \\ \lambda_i \geqslant 0 &\quad \text{(dual feasibility)} \\ \lambda_i x_i = 0 &\quad \text{(complementary slackness)} \\ f_i'(x_i) - \lambda_i - \mu = 0 &\quad \text{(dual optimality).} \end{aligned}$$

Eliminating the slack variables λ_i, we may rewrite the complementary slackness and the dual optimality conditions equivalently as

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

Recall that f_i' is strictly increasing, as $f_i'' > 0$. We now consider two cases for each i. First, suppose that $\mu > f_i'(0)$, or equivalently, $f_i'^{-1}(\mu) > 0$. In this case, the only conclusion consistent with the dual optimality condition is $x_i > 0$. But then, the complementary slackness condition implies that $f_i'(x_i) = \mu$, or equivalently, $x_i = f_i'^{-1}(\mu)$. We may interpret this finding as a Pareto equilibrium. Namely, for all positive resources $x_i > 0$, the marginal ratios of improvement $f_i'(x_i)$ must all be the same (μ is a common constant for all i). Otherwise, minor allocation adjustments on the resources could improve the overall objective.

Consider now the opposite case, when $\mu \leqslant f_i'(0)$, or equivalently, $f_i'^{-1}(\mu) \leqslant 0$. Suppose that $x_i > 0$ as in the previous case, so that by complementary slackness, $f_i'(x_i) = \mu \leqslant f_i'(0)$. But this contradicts the fact that f_i' is strictly increasing. Consequently, $x_i = 0$, and in summary

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

This proves claim (i) in the lemma. To verify (ii), observe that whenever $\mu \leqslant f_{i+1}'(0) \leqslant \cdots \leqslant f_n'(0)$ holds for some $i = 0, \ldots, n$, then $f_{i+1}'^{-1}(\mu), \ldots, f_n'^{-1}(\mu) \leqslant 0$, and, therefore, $x_{i+1} = \cdots = x_n = 0$. This argument is valid even for the invalid index $i = 0$, negating the possibility that $\mu \leqslant f_1'(0)$, which would lead to the zero solution, running contrary to the constraint $\sum x_i = 1$. ∎

We now proceed to obtain a closed-form solution for the privacy-redundancy function. Denote by $P_i = \sum_{j=1}^{i} p_j$ and $Q_i = \sum_{j=1}^{i} q_j$ the cumulative distributions corresponding to p and q. Define

$$\rho_i = 1 - \frac{p_i}{P_i q_i + p_i(1 - Q_i)}$$

for $i = 1, \ldots, n$ and $\rho_{n+1} = 1$. Observe that $\rho_1 = 0$, that $\rho_n = 1 - \frac{p_n}{q_n} = \rho_{\text{crit}}$, and consistently with Theorem 2, the solution corresponding to $i = n$ in this last theorem becomes $(1-\rho)\, q_j + \rho\, r_j^* = p_j$, for $j = 1, \ldots, n$. Define

$$\begin{aligned} \tilde{p} &= (P_i, p_{i+1}, \ldots, p_n) \\ \tilde{q} &= (Q_i, q_{i+1}, \ldots, q_n) \\ \tilde{r} &= (1, 0, \ldots, 0) \end{aligned}$$

distributions in the probability simplex of $n - i + 1$ dimensions.

Theorem 4: For any $i = 1, \ldots, n-1$, $\rho_i \leqslant \rho_{i+1}$, with equality if, and only if, $\frac{q_i}{p_i} = \frac{q_{i+1}}{p_{i+1}}$. For any $i = 1, \ldots, n$ and any $\rho \in [\rho_i, \rho_{i+1}]$, the optimal r is given by the equations

$$(1-\rho) q_j + \rho r_j^* = \frac{p_j}{P_i}((1-\rho) Q_i + \rho)$$

for $j = 1, \ldots, i$ and $r_j^* = 0$ for $j = i+1, \ldots, n$. The corresponding, minimum KL divergence yields the privacy-redundancy function

$$\mathcal{R}(\rho) = \mathrm{D}((1-\rho)\tilde{q} + \rho\tilde{r} \| \tilde{p}).$$

Proof: The first statement, regarding the monotonicity of the thresholds ρ_i, can be shown from their definition by routine algebraic manipulation, under the labeling assumption (3). To that end, it is helpful to observe that

$$P_{i+1} q_{i+1} + p_{i+1}(1 - Q_{i+1}) = P_i q_{i+1} + p_{i+1}(1 - Q_i).$$

We shall, however, give a more direct argument within the proof of the rest of this theorem.

Only the nontrivial case $\rho \in (0, 1)$ is shown. Using the definition of KL divergence, write the objective function as $\mathrm{D}(s\|p) = \sum_i s_i \ln \frac{s_i}{p_i}$, with $s = (1-\rho)q + \rho r$. This exposes the structure of the privacy-redundancy optimization problem as a special case of the resource allocation lemma, Lemma 3, with the strictly convex, twice differentiable functions $f_i(r_i) = s_i \ln \frac{s_i}{p_i}$ of r_i. In this special case, $f_i'(r_i) = \rho\left(\ln \frac{s_i}{p_i} + 1\right)$ and

$$f_i'^{-1}(\mu) = \frac{p_i e^{\frac{\mu}{\rho}-1} - (1-\rho) q_i}{\rho}$$

the (Pareto equilibrium) solution for r_i when $r_i > 0$.

Assumption (3) is equivalent to the assumption that $f_i'(0) \leqslant f_{i+1}'(0)$ in the lemma, because $f_i'(0) = \rho\left(\ln \frac{(1-\rho)q_i}{p_i} + 1\right)$ is a strictly increasing function of $\frac{q_i}{p_i}$. On account of the second part of the lemma

$$1 = \sum_{j=1}^{i} \frac{p_j e^{\frac{\mu}{\rho}-1} - (1-\rho) q_j}{\rho} = \frac{P_i e^{\frac{\mu}{\rho}-1} - (1-\rho) Q_i}{\rho}$$

thus

$$\mu = \rho\left(\ln \frac{(1-\rho)Q_i + \rho}{P_i} + 1\right).$$

Combining the expressions for μ and for the optimal r_j when $r_j > 0$, that is, $f_j'^{-1}(\mu)$, leads to the expression for the optimal solution for r in the theorem. It remains to confirm the interval of values of ρ in which it is defined.

