Un Criterio de Privacidad Basado en Teoría de la Información para la Generación de Consultas Falsas

David Rebollo-Monedero
Dpto. Ingeniería Telemática,
Universitat Politècnica de Catalunya
david.rebollo@entel.upc.edu

Javier Parra-Arnau
Dpto. Ingeniería Telemática,
Universitat Politècnica de Catalunya
javier.parra@entel.upc.edu

Jordi Forné
Dpto. Ingeniería Telemática,
Universitat Politècnica de Catalunya
jforne@entel.upc.edu

***Resumen*—En este artículo presentamos un criterio de privacidad basado en teoría de la información para la generación de consultas falsas en el ámbito de la recuperación de información privada. Medimos el riesgo de privacidad como la divergencia de Kullback y Leibler entre la distribución de consultas del usuario y la de la población, que incluye la entropía de la distribución del usuario como caso especial. Asimismo, llevamos a cabo una rigurosa justificación de nuestra métrica al interpretarla desde distintas perspectivas de teoría de la información, desde la propiedad de equipartición asintótica, pasando por los fundamentos sobre los que sustentan los métodos de maximización de la entropía, la minimización de la divergencia y la minimización de la ganancia de información, hasta el lema de Stein.**

I. Introducción

Durante las últimas dos décadas, Internet se ha ido integrando de manera gradual en nuestra vida diaria. Una de las actividades más frecuentes que llevan a cabo los usuarios cuando navegan por la Web es enviar una consulta a un motor de búsqueda. Los motores de búsqueda permiten a los usuarios recuperar información sobre una gran variedad de categorías, tales como hobbies, deportes, negocios o salud. Sin embargo, la mayoría de usuarios no son conscientes de los riesgos de privacidad que ello entraña [1].

De noviembre a diciembre de 2008, el 61 % de los adultos en Estados Unidos buscaron información en la red sobre una enfermedad en particular, un tratamiento específico, y otros temas relacionados [2]. Dichas consultas podrían revelar información sensible y ser utilizada para construir perfiles de usuario sobre enfermedades potenciales. Esta información privada podría acabar más tarde en las manos de un empresario y frustrar las esperanzas de uno de sus empleados.

En la literatura sobre sistemas de recuperación de información abundan los casos como el descrito, en los que se constata la importancia de la privacidad del usuario. Estos casos incluyen no sólo el riesgo de que los usuarios puedan ser caracterizados por un motor de búsquedas de Internet, sino también por proveedores de servicios basados en la localización (LBS, *location-based services*), o incluso la caracterización de empresas por parte de proveedores de bases de datos de patentes o mercados de valores. En este contexto, la falsificación de consultas, que consiste en acompañar consultas auténticas con consultas falsas, emerge como una posible solución para garantizar la privacidad del usuario hasta un cierto punto, a costa de una sobrecarga de tráfico y procesado.

Este artículo presenta un nuevo criterio de privacidad basado en teoría de la información para la generación de consultas en el ámbito de la recuperación de información. En concreto, nuestro criterio mide el riesgo de privacidad como una divergencia entre la distribución de consultas del usuario y la de la población, y contempla la entropía de la distribución del usuario como un caso particular. El objeto de este artículo es interpretar y justificar nuestra métrica de privacidad desde distintas perspectivas, a través de la propiedad de equipartición asintótica, el test de hipótesis y el lema de Stein.

La Sección II revisa las propuestas más relevantes en cuanto a recuperación de información privada y criterios de privacidad. La Sección III repasa algunos conceptos fundamentales relacionados con teoría de la información que ayudarán a entender la esencia de este trabajo. La Sección IV presenta una formulación de teoría de la información sobre el compromiso entre privacidad y redundancia para la falsificación de consultas en el contexto de recuperación de información privada. Esta sección muestra nuestra medida de privacidad, y posteriormente la interpreta y justifica. Finalmente, en la Sección V se presentan las conclusiones.

II. Estado del Arte en Recuperación de Información Privada

A lo largo de este artículo, utilizaremos el término recuperación de información privada (PIR, *private information retrieval*) en su sentido más amplio, queriendo decir con ello que no nos ceñiremos a las técnicas basadas en criptografía normalmente relacionadas con este acrónimo. Por consiguiente, nos referiremos a un escenario más genérico en el que los usuarios envían consultas de propósito general a un proveedor de servicios de información. Un ejemplo sería un usuario que enviase la consulta: "¿Cuál es la película más taquillera en la categoría de ciencia ficción?". A continuación, revisaremos las contribuciones más destacadas para PIR sobre la generación de consultas falsas y criterios de privacidad.

