Análisis en torno a las tecnologías de privacidad en redes. Anonimato en transmisión de datos

Por el Dr. Javier Areitio

Prof. Dr. Javier Areitio Bertolín – E.Mail: jareitio@eside.deusto.es
Catedrático de la Facultad de Ingeniería. ESIDE.
Director del Grupo de Investigación Redes y Sistemas.
Universidad de Deusto.

En el presente artículo se analizan diversas tecnologías de privacidad para redes. Es un hecho conocido que la interceptación de sesiones de comunicación usuales proporciona una gran cantidad de datos privados de tráfico incluso aunque se encuentre cifrada. Actualmente se observa un elevado crecimiento de aplicaciones del anonimato de las comunicaciones electrónicas en ámbitos como los negocios, nivel corporativo, personal además de su uso tradicional a nivel gubernamental. Tecnologías de seguridad clásicas en entornos de negocio como VPN (Virtual Private Network) tanto basadas en SSL (sin cliente) como con cliente basada en IPSec no proporcionan privacidad ya que son susceptibles de ataques a la identidad y a la localización. Hoy en día existen muchos negocios que esperan construirse sobre redes de anonimato. Se observa un incremento de proveedores de servicios de anonimato gratuitos y de pago para hosting de Email y de servidores físicos y virtuales, servicios de proxy inverso Internet, proveedores de pagos (tarjetas visa prepago), etc.