To this end, note that the condition $f_i'(0) < \mu$ in the lemma becomes

$$\rho\left(\ln \frac{(1-\rho)q_i}{p_i} + 1\right) < \rho\left(\ln \frac{(1-\rho)Q_i + \rho}{P_i} + 1\right)$$

or equivalently

$$\frac{(1-\rho)q_i}{p_i} < \frac{(1-\rho)Q_i + \rho}{P_i}$$

and after routine algebraic manipulation

$$\rho > 1 - \frac{p_i}{P_i q_i + p_i(1-Q_i)}.$$

One could proceed to carry out an analogous analysis on the upper bound condition $\mu \leqslant f'_{i+1}(0)$ of the lemma to determine the interval of values of ρ in which the solution is defined. However, it is simpler to realize that because a unique solution will exist for each ρ, then the intervals resulting from imposing $f'_i(0) < \mu \leqslant f'_{i+1}(0)$ must be contiguous and nonoverlapping, hence, of the form $(\rho_i, \rho_{i+1}]$. Further, because $\mathcal{R}(\rho)$ is continuous on $[0,1]$, one may write the intervals as $[\rho_i, \rho_{i+1}]$ in lieu of $(\rho_i, \rho_{i+1}]$. This argument also means that the (strict) monotonicity of $\frac{q_i}{p_i}$ is equivalent to the (strict) monotonicity of ρ_i, as stated at the beginning of the theorem.

To complete the proof, it is only left to express the privacy risk $\mathcal{R}(\rho) = \sum_{i=1}^{n} s_i \ln \frac{s_i}{p_i}$ in terms of the optimal distribution of forged queries. But first, we split the sum into two parts. The first term, corresponding to $s_j = \frac{p_j}{P_i}((1-\rho)\,Q_i + \rho)$, is

$$\sum_{j=1}^{i} s_j \ln \frac{s_j}{p_j} = ((1-\rho)Q_i + \rho) \ln \frac{(1-\rho)Q_i + \rho}{P_i}$$

where we exploit the fact that $\frac{s_j}{p_j}$ in the sum does not depend on j. The second term, corresponding to $s_j = (1-\rho)q_j$, is

$$\sum_{j=i+1}^{n} s_j \ln \frac{s_j}{p_j} = \sum_{j=i+1}^{n} (1-\rho) q_j \ln \frac{(1-\rho)q_j}{p_j}.$$

Finally, we immediately identify the terms of $\mathcal{R}(\rho)$ as a divergence between the distribution

$$((1-\rho)Q_i + \rho, (1-\rho)q_{i+1}, \ldots, (1-\rho)q_n)$$

and the distribution $(P_i, p_{i+1}, \ldots, p_n)$. ∎

The optimal forgery strategy of Theorem 4 lends itself to an intuitive interpretation. On the one hand, only queries corresponding to the categories $j = 1, \ldots, i$ are forged, precisely those corresponding to the smallest ratios $\frac{q_j}{p_j}$, loosely speaking, those with probabilities furthest away from the population's distribution. On the other, the optimal user's apparent query distribution within those categories is proportional to the population's, which means that, given that a query belongs to one of these categories, the conditional distribution of submitted queries is equal to the conditional distribution of the population.

A number of conclusions can be drawn from the closed-form solution in Theorem 4. In the following two sections, we focus on the behavior of the privacy-redundancy function at low redundancies on the one hand, and low risk on the other.

D. Low-Redundancy Case

In this section, we characterize $\mathcal{R}(\rho)$ for $\rho \simeq 0$.

Proposition 5 (Low Redundancy): In the nontrivial case when $p \neq q$, there exists a positive integer i with redundancy thresholds satisfying $0 = \rho_1 = \cdots = \rho_i < \rho_{i+1}$. For all $\rho \in [0, \rho_{i+1}]$, the optimal forgery strategy r^* contains i nonzero components, and the slope of the privacy-redundancy function at the origin is $\mathcal{R}'(0) = \ln \frac{q_1}{p_1} - \mathrm{D}(q\|p)$.

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

$$\frac{q_1}{p_1} = \cdots = \frac{q_i}{p_i} < \frac{q_{i+1}}{p_{i+1}} \leqslant \cdots \leqslant \frac{p_n}{q_n}.$$

At this point, we observe that the ratio between the ith components of the cumulative distributions equals the common ratio $\frac{q_1}{p_1}$

$$\frac{Q_i}{P_i} = \frac{q_1 + \cdots + q_i}{p_1 + \cdots + p_i} = \frac{\frac{q_1}{p_1}p_1 + \cdots + \frac{q_1}{p_1}p_i}{p_1 + \cdots + p_i} = \frac{q_1}{p_1}.$$

On account of Theorem 4

$$0 = \rho_1 = \cdots = \rho_i < \rho_{i+1} \leqslant \cdots \leqslant \rho_n$$

and for all $\rho \in [0, \rho_{i+1}]$

$$\mathcal{R}(\rho) = \mathrm{D}\left((1-\rho)\tilde{q} + \rho(1,0,\ldots,0)\|\tilde{p}\right).$$

It is routine to check that

$$\frac{\mathrm{d}}{\mathrm{d}\rho}\mathrm{D}((1-\rho)\tilde{q} + \rho\tilde{r}\|\tilde{p})\bigg|_{\rho=0} = \sum_j (\tilde{r}_j - \tilde{q}_j) \ln \frac{\tilde{q}_j}{\tilde{p}_j}$$

and to compute the slope of the privacy-redundancy function at the origin

$$\begin{aligned}\mathcal{R}'(0) &= \ln \frac{\tilde{q}_1}{\tilde{p}_1} - \mathrm{D}(\tilde{q}\|\tilde{p}) \\ &= \ln \frac{Q_i}{P_i} - \mathrm{D}((Q_i, q_{i+1}, \ldots, q_n)\|(P_i, p_{i+1}, \ldots, p_n))\end{aligned}$$

but $\frac{Q_i}{P_i} = \frac{q_1}{p_1}$, therefore

$$\begin{aligned}&\mathrm{D}((Q_i, q_{i+1}, \ldots, q_n)\|(P_i, p_{i+1}, \ldots, p_n)) \\ &\quad = Q_i \ln \frac{Q_i}{P_i} + \sum_{j=i+1}^{n} q_j \ln \frac{q_j}{p_j} = \\ &\quad = (q_1 + \cdots + q_i) \ln \frac{q_1}{p_1} + \sum_{j=i+1}^{n} q_j \ln \frac{q_j}{p_j} = \mathrm{D}(q\|p)\end{aligned}$$

and finally, $\mathcal{R}'(0) = \ln \frac{q_1}{p_1} - \mathrm{D}(q\|p)$. ∎

Define the relative decrement factor

$$\delta = -\frac{\mathcal{R}'(0)}{\mathcal{R}(0)} = 1 + \frac{\ln \frac{p_1}{q_1}}{\mathrm{D}(q\|p)}.$$