II-A. Recuperación de Información Privada

En el ámbito de la recuperación de información privada, existen una gran variedad de propuestas. Algunas de ellas se basan en terceras partes de confianza (TTPs, *trusted third parties*) que actúan como intermediario entre los usuarios y el proveedor de servicios de información [3]. Aunque este tipo

de soluciones garantizan la privacidad del usuario gracias a que su identidad es, de hecho, desconocida para el proveedor de servicios, la confianza del usuario únicamente se traslada de una entidad a otra.

Como alternativa, algunas propuestas que no se basan en TTPs, utilizan técnicas de perturbación. En el caso concreto de LBS, los usuarios perturban su información de localización al consultar a un proveedor de servicios [4]. Esto proporciona a los usuarios un cierto nivel de privacidad en términos de localización, pero no así en cuanto al contenido de las consultas y la actividad del usuario. Asimismo, esta técnica plantea un compromiso entre privacidad y utilidad de los datos: cuanto mayor es la perturbación de la localización, mayor es la privacidad del usuario, pero menor la precisión de las respuestas del proveedor de servicios. Como alternativa, los métodos criptográficos para PIR permiten a un usuario recuperar, de forma privada, el contenido de una base datos indexado por una dirección de memoria enviada por el usuario, haciendo que sea inviable por parte del proveedor de la base de datos averiguar qué entradas fueran recuperadas [5]. Desafortunadamente, este tipo de métodos requieren la cooperación del proveedor en el protocolo de privacidad, se restringen hasta cierto punto a funciones de consulta-respuesta en forma de tablas de búsqueda de longitud finita con respuestas precomputadas, y conllevan una significativa carga computacional.

La generación de consultas falsas, que el centro de nuestra discusión, aparece como una alternativa a los métodos anteriores. La idea subyacente consiste en enviar consultas originales junto con consultas falsas. A pesar de la sencillez de este método, la falsificación de consultas es capaz de garantizar la privacidad del usuario hasta un cierto punto, a costa de una sobrecarga de tráfico y procesado, aunque sin tener que tener confiar ni en el proveedor de información ni en el operador de red.

Basándose en este principio, se han propuesto e implementado varios protocolos PIR. En [6], [7], se presenta una solución que pretende preservar la privacidad de un grupo de usuarios que navegan por Internet compartiendo un punto de acceso a la Web. Los autores proponen la generación de transacciones falsas, i.e., accesos a páginas web para frustrar a un atacante en su intento por caracterizar al grupo. La privacidad se mide como la similitud entre el perfil real de un grupo de usuarios y el observado por el atacante [6].

Además de las implicaciones legales, existen distintas consideraciones técnicas para la preservación de la privacidad a través de la generación de consultas falsas [8], puesto que los atacantes podrían analizar no sólo el contenido de las consultas sino también la actividad, el ritmo de generación, el enrutamiento o cualquier otro parámetro del protocolo de transmisión, por medio de varias consultas o a través de diversos servicios de información. Asimismo, se espera que tanto los proveedores de información como los de la red se muestren reticentes a la generación automática de consultas falsas, con lo que cualquier esquema que se precie debe tener en cuenta la sobrecarga de tráfico.

II-B. Criterios de Privacidad

En esta sección revisaremos una serie de técnicas propuestas originalmente para el control de revelación estadístico (SDC, *statistical disclosure control*), pero igualmente aplicables a PIR, la aplicación que motiva nuestro trabajo. En privacidad de bases de datos, se define un *conjunto de microdatos* como una tabla de base de datos cuyos registros contienen información sobre encuestados individuales. Específicamente, este conjunto contiene atributos clave, es decir, atributos que, utilizados conjuntamente, se pueden relacionar con información externa para reidentificar a los encuestados a los que se refieren los registros en el conjunto de microdatos. Como ejemplo, los atributos clave podrían ser trabajo, dirección, edad, género, peso y altura. De igual modo, el conjunto de microdatos contiene atributos confidenciales con información sensible sobre el encuestado, tales como sueldo, religión o afiliación política.