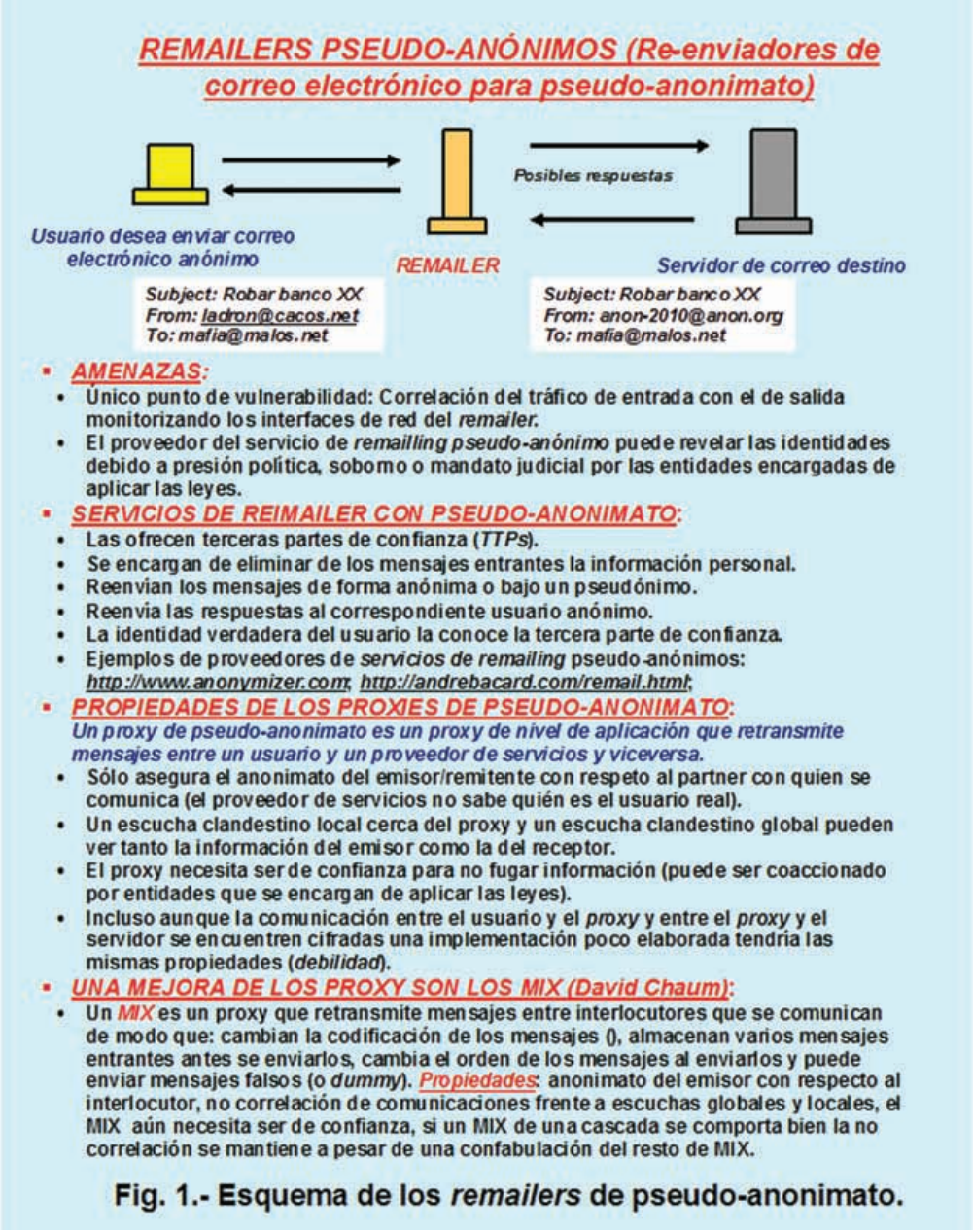


Fig. 1.- Esquema de los *remailers* de pseudo-anonimato.

Introducción

La privacidad es fundamental a la hora de poder protegerse contra todo tipo de usuarios maliciosos y actividades fraudulentas, en entornos de colaboración e interacción confiable. La privacidad es necesaria para proteger los siguientes elementos:
(i) A la fuente de la información.
(ii) Al destino de la información.
(iii) A la ruta de transmisión de diseminación de la información.
(iv) Al propio contenido de la información.

Examinemos algunos aspectos relacionadas con la privacidad:

(1) Debido a que la semántica de la información cambia con el tiempo, contexto e interpretación de las personas, algunas consideraciones útiles para privacidad son:
(a) La replicación, equivalencia y semejanza.
(b) La acumulación y generalización.
(c) La exageración y mutilación.
(d) El anonimato y Crowds (multitudes para ocultar un sujeto).
(e) Los permisos de acceso, la autenticación y las vistas.

(2) Debido a que la dirección exacta sólo la puede conocer el vecino de un nodo (correspondiente), algunas consideraciones útiles para privacidad son:
(a) La petición se reenvía hacia una dirección y posición apropiada.
(b) La granularidad de la localización puede cambiarse.
(c) Eliminar la asociación entre el contenido de la información y la identidad de la fuente de información.

(d) Alguien puede saber la fuente mientras otros pueden conocer el contenido pero no ambas.
(e) Se necesitan informes de posición a tiempo para mantener la traza-rastro de un nodo, pero esto conduce a la revelación de la trayectoria del movimiento del nodo.
(f) Un algoritmo AO2P (Ad hoc On-Demand Position-based Private routing) puede utilizar la posición de un punto de referencia abstracto en vez de la posición del destino.
(g) El anonimato como una medida de privacidad puede basarse en la probabilidad de que coincida una posición de un nodo con su identificador y el número de nodos de un área concreta que represente una posición.
(h) Utilizar proxies confiables para proteger la privacidad.

(3) Algunas personas o sitios pueden ser más confiables que otros debido a razones de evidencia, credibilidad, interacciones y recomendaciones pasadas, algunas consideraciones útiles para privacidad son:
(a) Desarrollar medidas de confiabilidad y privacidad.
(b) Ofrecer información privada en incrementos durante un período de tiempo.
(c) Comerciar la privacidad para confiabilidad.

(4) Es difícil especificar políticas que preserven la privacidad de una manera legal, precisa y correcta. Es incluso más difícil aplicar políticas de privacidad. Algunas consideraciones útiles de cara a la privacidad son:
(a) Desarrollar lenguajes para especificar políticas.
(b) Utilizar obligaciones y penalizaciones.
(c) Especificar cuando, quién y cuantas veces la información privada puede diseminarse.
(d) Utilizar la apoteosis para destruir la información privada.
(e) Unir y vincular datos con restricciones de política.

Concepto de anonimato. Su necesidad

El anonimato puede definirse como:
(i) La no vinculación entre acción e identidad. Por ejemplo, el remitente y su correo electrónico no deben estar más relacionados dentro del sistema de lo que están en un conocimiento a-priori.
(ii) La no observabilidad. Cualquier entidad de interés (mensaje, evento, acción, persona, etc.) es indistinguible de cualquier otra de interés.
(iii) La capacidad para hacer algo sin ser cogido.

El anonimato se define por acción. El anonimato trata de ocultar información como por ejemplo: la identidad del usuario, la relación entre usuarios, el hecho de que un usuario no pueda ser identificado dentro de un conjunto de sospechosos. Un concepto relacionado con el anonimato es el pseudo-anonimato que ss la capacidad de hacer algo sin ser descubierto, pero el adversario detecta repetición de acciones.