Proposition 5 means that

$$\mathcal{R}(\rho) \simeq (1-\rho)\mathrm{D}(q\|p) + \rho \ln \frac{q_1}{p_1}$$

for $\rho \simeq 0$ or, in terms of relative decrement

$$\frac{\mathrm{D}(q\|p) - \mathcal{R}(\rho)}{\mathrm{D}(q\|p)} \simeq \delta\rho. \tag{4}$$

Conceptually speaking, the ratio $\frac{q_1}{p_1}$ characterizes the privacy gain at low redundancy, together with $\mathrm{D}(q\|p)$, in contrast to the fact that the ratio $\frac{q_n}{p_n}$ determines ρ_{crit}, the maximum redundancy

for zero privacy risk, defined in Section IV-B. We mentioned in that section that $\frac{q_n}{p_n} \geqslant 1$. An entirely analogous argument shows that $\frac{q_1}{p_1} \leqslant 1$, thus $\ln \frac{q_1}{p_1} \leqslant 0$ in our first order approximations, and consequently $\delta > 1$. In other words, the relative risk reduction (4) is conveniently greater than the redundancy introduced. The risk decrement at low redundancies becomes less noticeable with worse ratio $\frac{q_1}{p_1}$ (closer to 1), for a fixed $\mathrm{D}(q\|p)$. We may, however, improve our bound on the relative decrement factor δ, as the next proposition shows.

Proposition 6 (Relative Decrement): In the nontrivial case when $p \neq q$, $\delta\rho_{\mathrm{crit}} \geqslant 1$, with equality if, and only if, $\mathcal{R}(\rho)$ is affine.

Proof: The statement of the proposition is a consequence of Theorems 1 and 2. Since $p \neq q$, it is clear that $\mathcal{R}(0) = \mathrm{D}(q\|p) > 0$, and, as argued in Section IV-B, $\rho_{\mathrm{crit}} > 0$. Consider now any continuous, convex function $\mathcal{R}(\rho)$, defined at least on $[0, \rho_{\mathrm{crit}}]$, satisfying $\mathcal{R}(0) > 0$, with ρ_{crit} a positive root. A straight line with slope at the origin $\mathcal{R}'(0)$ must fall under the segment connecting $\mathcal{R}(0)$ at $\rho = 0$, and 0 at $\rho = \rho_{\mathrm{crit}}$, strictly so unless the function is affine. Mathematically, $\mathcal{R}'(0) \leqslant \frac{-\mathcal{R}(0)}{\rho_{\mathrm{crit}}}$, or equivalently, $\delta = \frac{-\mathcal{R}'(0)}{\mathcal{R}(0)} \geqslant \frac{1}{\rho_{\mathrm{crit}}}$. ∎

The proof of Proposition 6 suggests that $\delta\,\rho_{\mathrm{crit}} \geqslant 1$ is a measure of the convexity of $\mathcal{R}(\rho)$, and ties the behavior of the function at low redundancies and low risk. On account of this bound, the relative risk reduction (4) satisfies $\delta\rho \geqslant \frac{\rho}{\rho_{\mathrm{crit}}}$, which means that the relative risk reduction cannot be smaller than the redundancy introduced, relative to its critical value. This convenient result is fairly intuitive from the graphical representation of Fig. 2, in Section IV-B. An explicit restatement of the proposition leads to the interesting inequality $\mathrm{D}(q\|p) \leqslant \left(\frac{q_n}{p_n} - 1\right) \ln \frac{p_1}{q_1}$.

E. Low-Risk Case

We now turn to the case when $\rho \simeq \rho_{\mathrm{crit}}$, and thus, $\mathcal{R}(\rho) \simeq 0$. Use $i = n - 1$ to confirm that, if $p \neq q$

$$\mathcal{R}(\rho) = \mathrm{D}((1-\rho)(1-q_n, q_n) + \rho(1,0)\|(1-p_n, p_n)) > 0$$

whenever $\rho \in [\rho_{n-1}, \rho_{\mathrm{crit}}]$. (Recall that $\rho_n = \rho_{\mathrm{crit}}$.) Evidently, we are assuming that $\frac{q_{n-1}}{p_{n-1}} \neq \frac{q_n}{p_n}$ so that, on account of Theorem 4, $\rho_{n-1} < \rho_{\mathrm{crit}}$, to avoid an empty interval. More explicitly

$$\mathcal{R}(\rho) = (1-(1-\rho)q_n) \ln \frac{1-(1-\rho)q_n}{1-p_n} + (1-\rho)q_n \ln \frac{(1-\rho)q_n}{p_n}.$$

From this expression, it is routine to conclude that $\mathcal{R}'(\rho_{\mathrm{crit}}) = 0$ and

$$\mathcal{R}''(\rho_{\mathrm{crit}}) = \frac{q_n^2}{p_n - p_n^2} = \frac{p_n}{1-p_n} \frac{1}{(1-\rho_{\mathrm{crit}})^2}$$

(left-side differentiation), and finally

$$\mathcal{R}(\rho) \simeq \frac{1}{2}\mathcal{R}''(\rho_{\mathrm{crit}})(\rho - \rho_{\mathrm{crit}})^2.$$

We would like to remark that the fact that $\mathcal{R}(\rho)$ admits a quadratic approximation for $\rho \simeq \rho_{\mathrm{crit}}$, with $\mathcal{R}'(\rho_{\mathrm{crit}}) = 0$, may be concluded immediately from the fundamental properties of the Fisher information [27]. 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 the Fisher information. Denote by $s^*_\rho = (1-\rho)\,q + \rho\, r^*$ the family of optimal apparent distributions, indexed by the redundancy. Theorem 2 guarantees that $s^*_{\rho_{\mathrm{crit}}} = p$, thus we may write $\mathcal{R}(\rho) = \mathrm{D}(s^*_\rho\|s^*_{\rho_{\mathrm{crit}}})$. Under this formulation, it is clear that the Fisher information associated with the redundancy is $\mathrm{I}(\rho_{\mathrm{crit}}) = \mathcal{R}''(\rho_{\mathrm{crit}})$.

Finally, the observation at the end of Section II that $r^*_n = 0$ at $\rho = \rho_{\mathrm{crit}}$ is consistent with the fact that ρ_{crit} is the endpoint of the interval corresponding the the solution for r^* with $n-1$ nonzero components in Theorem 4.

F. Maximizing the Entropy of the User's Query Distribution

We mentioned in Section III-A that Shannon's entropy could be regarded as a special case of KL divergence. As in that section, let $p = u$ be the uniform distribution on $\{1, \ldots, n\}$, so that the privacy risk becomes

$$\mathrm{D}((1-\rho)q + \rho r\|u) = \ln n - \mathrm{H}((1-\rho)q + \rho r).$$

This of course means that our theoretical analysis addresses the problem of maximizing the entropy of the user's apparent query distribution. In this case, our assumption on the labeling of the probabilities becomes $q_1 \leqslant \cdots \leqslant q_n$.