Un planteamiento habitual en SDC es la microagregación, que consiste, primero, en dividir el conjunto de datos en grupos de registros con tuplas de valores de atributos clave similares, y segundo, en reemplazar las tuplas de cada registro en cada uno de los grupos por una tupla representativa del grupo. Uno de los criterios de privacidad más populares en la anonimización de bases de datos, es k-anonimato [9]. Este criterio se puede lograr a través de la microagregación, ya que requiere que cada combinación de atributos clave sea compartida por al menos k registros en el conjunto de microdatos. Sin embargo, el principal inconveniente de este criterio y de sus posteriores mejoras [10]–[12] es su vulnerabilidad ante los ataques de similitud y *skewness* [13]. Con el objeto de superar estas deficiencias, [14] propone otro criterio de privacidad. Concretamente, un conjunto de datos satisface t-closeness si, para cada grupo de registros que comparten una combinación de atributos clave, la divergencia de Kullback y Leibler (KL) entre la distribución de atributos confidenciales dentro de un grupo y la distribución de estos atributos en el conjunto de datos global no supera un umbral t. Inspirados en esta idea, [15], [16] definen riesgo de privacidad como una versión promediada del requisito impuesto por t-closeness sobre el conjunto de grupos agregados. Otro criterio de privacidad basado en teoría de la información propone medir el grado de anonimato observable por un atacante como la entropía de la distribución de probabilidad de los posibles emisores de un determinado mensaje [17], [18].

A pesar de las propuestas citadas anteriormente, querríamos poner énfasis en la posible necesidad, por parte de algunas aplicaciones, de criterios de privacidad basados en teoría de la información más sofisticados que k-anonimato o sus respectivas mejoras.

III. Introducción a Conceptos de Teoría de la Información

A lo largo de este artículo, denominaremos alfabeto al espacio medible en el que una variable aleatoria (v.a.) toma valores. Seguiremos la convención de utilizar mayúsculas para las v.a.'s, y minúsculas para los valores particulares que éstas

pueden tomar. Las funciones de densidad de probabilidad (PDFs, *probability density functions*) y las funciones de masa de probabilidad (PMFs, *probability mass functions*) son denotadas por p, subindexadas por sus correspondientes v.a.'s en caso de ambigüedad. Por ejemplo, tanto $p_X(x)$ como $p(x)$ indican el valor de la función p_X en x, lo que ayuda a escribir ecuaciones más concisas. De manera informal, nos referiremos ocasionalmente a la función p como $p(x)$. Asimismo, utilizaremos la notación $p_{X|Y}$ y $p(x|y)$ de manera equivalente.

En este artículo, adoptamos la misma notación utilizada en [19] para cantidades de teoría de la información. En concreto, el símbolo H se referirá a la entropía y D a la entropía relativa o divergencia KL. A continuación recordamos muy brevemente varios conceptos de teoría de la información para aquellos lectores que no estén íntimamente familiarizados con este campo. Por simplicidad, utilizaremos logaritmos neperianos.

- La *entropía* $\mathrm{H}(X)$ de una v.a. discreta X con distribución de probabilidad p es una medida de su incertidumbre, y se define como

$$\mathrm{H}(X) = -\,\mathrm{E}\,\ln p(X) = -\sum_x p(x)\ln p(x),$$

  donde E es el operador esperanza. Este operador es sustituido por la integral cuando p es una PDF.

- Dadas dos distribuciones de probabilidad $p(x)$ y $q(x)$ sobre el mismo alfabeto, la *divergencia KL* o *entropía relativa* $\mathrm{D}(p\,\|\,q)$ se define, en el caso discreto, como

$$\mathrm{D}(p\,\|\,q) = \mathrm{E}_p\,\ln\frac{p(X)}{q(X)} = \sum_x p(x)\ln\frac{p(x)}{q(x)}.$$

  Cuando p y q son PDFs, la esperanza se transforma en una integral.

Aunque la divergencia KL no satisface la propiedad de simetría y la desigualdad triangular, nos da una medida de la distancia o discrepancia entre distribuciones, en el sentido que $\mathrm{D}(p\,\|\,q) \geq 0$, con igualdad si y sólo si $p = q$.

Este intuitivo sentido de distancia se hace más evidente al examinar el lema de Stein. Suponga que observamos una secuencia de k v.a.'s independientes e idénticamente distribuidas (i.i.d.'s), y que necesitamos evaluar si éstas han sido generadas según una distribución de probabilidad p_1, hipótesis $\mathscr{H}_1$, o p_2, hipótesis $\mathscr{H}_2$. Dadas estas dos hipótesis, definimos la *región de aceptación* $\mathscr{A}_k$ como el conjunto de secuencias que, una vez observadas, nos llevan a aceptar $\mathscr{H}_1$. De forma análoga, definimos el complemento de este conjunto, $\bar{\mathscr{A}_k}$, como el conjunto de secuencias que nos decantan por $\mathscr{H}_2$. A continuación, contemplamos las siguientes probabilidades de error:

- la probabilidad de un falso negativo α_k, definido como la probabilidad de aceptar $\mathscr{H}_2$ cuando $\mathscr{H}_1$ es cierta,
- y la probabilidad de un falso positivo β_k, definido como la probabilidad de aceptar $\mathscr{H}_1$ cuando $\mathscr{H}_2$ es cierta.