Clearly, the critical redundancy of Theorem 2 becomes $\rho_{\mathrm{crit}} = 1 - \frac{1}{nq_n}$. Theorem 4 gives the forgery distribution r maximizing the entropy, namely

$$(1-\rho)q_j + \rho r^*_j = \frac{1}{i}((1-\rho)Q_i + \rho)$$

for $j = 1, \ldots, i$, and $r^*_j = 0$ otherwise. Observe that the resulting apparent distribution is constant, water filled if you wish, for $j = 1, \ldots, i$, the indices corresponding to the smallest values of q. The thresholds of the intervals of ρ are $\rho_i = 1 - \frac{1}{iq_i + 1 - Q_i}$, and the maximum privacy attainable

$$\mathrm{H}((1-\rho)q + \rho r^*) = ((1-\rho)Q_i + \rho)\ln i + \mathrm{H}((1-\rho)\tilde{q} + \rho\tilde{r}).$$

V. Simple, Conceptual Examples

In this section, we illustrate the formulation of Section II and the theoretic analysis of Section IV with numerical results for two simple, intuitive examples. First, we contemplate the special case of entropy maximization described in Section IV-F, and secondly, we address the more general case of divergence minimization in Section IV-C.

A. Entropy Maximization

We set the population's distribution to the uniform distribution

$$p = u = (1/3, 1/3, 1/3) \simeq (0.333, 0.333, 0.333)$$

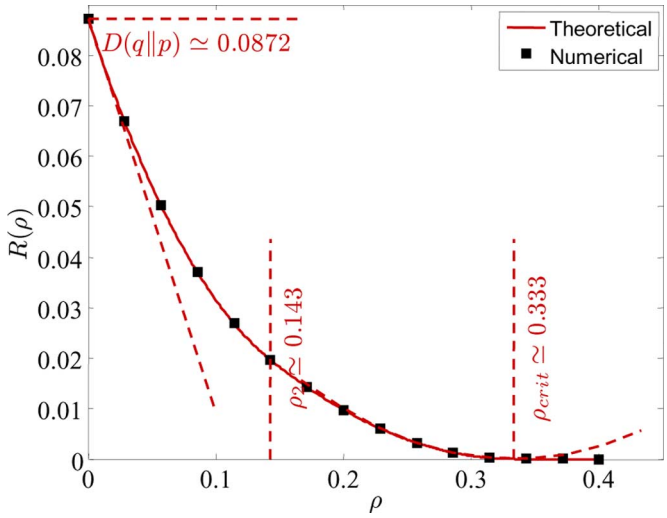


Fig. 5. Entropy maximization example. Privacy-redundancy function for $p = (1/3, 1/3, 1/3)$ and $q = (1/6, 1/3, 1/2)$. $\mathcal{R} \simeq 1.10 - \mathrm{H}(s)$.

across three categories, so that

$$\mathcal{R} = \mathrm{D}(s\|u) = \ln 3 - \mathrm{H}(s) \simeq 1.10 - \mathrm{H}(s)$$

and assume the user's distribution

$$q = (1/6, 1/3, 1/2) \simeq (0.167, 0.333, 0.5).$$

The three categories are sorted to satisfy the labeling assumption (3).

The redundancy thresholds ρ_i of Theorem 4 are $\rho_1 = 0$, $\rho_2 \simeq 0.143$ and $\rho_3 = \rho_{\mathrm{crit}} = 1/3 \simeq 0.333$. The initial privacy risk, without query forgery, is $\mathcal{R}(0) \simeq 0.0872 (\mathrm{H}(q) \simeq 1.01)$, and the first and second-order approximations of Sections IV-D and IV-E, characterizing the privacy-redundancy function $\mathcal{R}(\rho)$ (1) at extreme values of the redundancy ρ, are determined by the quantities $\mathcal{R}'(0) \simeq -0.780$ and $\mathcal{R}''(\rho_{\mathrm{crit}}) \simeq 1.13$. The resulting function $\mathcal{R}(\rho)$ has been computed both theoretically, applying Theorem 4, and numerically.[1] The function is depicted in Fig. 5, along with the corresponding thresholds and approximations.

We now turn to analyze the optimal apparent distribution $s^* = (1 - \rho)q + \rho r^*$ for several interesting values of ρ, which contains the optimal forged query distribution r^*. The population's distribution p, the user's distribution q and the apparent distribution s^* are shown in the probability simplexes represented in Fig. 6. The contours correspond to the divergence $\mathrm{D}(\cdot\|p)$ from a point in the simplex to the reference distribution p. The smaller triangle depicts the subsimplex $\{s = (1 - \rho)q + \rho r\}$ of possible apparent query distributions, not necessarily optimal, for a fixed ρ. In Fig. 6(a), a redundancy $\rho < \rho_2$ below the first nonzero threshold has been chosen to verify that in this case $r^* = (1, 0, 0)$. In the notation of Theorem 4, r^* has exactly $i = 1$ nonzero components, which, geometrically, places the solution s^* at one vertex of the subsimplex. It is also interesting to notice that a redundancy of just 8% lowers the privacy risk to a 44% of the original risk $\mathrm{D}(q\|p)$.

[1]The numerical method chosen is the interior-point optimization algorithm [47] implemented by the Matlab R2008a function `fmincon`.

The coefficient in the relative decrement formula (4) for low redundancies of Section IV-D is $\delta \simeq 8.95$, conveniently high, and $\delta\, \rho_{\mathrm{crit}} \simeq 2.98 > 1$, consistently with Proposition 6.

In Fig. 6(b), $\rho_2 < \rho < \rho_{\mathrm{crit}}$, thus r^* contains $i = 2$ nonzero components, which places the solution s^* on one edge of the subsimplex. In this case, a redundancy of 25% reduces the risk to a mere 4% of its original value. The case $\rho = \rho_{\mathrm{crit}}$, for which $\mathcal{R}(\rho) = 0$, is shown in Fig. 6(c), where still $i = 2$, thus $r_3^* = 0$, as argued at the end of Section IV-B. The impractical case when $\rho > \rho_{\mathrm{crit}}$ is represented in Fig. 6(d), which leads to $i = 3$ and s^* in the interior of the subsimplex.