Suponga que elegimos una región de aceptación con la intención de minimizar β_k, mientras que no permitimos que α_k exceda un cierto umbral ϵ. En términos generales, el lema de Stein afirma que la tasa de error óptima, β_k^ϵ, es aproximadamente $e^{-k\,\mathrm{D}(p_1\,\|\,p_2)}$, para valores de k grandes y ϵ pequeños.

A modo de ejemplo, considere el test de hipótesis en el que observamos una secuencia $X_1, \ldots, X_k$ de k lanzamientos i.i.d.'s de una v.a., e intentamos averiguar si se han producido de acuerdo con una distribución gaussiana $p_1 = \mathscr{N}(d/2, \sigma^2)$ o $p_2 = \mathscr{N}(-d/2, \sigma^2)$. Teniendo en cuenta estas distribuciones, elegiríamos la región de decisión óptima $\mathscr{A}_k$ dada por el lema de Neyman-Pearson [19], y calcularíamos la probabilidad de un falso positivo como la integral de $p_2(x_1, \ldots, x_k)$ sobre $\mathscr{A}_k$. Resulta que, a partir del lema de Stein, esta probabilidad es aproximadamente $e^{\frac{-k\,d^2}{2\sigma^2}}$, ya que $\mathrm{D}(p_1\,\|\,p_2) = \frac{d^2}{2\sigma^2}$, lo que hace más palpable esta noción de distancia: cuanto mayor es la distancia real d entre las medias de las dos distribuciones, mayor es la divergencia KL, y menor la probabilidad de error al distinguir entre ambas distribuciones.

IV. Un Criterio de Privacidad de Teoría de la Información para la Falsificación de Consultas

Esta sección presenta la principal contribución de este trabajo, un nuevo criterio de privacidad basado en una cantidad de teoría de la información para la falsificación de consultas en PIR. En concreto, la Sección IV-A introduce nuestra medida de privacidad, lo que nos conduce al problema de optimización mostrado en la Sección IV-B en el que se presenta el compromiso óptimo entre riesgo de privacidad y redundancia. Posteriormente, interpretamos y justificamos nuestra medida de privacidad desde distintos puntos de vista. En particular, la Sección IV-C investiga los fundamentos sobre los que se sustentan los métodos de maximización de la entropía, la minimización de la divergencia y la minimización de la ganancia de información. Para comprender estos argumentos, revisamos la propiedad de equipartición asintótica, el test de hipótesis y el lema de Stein.

IV-A. Criterio de Privacidad

Nuestro modelo matemático representa las *consultas* de usuario como v.a.'s que toman valores en un alfabeto común. Asumiremos que las consultas de usuario no son elaboradas o detalladas. En su lugar, éstas se referirán a un conjunto de categorías o temas, o de forma equivalente, podrán representar palabras clave en un conjunto indexable reducido. Por consiguiente, consideraremos que el alfabeto es finito. En concreto, asumiremos que las consultas toman valores en el alfabeto $\mathscr{X} = \{1, \ldots, n\}$ para algún $n \in \mathbb{Z}^+$.

Teniendo en cuenta estas consideraciones, definiremos p como la distribución de consultas de la *población*, q como la distribución real de un *usuario* en particular, y r como la distribución de las consultas *falsificadas* de ese usuario. Asimismo, consideraremos un parámetro de *redundancia* de consultas $0 \leqslant \rho \leqslant 1$, que será el ratio entre consultas falsificadas y consultas totales. De acuerdo con esto, definiremos la

distribución de consultas *aparente* del usuario s como la combinación convexa $(1-\rho)\,q+\rho\,r$, que será la distribución que en realidad observará el proveedor de servicios de información, o simplemente, un atacante de la privacidad. Un atacante será capaz de comprometer la privacidad de un usuario siempre que la distribución de consultas aparente de este usuario difiera de la distribución de consultas de la población.

Inspirados por los criterios de privacidad propuestos en [14]–[17], definimos el *riesgo de privacidad inicial* como la divergencia KL entre la distribución del usuario y la de la población, es decir, $\mathcal{R}_0 = \mathrm{D}(q \,\|\, p)$. De forma similar, definimos el *riesgo de privacidad final* $\mathcal{R}$ como la divergencia KL entre la distribución aparente y la distribución de la población, es decir,

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

IV-B. Compromiso Óptimo entre Privacidad de Consultas y Redundancia

Esta sección muestra una formulación del compromiso entre privacidad y redundancia para la generación de consultas falsas, que surge de la medida de privacidad presentada en la Sección IV-A. Partiendo de la definición de nuestro criterio de privacidad, supondremos que la población es suficientemente grande como para despreciar el impacto de la elección de r en p. De esta forma, definimos la función *privacidad-redundancia*

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

que representa el compromiso óptimo entre riesgo de privacidad de consultas y redundancia.