B. Divergence Minimization

We assume that the population's distribution is

$$p = (5/12, 1/3, 1/4) \simeq (0.417, 0.333, 0.25)$$

and the user's distribution

$$q = (1/6, 1/3, 1/2) \simeq (0.167, 0.333, 0.5).$$

The three categories are sorted to satisfy the labeling assumption (3)

$$\left(\frac{q_i}{p_i}\right)_i = (2/5, 1, 2) = (0.4, 1, 2).$$

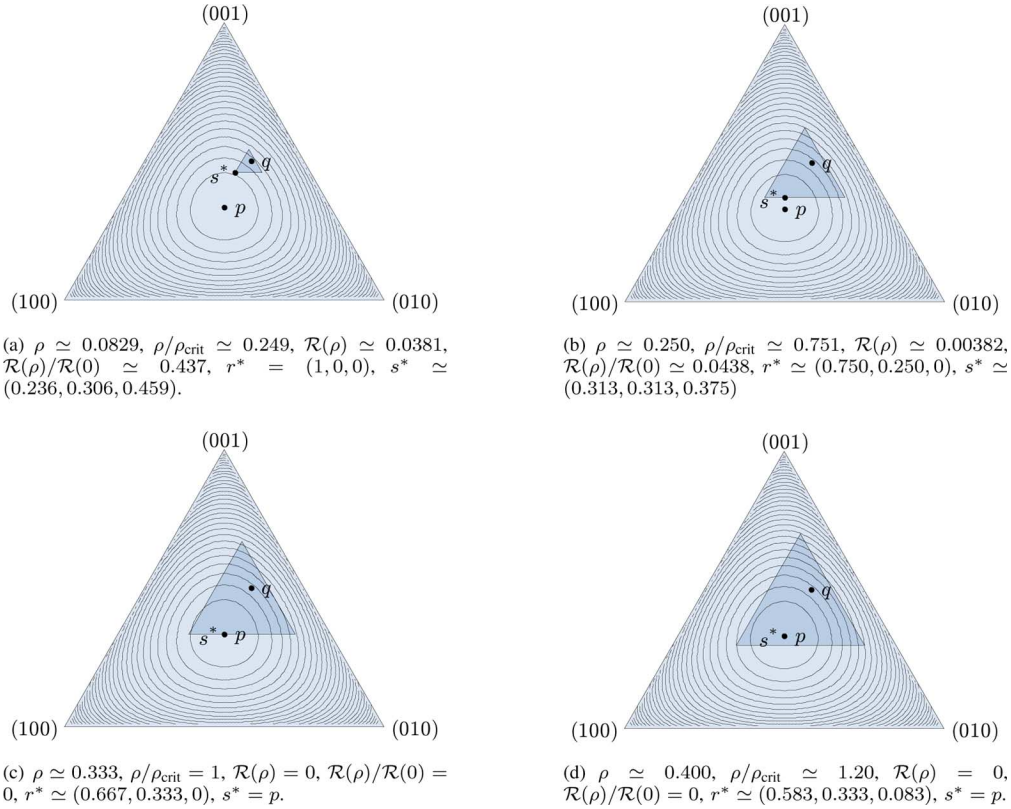
The redundancy thresholds ρ_i are $\rho_1 = 0$, $\rho_2 = 0.2$ and $\rho_3 = \rho_{\mathrm{crit}} = 0.5$. The initial privacy risk is $\mathcal{R}(0) \simeq 0.194$, and the first and second-order approximations are determined by $\mathcal{R}'(0) \simeq -1.11$ and $\mathcal{R}''(\rho_{\mathrm{crit}}) \simeq 1.33$. The resulting function $\mathcal{R}(\rho)$ is depicted in Fig. 7, along with the corresponding thresholds and approximations.

We now turn to analyze the optimal apparent distribution s^* for several interesting values of ρ, which contains the optimal forged query distribution r^*. The population's distribution p, the user's distribution q and the apparent distribution s^* are shown in the probability simplexes represented in Fig. 8. In Fig. 8(a), a redundancy $\rho < \rho_2$ below the first nonzero threshold has been chosen to verify that in this case $r^* = (1, 0, 0)$. Observe that r^* has exactly $i = 1$ nonzero components, which, geometrically, places the solution s^* at one vertex of the subsimplex. It is also interesting to notice that a redundancy of just 12% lowers the privacy risk to a 46% of the original risk $\mathrm{D}(q\|p)$. The coefficient in the relative decrement formula for low redundancies is $\delta \simeq 5.73$, again conveniently high, and $\delta\, \rho_{\mathrm{crit}} \simeq 2.87 > 1$.

In Fig. 8(b), $\rho_2 < \rho < \rho_{\mathrm{crit}}$, thus r^* contains $i = 2$ nonzero components, which places the solution s^* on one edge of the subsimplex. In this case, a redundancy of 38% reduces the risk to a mere 5% of its original value. The case $\rho = \rho_{\mathrm{crit}}$, for which $\mathcal{R}(\rho) = 0$, is shown in Fig. 8(c), where still $i = 2$, thus $r_3^* = 0$. The impractical case when $\rho > \rho_{\mathrm{crit}}$ is represented in Fig. 8(d), which leads to $i = 3$ and s^* in the interior of the subsimplex.

VI. Concluding Remarks

There exists a large number of solutions to the problem of PIR, in the widest sense of the term, each one with its own advantages and disadvantages. Query forgery is by itself a simple strategy in terms of infrastructure requirements, which does not involve placing the user's trust on an external entity, namely a



(a) $\rho \simeq 0.0829$, $\rho/\rho_{\text{crit}} \simeq 0.249$, $\mathcal{R}(\rho) \simeq 0.0381$, $\mathcal{R}(\rho)/\mathcal{R}(0) \simeq 0.437$, $r^* = (1,0,0)$, $s^* \simeq (0.236, 0.306, 0.459)$.

(b) $\rho \simeq 0.250$, $\rho/\rho_{\text{crit}} \simeq 0.751$, $\mathcal{R}(\rho) \simeq 0.00382$, $\mathcal{R}(\rho)/\mathcal{R}(0) \simeq 0.0438$, $r^* \simeq (0.750, 0.250, 0)$, $s^* \simeq (0.313, 0.313, 0.375)$

(c) $\rho \simeq 0.333$, $\rho/\rho_{\text{crit}} = 1$, $\mathcal{R}(\rho) = 0$, $\mathcal{R}(\rho)/\mathcal{R}(0) = 0$, $r^* \simeq (0.667, 0.333, 0)$, $s^* = p$.

(d) $\rho \simeq 0.400$, $\rho/\rho_{\text{crit}} \simeq 1.20$, $\mathcal{R}(\rho) = 0$, $\mathcal{R}(\rho)/\mathcal{R}(0) = 0$, $r^* \simeq (0.583, 0.333, 0.083)$, $s^* = p$.