El término *mínimo* que aparece en la definición de la función privacidad-redundancia está justificado por el hecho de que el problema de optimización planteado implica una función acotada inferiormente y semi-continua inferiormente sobre un conjunto compacto, que es el símplex de probabilidad al que pertenece r.

Teniendo en cuenta esta formulación, conviene apreciar que es posible obtener resultados teóricos análogos para una definición alternativa del riesgo de privacidad, dada por la inversión de los argumentos de la divergencia KL. En la Sección IV-C2 se dan más detalles sobre esta formulación alternativa.

IV-C. Interpretación y Justificación

En esta sección interpretaremos y justificaremos la divergencia KL como criterio de privacidad en la definición de la función privacidad-redundancia. En concreto, examinaremos los argumentos en la literatura que abogan por la maximización de la entropía, y la minimización de la divergencia y la ganancia de información.

Antes de proceder a la interpretación y justificación de nuestro criterio de privacidad, querríamos comentar que, aunque nuestra propuesta surge de una cantidad de teoría de la información y resulta matemáticamente tratable, la adecuación de nuestra formulación está supeditada a la adaptación de los criterios optimizados, que a su vez depende de varios factores tales como la propia aplicación, el modelo de adversario y los mecanismos en contra de la privacidad que se hayan contemplado. Las interpretaciones y justificaciones que aquí se detallan tienen por objeto ayudar a los diseñadores y usuarios de sistemas a evaluar la adecuación de nuestra propuesta a una aplicación específica de recuperación de información.

Asimismo, querríamos poner énfasis en que, a pesar de que nuestro criterio de privacidad se basa en una cantidad fundamental de teoría de la información, la convergencia de estos dos campos en absoluto es nueva. De hecho, el trabajo de Shannon en los años cincuenta ya introdujo el concepto de *equivocación* como la entropía condicional de un mensaje privado dada la observación de un criptograma [20], utilizada más tarde en la formulación del problema *wiretap channel* [21], [22] como una medida de confidencialidad. Del mismo modo, podemos mencionar la interpretación basada en teoría de la información de la divergencia entre las distribuciones a priori y a posterior, denominada *ganancia de información promedio* en algunos campos de estadística [23], [24]. Asimismo, estudios recientes [17] reafirman la adecuación y aplicabilidad del concepto de entropía como medida de privacidad, tal y como comentamos en la Sección II.

IV-C1. Maximización de la Entropía: Nuestra primera interpretación está basada, de hecho, en la idea de que la entropía de Shannon se puede considerar como un caso particular del criterio propuesto en este artículo. Para comprender esta conexión, suponga que la distribución de consultas de la población es la distribución uniforme u sobre el alfabeto $\mathscr{X}$, es decir, que $u_i = 1/n$ para todo $i \in \mathscr{X}$. En este supuesto, el riesgo de privacidad se puede expresar como

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

Por tanto, minimizar la divergencia KL es equivalente a maximizar la entropía de la distribución de consultas aparente del usuario:

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

Esta equivalencia nos conduce a las siguientes dos implicaciones. En primer lugar, el criterio de privacidad $\mathrm{H}((1-\rho)\,q+\rho\,r)$ es una medida de *ganancia* de privacidad, más que de *riesgo* de privacidad. En segundo lugar, se trata de una medida de privacidad *absoluta*, en contraste con nuestro criterio más general, en el sentido que es una métrica *relativa* a cualquier distribución de referencia.

El hecho de considerar esta medida absoluta de ganancia de privacidad permite acercarnos a los fundamentos sobre los que se apoyan los métodos de *maximización de la entropía*. Algunos de estos argumentos están relacionados con el mayor número de permutaciones con repetición asociado a una distribución empírica [25]. Sin embargo, el argumento más apropiado para la justificación de la maximización de la entropía, considerada como una medida de privacidad,

viene dado por la propiedad de equipartición asintótica (AEP, *asymptotic equipartition property*) [19, §3].