Fig. 6. Entropy maximization example. Probability simplexes showing p, q and s^* for several values of ρ. (a) $\rho \simeq 0.0829, \rho/\rho_{\text{crit}} \simeq 0.249, \mathcal{R}(\rho) \simeq 0.0381, \mathcal{R}(\rho)/\mathcal{R}(0) \simeq 0.437, r^* = (1,0,0), s^* \simeq (0.236, 0.306, 0.459)$. (b) $\rho \simeq 0.250, \rho/\rho_{\text{crit}} \simeq 0.751, \mathcal{R}(\rho) \simeq 0.00382, \mathcal{R}(\rho)/\mathcal{R}(0) \simeq 0.0438, r^* = (0.750, 0.250, 0), s^* \simeq (0.313, 0.313, 0.375)$. (c) $\rho \simeq 0.333, \rho/\rho_{\text{crit}} = 1, \mathcal{R}(\rho) = 0, \mathcal{R}(\rho)/\mathcal{R}(0) = 0, r^* \simeq (0.667, 0.333, 0), s^* = p$. (d) $\rho \simeq 0.400, \rho/\rho_{\text{crit}} \simeq 1.20, \mathcal{R}(\rho) = 0, \mathcal{R}(\rho)/\mathcal{R}(0) = 0, r^* \simeq (0.583, 0.333, 0.083), s^* = p$.

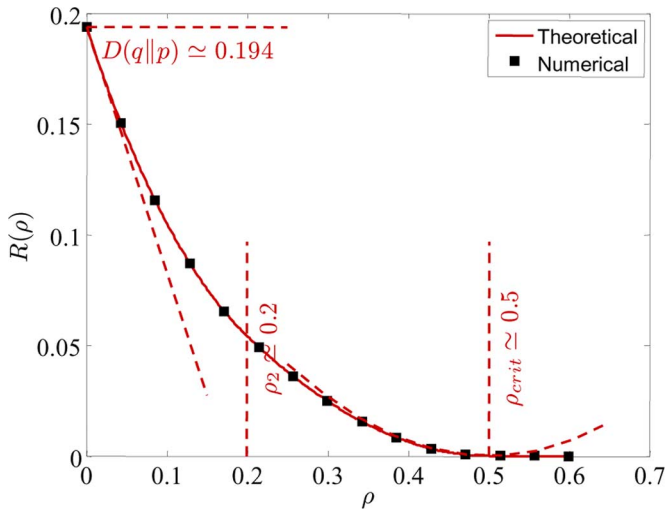


Fig. 7. Divergence minimization example. Privacy-redundancy function for $p = (5/12, 1/3, 1/4)$ and $q = (1/6, 1/3, 1/2)$.

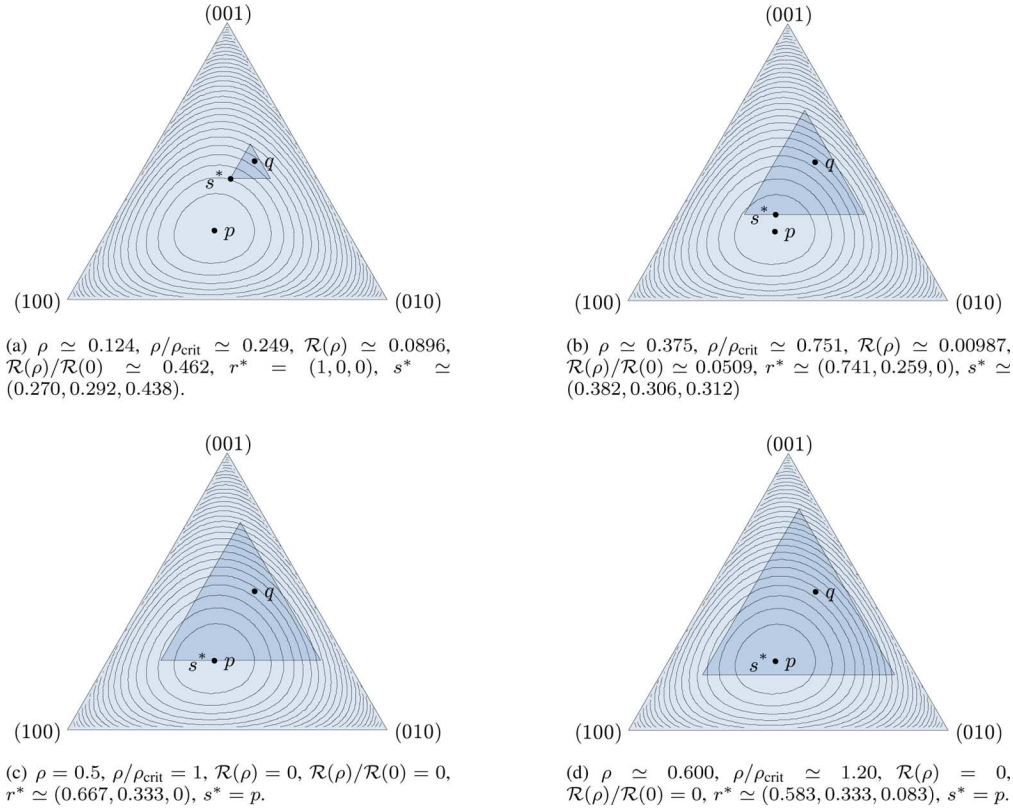
TTP. It is also part of more complex protocols, such as [18]. However, query forgery comes at the cost of traffic and processing overhead. In other words, there is a patent trade-off between privacy and redundancy.

Our main contribution is a systematic, mathematical approach to the problem of optimal query forgery for PIR. Precisely, we carefully justify a measure of privacy, and formulate and solve an optimization problem modeling the privacy-redundancy trade-off. Inspired by our previous work on statistical disclosure control [23], the privacy risk is measured as the KL divergence between the user's apparent query distribution, containing dummy queries, and the population's. Our formulation contemplates, as a special case, the maximization of the entropy of the user's distribution.

Queries are modeled fairly generally by random variables, which might in fact not only represent complete queries, but also query categories, individual keywords in a small indexable set, parts of queries such as coordinates sent to an LBS provider, and even sequences of related queries. This work, however, is limited to relative frequencies, relevant against content-based attacks. That is to say, it does not address differences in absolute frequencies, which could be exploited in traffic analysis.

We justify our privacy criterion by interpreting it from different perspectives, and by connecting it to the extensive rationale behind entropy maximization and divergence minimization in the literature. Our interpretations are based on the AEP, hypothesis testing and Stein's lemma, axiomatizations of (relative) entropy, Bayesian inference, and the average information gain criteria in [23].

In spite of its information-theoretic appeal and mathematical tractability, we must acknowledge that the adequacy of our formulation ultimately lies in the appropriateness of the criteria optimized, which in turn depends on the specific application, on



(a) $\rho \simeq 0.124$, $\rho/\rho_{\text{crit}} \simeq 0.249$, $\mathcal{R}(\rho) \simeq 0.0896$, $\mathcal{R}(\rho)/\mathcal{R}(0) \simeq 0.462$, $r^* = (1,0,0)$, $s^* \simeq (0.270, 0.292, 0.438)$.