Suponga una secuencia $X_1, \ldots, X_k$ de k consultas i.i.d.'s, que toman valores en $\mathscr{X}$, y son generadas de acuerdo con la distribución de consultas aparente del usuario $s = (1 - \rho)\, q + \rho\, r$. Para k suficientemente grande, la AEP sostiene que es muy probable que la secuencia de consultas $x_1, \ldots, x_k$ pertenezca a un subconjunto $\mathscr{T}^{(k)}$ del conjunto de todas las posibles secuencias, denominado *conjunto típico*, que satisface estas propiedades: la probabilidad de este conjunto es aproximadamente 1, todos los elementos son casi equiprobables, y el número de elementos es prácticamente $e^{k\,\mathrm{H}(s)}$. Resulta que la entropía está acotada superiormente por $\ln n$, como consecuencia de la no negatividad de la divergencia KL, y alcanza su valor máximo cuando la distribución aparente es la distribución uniforme. A partir de esta observación, podemos deducir que la distribución uniforme maximiza el conjunto típico $\mathscr{T}^{(k)}$ y, cuando esto sucede, éste se convierte en el conjunto de todos los posibles resultados, conteniendo n^k secuencias. Puesto que $\mathrm{H}(s)$ caracteriza completamente esta aproximación, cualquier medida de privacidad con sentido acabaría siendo básicamente equivalente a ésta.

Teniendo en cuenta esta conexión entre entropía y tamaño del conjunto típico, ahora describiremos la siguiente amenaza de privacidad. Suponga que un atacante intenta adivinar una secuencia de k consultas de un usuario en particular a partir de la observación de secuencias previas. Cuanto mayor sea la entropía $\mathrm{H}(s)$ de la distribución de consultas aparente del usuario, mayor será el tamaño del conjunto típico $\mathscr{T}^{(k)}$ de secuencias posibles e igualmente probables de k consultas, y mayor la probabilidad de que la secuencia a adivinar sea significativamente diferente de las anteriores. Este escenario nos permite concluir que los métodos de maximización de la entropía contribuyen ampliamente a la protección de la privacidad del usuario.

IV-C2. Minimización de la Divergencia: En la sección anterior examinamos los argumentos que abogan por la maximización de la entropía. En esta sección, recurriremos al lema de Stein, revisado en la Sección III, para nuestra interpretación de la divergencia como falsos positivos y falsos negativos. En concreto, describiremos un escenario en el que un atacante utiliza test de hipótesis para comprometer la privacidad del usuario.

En el resto de la sección, consideraremos nuestro criterio de privacidad en su sentido más amplio, es decir, la distribución de consultas del usuario no se comparará necesariamente, en términos de la divergencia KL, con la distribución uniforme.

Nuestra interpretación contempla el escenario en el que un atacante conoce, o es capaz de estimar, la distribución de consultas aparente s de un usuario determinado. Además, suponemos que el atacante observa una secuencia de k consultas i.i.d.'s, e intenta adivinar si éstas han sido generadas por ese usuario o no. Exactamente, el atacante considera el test de hipótesis binario entre dos alternativas: si las consultas se han producido de acuerdo con la distribución aparente del usuario s, hipótesis $\mathscr{U}$, o la distribución general de la población p, hipótesis $\mathscr{P}$.

Llegados a este punto, un atacante podría llevar a cabo dos estrategias mutuamente excluyentes. La primera estrategia considera que el atacante está interesado en acotar la probabilidad de un falso negativo $\mathrm{P}(\mathscr{P}|\mathscr{U})$, dado que su objetivo es que el usuario no pase desapercibido. A partir del lema de Stein, encontramos que la probabilidad $\mathrm{P}(\mathscr{P}|\mathscr{U})$ de un falso positivo es aproximadamente $e^{-k\,\mathrm{D}(s\,\|\,p)}$ para k grande. Por consiguiente, la minimización de $\mathrm{D}(s\,\|\,p)$ en la definición de la función privacidad-redundancia (1) implica la maximización del exponente en la tasa de error de falsos positivos. Dicho de otra forma, la distribución óptima de consultas falsas r^* frustra a un atacante en su esfuerzo por reconocer a un usuario de entre la población, y por tanto, comprometer la privacidad del usuario.

Más que fijar la probabilidad de un falso negativo, ahora el objetivo del atacante es minimizar la probabilidad de error global

$$P_{\mathrm{T}} = \mathrm{P}(\mathscr{U})\,\mathrm{P}(\mathscr{P}|\mathscr{U}) + \mathrm{P}(\mathscr{P})\,\mathrm{P}(\mathscr{U}|\mathscr{P}).$$

Aprovechándose del hecho de que la actividad de la población global es mucho mayor que la de un único usuario, el atacante está interesado en acotar $\mathrm{P}(\mathscr{U}|\mathscr{P})$, y hacer lo posible para minimizar $\mathrm{P}(\mathscr{P}|\mathscr{U})$. Resulta que la probabilidad de un falso negativo dado por el lema de Stein es aproximadamente $e^{-k\,\mathrm{D}(p\,\|\,s)}$, lo que justifica una definición alternativa de la función privacidad-redundancia dada por la inversión de los dos argumentos de la divergencia KL. De acuerdo con esta observación, la estrategia de falsificación de consultas r^* que minimiza $\mathrm{D}(p\,\|\,s)$, conduce a la maximización de la probabilidad de error global del atacante y contribuye a proteger la privacidad del usuario.

A modo de aclaración, nos gustaría destacar que, a pesar de que esta definición alternativa resulta oportuna en el último escenario propuesto, nosotros creemos que la formulación original es más apropiada, ya que incluye, como caso particular, los métodos de maximización de la entropía descritos en la Sección IV-C1.

IV-C3. Minimización de la Ganancia de Información: Una vez analizados los principales argumentos en pro de la maximización de la entropía y la minimización de la divergencia, ahora estableceremos una conexión entre nuestro criterio de privacidad y el criterio propuesto en [16].

Considere $p_{Q|U}(q|u)$ la distribución de consultas del usuario u, donde U es una variable aleatoria que identifica a un usuario en particular y toma el valor u. Asimismo, Q es una variable aleatoria que representa una consulta en particular, y toma el valor q.

Sea $p_Q(q)$ la distribución de probabilidad sin condicionar que modela la distribución de consultas de la población. Naturalmente, $p_U(u)$ sería la probabilidad de usuario, posiblemente ponderada por su actividad. En esta notación, nuestra medida de riesgo de privacidad para el usuario u se puede escribir como $\mathrm{D}(p_{Q|U}(\cdot|u)\|p_Q)$. De forma similar, podemos aplicarla

para redefinir el concepto de t-closeness. Una distribución satisface t-closeness si y sólo si $\mathrm{D}(p_{Q|U}(\cdot|u)\|p_Q) \leq t$ para todos los valores u de U, lo que sugiere medir el riesgo de privacidad como un máximo sobre divergencias. Inspirados por t-closeness, [16] presenta un planteamiento más interesante en el sentido que nos permite conectar con la ganancia de información promedio. En concreto, el criterio de privacidad propuesto en [16] es la divergencia KL condicional

$$\mathrm{D}(p_{Q|U}\|p_Q) = \mathrm{E}_U\, \mathrm{D}(p_{Q|U}(\cdot|U)\|p_Q),$$

es decir, la información mutua entre Q y U, o de forma equivalente, el promedio entre usuarios del criterio de privacidad definido en este artículo. En contraste con este criterio, nuestra medida de privacidad contempla un único usuario, pero podría, en principio, generalizarse a escenarios multiusuarios en una futura propuesta.

V. Conclusiones

Existe una gran variedad de propuestas para PIR, considerado aquí en el sentido más amplio del término. Dentro de estas soluciones, la generación de consultas falsas surge como una estrategia simple en términos de requisitos de infraestructura, ya que los usuarios no necesitan una entidad externa en la que confiar. Sin embargo, esta solución plantea un compromiso entre la privacidad y el coste de la sobrecarga de tráfico y procesado.

Nuestra principal contribución es un criterio de privacidad basado en teoría de la información para la falsificación de consultas en PIR, que emerge de la formulación del compromiso entre privacidad y redundancia. Inspirados por el trabajo en [16], medimos el riesgo de privacidad como la divergencia KL entre la distribución de consultas aparente del usuario, que contiene consultas falsas, y la de la población. Nuestra formulación contempla, como caso especial, la maximización de la entropía de la distribución del usuario.

En este artículo justificamos nuestro criterio de privacidad al interpretarlo desde distintas perspectivas, y al conectarlo con los argumentos en la literatura que abogan por la maximización de la entropía, la minimización de la divergencia y la minimización de la ganancia de información. Nuestras interpretaciones están basadas en la AEP, el test de hipótesis y el lema de Stein, y el criterio de ganancia de información promedio propuesto en [16].

Aunque nuestra propuesta surge de una medida de teoría de la información y resulta matemáticamente tratable, la adecuación de nuestra formulación está supeditada a la adaptación de los criterios optimizados, que a su vez depende de varios factores tales como la propia aplicación, la estadística de consultas de los usuarios, la sobrecarga de red y de procesado provocados por la consultas falsas, el modelo de adversario y los mecanismos en contra de la privacidad que se hayan contemplado.