(b) $\rho \simeq 0.375$, $\rho/\rho_{\text{crit}} \simeq 0.751$, $\mathcal{R}(\rho) \simeq 0.00987$, $\mathcal{R}(\rho)/\mathcal{R}(0) \simeq 0.0509$, $r^* \simeq (0.741, 0.259, 0)$, $s^* \simeq (0.382, 0.306, 0.312)$

(c) $\rho = 0.5$, $\rho/\rho_{\text{crit}} = 1$, $\mathcal{R}(\rho) = 0$, $\mathcal{R}(\rho)/\mathcal{R}(0) = 0$, $r^* \simeq (0.667, 0.333, 0)$, $s^* = p$.

(d) $\rho \simeq 0.600$, $\rho/\rho_{\text{crit}} \simeq 1.20$, $\mathcal{R}(\rho) = 0$, $\mathcal{R}(\rho)/\mathcal{R}(0) = 0$, $r^* \simeq (0.583, 0.333, 0.083)$, $s^* = p$.

Fig. 8. Divergence minimization example. Probability simplexes showing p, q and s^* for several values of ρ. (a) $\rho \simeq 0.124, \rho/\rho_{\text{crit}} \simeq 0.249, \mathcal{R}(\rho) \simeq 0.0896, \mathcal{R}(\rho)/\mathcal{R}(0) \simeq 0.462, r^* = (1,0,0), s^* \simeq (0.270, 0.292, 0.438)$. (b) $\rho \simeq 0.375, \rho/\rho_{\text{crit}} \simeq 0.751, \mathcal{R}(\rho) \simeq 0.00987, \mathcal{R}(\rho)/\mathcal{R}(0) \simeq 0.0509, r^* = (0.741, 0.259, 0), s^* \simeq (0.382, 0.306, 0.312)$. (c) $\rho = 0.5, \rho/\rho_{\text{crit}} = 1, \mathcal{R}(\rho) = 0, \mathcal{R}(\rho)/\mathcal{R}(0) = 0, r^* \simeq (0.667, 0.333, 0), s^* = p$. (d) $\rho \simeq 0.600, \rho/\rho_{\text{crit}} \simeq 1.20, \mathcal{R}(\rho) = 0, \mathcal{R}(\rho)/\mathcal{R}(0) = 0, r^* \simeq (0.583, 0.333, 0.083), s^* = p$.

the query statistics of the users, on the actual network and processing overhead incurred by introducing forged queries, and last but not least, on the adversarial model and the mechanisms against privacy contemplated. In a way, this is not unlike the occasionally controversial issue of whether mean-squared error is a suitable distortion measure in lossy compression.

We present a closed-form solution for the optimal forgery strategy and a privacy-redundancy function $\mathcal{R}(\rho)$ characterizing the optimal trade-off. Our theoretical analysis bears certain resemblance to the water-filling problem in rate-distortion theory, and is restricted to the discrete case of n query categories. We show that the privacy-redundancy function $\mathcal{R}(\rho)$ is convex, and that there exists a critical redundancy ρ_{crit} beyond which perfect privacy is attainable. This ρ_{crit} only depends on the largest ratio $\frac{q_j}{p_j}$ of probabilities between the user's query distribution q and the population's p. For a given redundancy ρ, the optimal query forgery distribution contains i nonzero components associated with the $i = 1, \ldots, n$ smallest ratios $\frac{q_j}{p_j}$, or in the entropy case, associated with the i smallest q_j. The number of nonzero categories i increases with ρ, being $i = n - 1$ for $\rho = \rho_{\text{crit}}$. Intuitively, this is a greedy approach. The optimal user's apparent query distribution within those categories is proportional to the population's, which means that, given that a query belongs to one of these categories, the conditional distribution of submitted queries is equal to the conditional distribution of the population.

Further, we characterize $\mathcal{R}(\rho)$ at low redundancies and low risks. We provide a first-order approximation for $\rho \simeq 0$ whenever $p \neq q$, which turns out to be a convex combination, governed by ρ, between $\mathcal{R}(0)$ and the smallest log-ratio $\frac{q_j}{p_j}$. A convenient consequence of the convexity of $\mathcal{R}(\rho)$ is that the relative risk decrement at low redundancies cannot be smaller than the redundancy introduced, relative to its critical value. We provide a second-order approximation for $\rho \simeq \rho_{\text{crit}}$, assuming there is a strictly largest ratio $\frac{q_j}{p_j}$. We interpret the fact that $\mathcal{R}'(\rho_{\text{crit}}) = 0$ as a consequence of a fundamental property of the Fisher information.

ACKNOWLEDGMENT

The authors would like to thank J. Parra-Arnau, with the Department of Telematics Engineering at the Technical University of Catalonia, the editor of this journal, and the anonymous reviewers for their careful reading and helpful comments.

REFERENCES

[1] R. Dingledine, Free Haven's Anonymity Bibliography 2009 [Online]. Available: http://www.freehaven.net/anonbib/

[2] Y. Elovici, C. Glezer, and B. Shapira, "Enhancing customer privacy while searching for products and services on the World Wide Web," *Internet Res.*, vol. 15, no. 4, pp. 378–399, 2005.

[3] R. Puzis, D. Yagil, Y. Elovici, and D. Braha, "Collaborative attack on Internet user's anonymity," *Internet Res.*, vol. 19, no. 1, pp. 60–77, 2009.