Agradecimientos

Este trabajo ha sido financiado en parte por el gobierno español mediante los proyectos CONSOLIDER INGENIO 2010 CSD2007-00004 “ARES” y TSI2007-65393-C02-02 “ITAC”, y por el gobierno catalán bajo la subvención 2009 SGR 1362.

Referencias

[1] D. Fallows, “Search engine users,” Pew Internet and American Life Project, Tech. Rep., Jan. 2005.

[2] S. Fox and S. Jones, “The social life of health information,” Pew Internet and American Life Project, Tech. Rep., Jun. 2009.

[3] C. C. M. F. Mokbel and W. G. Aref, “The new casper: query processing for location services without compromising privacy,” in *Proc. Int. Conf. on Very Large Data Bases*. VLDB J., 2006, pp. 763–774.

[4] 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.

[5] 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)*, ser. Lecture Notes Comput. Sci. (LNCS), vol. 4450. Beijing, China: Springer-Verlag, Sep. 2007, pp. 393–411.

[6] B. S. Y. Elovici and A. Maschiach, “A new privacy model for hiding group interests while accessing the web,” in *Proc. ACM Workshop on Privacy in the Electron. Society*. ACM, 2002, pp. 63–70.

[7] B. Shapira, Y. Elovici, A. Meshiach, and T. Kuflik, “PRAW – The model for PRivAte Web,” *J. Amer. Soc. Inform. Sci., Technol.*, vol. 56, no. 2, pp. 159–172, 2005.

[8] C. Soghoian, “The problem of anonymous vanity searches,” *I/S: J. Law, Policy Inform. Soc. (ISJLP)*, Jan. 2007.

[9] P. Samarati and L. Sweeney, “Protecting privacy when disclosing information: k-Anonymity and its enforcement through generalization and suppression,” SRI Int., Tech. Rep., 1998.

[10] X. Sun, H. Wang, J. Li, and T. M. Truta, “Enhanced p-sensitive k-anonymity models for privacy preserving data publishing,” *Trans. Data Privacy*, vol. 1, no. 2, pp. 53–66, 2008.

[11] A. Machanavajjhala, J. Gehrke, D. Kiefer, and M. Venkitasubramanian, “l-Diversity: Privacy beyond k-anonymity,” in *Proc. IEEE Int. Conf. Data Eng. (ICDE)*, Atlanta, GA, Apr. 2006, p. 24.

[12] H. Jian-min, C. Ting-ting, and Y. Hui-qun, “An improved V-MDAV algorithm for l-diversity,” in *Proc. IEEE Int. Symp. Inform. Processing (ISIP)*, Moscow, Russia, May 2008, pp. 733–739.

[13] J. Domingo-Ferrer and V. Torra, “A critique of k-anonymity and some of its enhancements,” in *Proc. Workshop Privacy, Security, Artif. Intell. (PSAI)*, Barcelona, Spain, 2008, pp. 990–993.

[14] 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.

[15] D. Rebollo-Monedero, J. Forné, and J. Domingo-Ferrer, “From t-closeness to PRAM and noise addition via information theory,” in *Privacy Stat. Databases (PSD)*, ser. Lecture Notes Comput. Sci. (LNCS). Istambul, Turkey: Springer-Verlag, Sep. 2008, pp. 100–112.

[16] ——, “From t-closeness-like privacy to postrandomization via information theory,” *IEEE Trans. Knowl. Data Eng.*, Oct. 2009. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TKDE.2009.190

[17] C. Díaz, S. Seys, J. Claessens, and B. Preneel, “Towards measuring anonymity,” in *Proc. Workshop Privacy Enhanc. Technol. (PET)*, ser. Lecture Notes Comput. Sci. (LNCS), vol. 2482. Springer-Verlag, Apr. 2002.

[18] C. Díaz, “Anonymity and privacy in electronic services,” Ph.D. dissertation, Katholieke Univ. Leuven, Dec. 2005.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[20] C. E. Shannon, “Communication theory of secrecy systems,” Bell Syst., Tech. J., 1949.

[21] A. Wyner, “The wiretap channel,” Bell Syst., Tech. J. 54, 1975.

[22] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” *IEEE Trans. Inform. Theory*, vol. 24, pp. 339–348, May 1978.

[23] P. M. Woodward, “Theory of radar information,” in *Proc. London Symp. Inform. Theory, Ministry of Supply*, London, UK, 1950, pp. 108–113.

[24] D. V. Lindley, “On a measure of the information provided by an experiment,” *Annals Math. Stat.*, vol. 27, no. 4, pp. 986–1005, 1956.

[25] E. T. Jaynes, “On the rationale of maximum-entropy methods,” *Proc. IEEE*, vol. 70, no. 9, pp. 939–952, Sep. 1982.