[4] D. Chaum, "Security without identification: Transaction systems to make big brother obsolete," *Commun. ACM*, vol. 28, no. 10, pp. 1030–1044, Oct. 1985.
[5] V. Benjumea, J. López, and J. M. T. Linero, "Specification of a framework for the anonymous use of privileges," *Telemat., Informat.*, vol. 23, no. 3, pp. 179–195, Aug. 2006.
[6] G. Bianchi, M. Bonola, V. Falletta, F. S. Proto, and S. Teofili, "The SPARTA pseudonym and authorization system," *Sci. Comput. Program.*, vol. 74, no. 1–2, pp. 23–33, 2008.
[7] M. Duckham, K. Mason, J. Stell, and M. Worboys, "A formal approach to imperfection in geographic information," *Comput., Environ., Urban Syst.*, vol. 25, no. 1, pp. 89–103, 2001.
[8] C. Chow, M. F. Mokbel, and 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, Nov. 2006, pp. 171–178.
[9] P. Samarati and L. Sweeney, Protecting Privacy When Disclosing Information k-Anonymity and its Enforcement Through Generalization and Suppression, SRI Int, 1998, Tech. Rep.
[10] P. Samarati, "Protecting respondents' identities in microdata release," *IEEE Trans. Knowl. Data Eng.*, vol. 13, no. 6, pp. 1010–1027, 2001.
[11] J. Domingo-Ferrer, "Microaggregation for database and location privacy," in *Proc. Int. Workshop Next-Gen. Inform. Technol., Syst. (NGITS)*, Kibbutz Shefayim, Israel, Jul. 2006, vol. 4032, pp. 106–116.
[12] A. Solanas and A. Martínez-Ballesté, "A TTP-free protocol for location privacy in location-based services," *Comput. Commun.*, vol. 31, no. 6, pp. 1181–1191, Apr. 2008.
[13] R. Ostrovsky and W. E. Skeith, III, "A survey of single-database PIR: Techniques and applications," in *Proc. Int. Conf. Practice, Theory Public-Key Cryptogr. (PKC)*, Beijing, China, Sep. 2007, vol. 4450, pp. 393–411.
[14] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private queries in location based services: Anonymizers are not necessary," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Vancouver, BC, Canada, Jun. 2008, pp. 121–132.
[15] T. Kuflik, B. Shapira, Y. Elovici, and A. Maschiach, "Privacy preservation improvement by learning optimal profile generation rate," in *User Model.*, 2003, vol. 2702/2003, pp. 168–177.
[16] B. Shapira, Y. Elovici, A. Meshiach, and T. Kuflik, "PRAW—The model for PRivAte Web," *J. Amer. Soc. Inf. Sci., Technol.*, vol. 56, no. 2, pp. 159–172, 2005.
[17] H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of location privacy using dummies for location-based services," in *Proc. IEEE Int. Conf. Data Eng. (ICDE)*, Washington, DC, Oct. 2005, pp. 1248–1248.
[18] D. Rebollo-Monedero, J. Forné, A. Solanas, and T. Martínez-Ballesté, "Private location-based information retrieval through user collaboration," *Comput. Commun.*, vol. 33, no. 6, pp. 762–774, 2010.
[19] D. Rebollo-Monedero, J. Forné, L. Subirats, A. Solanas, and A. Martínez-Ballesté, "A collaborative protocol for private retrieval of location-based information," presented at the IADIS Int. Conf. e-Society, Barcelona, Spain, Feb. 2009.
[20] D. C. Howe and H. Nissenbaum, TrackMeNot 2006 [Online]. Available: http://mrl.nyu.edu/~dhowe/trackmenot
[21] V. Toubiana, SquiggleSR 2007 [Online]. Available: http://www.squigglesr.com
[22] C. Soghoian, "The problem of anonymous vanity searches," *I/S: J. Law, Policy Inf. Soc. (ISJLP)*, Jan. 2007.
[23] D. Rebollo-Monedero, J. Forné, and J. Domingo-Ferrer, "From t-closeness-like privacy to postrandomization via information theory," *IEEE Trans. Knowl. Data Eng.* Oct. 2009.
[24] D. Rebollo-Monedero, J. Forné, and J. Domingo-Ferrer, "From t-closeness to PRAM and noise addition via information theory," in *Proc. Privacy Stat. Databases (PSD)*, Istanbul, Turkey, Sep. 2008, pp. 100–112.
[25] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *Proc. IEEE Int. Conf. Data Eng. (ICDE)*, Istanbul, Turkey, Apr. 2007, pp. 106–115.
[26] C. Díaz, S. Seys, J. Claessens, and B. Preneel, "Towards measuring anonymity," in *Proc. Workshop Privacy Enhanc. Technol. (PET)*, Apr. 2002, vol. 2482.
[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.
[28] Google Insights [Online]. Available: http://www.google.com/insights/search
[29] "Communication theory of secrecy systems," *Bell Syst.*, 1949.
[30] "The wiretap channel," *Bell Syst.*, 1975.
[31] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 339–348, May 1978.
[32] P. M. Woodward, "Theory of radar information," in *Proc. London Symp. Inf. Theory, Ministry of Supply*, London, U.K., 1950, pp. 108–113.
[33] D. V. Lindley, "On a measure of the information provided by an experiment," *Ann. Math. Statist.*, vol. 27, no. 4, pp. 986–1005, 1956.
[34] C. Díaz, "Anonymity and Privacy in Electronic Services," Ph.D. dissertation, Katholieke Univ. Leuven, Belgium, 2005.
[35] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev. Ser. II*, vol. 106, no. 4, pp. 620–630, 1957.
[36] E. T. Jaynes, "Information theory and statistical mechanics II," *Phys. Rev. Ser. II*, vol. 108, no. 2, pp. 171–190, 1957.
[37] J. Uffink, "Can the maximum entropy principle be explained as a consistency requirement?," *Stud. Hist., Philos. Mod. Phys.*, vol. 26B, pp. 223–261, 1995.
[38] E. T. Jaynes, "On the rationale of maximum-entropy methods," *Proc. IEEE*, vol. 70, pp. 939–952, Sep. 1982.
[39] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 1, pp. 26–37, Jan. 1980.
[40] J. Shore and R. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Inf. Theory*, vol. IT-27, pp. 472–482, Jul. 1981.
[41] J. Skilling, *Maximum-Entropy and Bayesian Methods in Science and Engineering*. Dordrecht, The Netherlands: Kluwer, 1988.
[42] "A mathematical theory of communication," *Bell Syst.*, 1948.
[43] D. K. Fadeev, "Zum begriff der entropie einer endlichen wahrscheinlichkeitsschemnas," *Arbeiten Zur Informationstheorie*, vol. 1, pp. 85–90, 1957.
[44] A. Feinstein, *Foundations of Information Theory*. New York: McGraw-Hill, 1958.
[45] A. Rényi, "On measures of entropy and information," in *Proc. Berkeley Symp. Math., Stat., Prob.*, Berkeley, CA, Jun. 1961, pp. 547–561.
[46] A. Giffin and A. Caticha, "Updating probabilities with data and moments," in *Proc. Int. Workshop Bayesian Infer., Max. Ent. Methods Sci., Eng. (MaxEnt)*, Saratoga Springs, NY, Jul. 2007.
[47] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

David Rebollo-Monedero received the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, 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, Barcelona, Spain, from 1997 to 2000. He is currently a postdoctoral researcher with the Information Security Group, Department of Telematics, Technical University of Catalonia, 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 and the Ph.D. degree from the Technical University of Catalonia (UPC), Spain, in 1992 and 1997, respectively.

In 1991, he joined the Cryptography and Network Security Group, Department of Applied Mathematics and Telematics. Currently, he is with the Information Security Group, Department of Telematics Engineering, UPC. His research interests span a number of subfields within information security and privacy, including network security, electronic commerce, and public key infrastructures. Currently, he is an Associate Professor at the Telecommunications Engineering School, UPC.