El anonimato de localización es el estado de no ser identificable dentro de un conjunto de sujetos, el conjunto de posibles sujetos se denomina conjunto de anonimato; cuando el número de sujetos es grande, el nivel de anonimato es elevado. Para proteger la privacidad de la localización del usuario (es crucial ya que los datos de posición de una persona son datos personales significativos; los datos de posición permiten invasión de la privacidad del usuario) en servicios basados en localización (LBS,

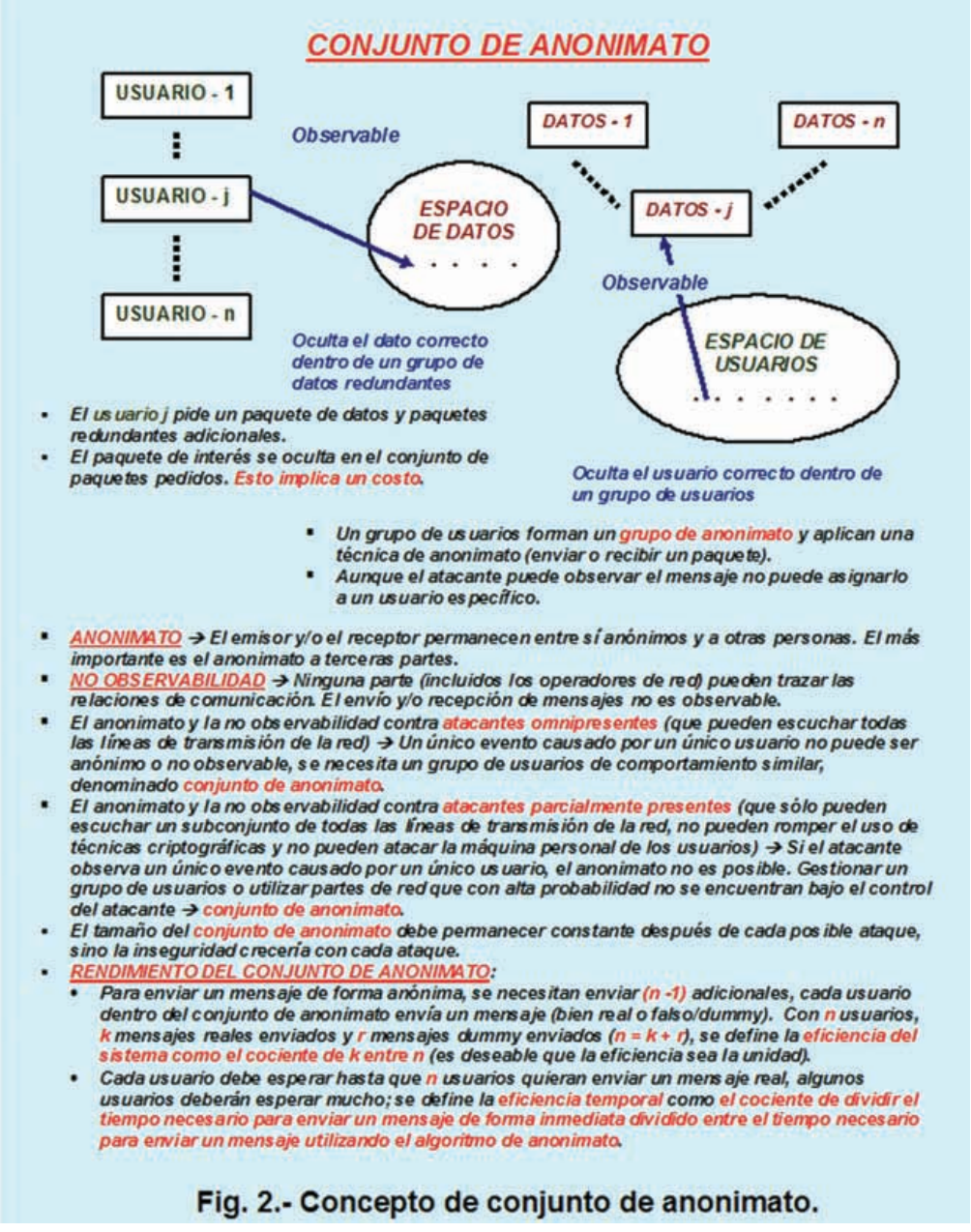


Fig. 2.- Concepto de conjunto de anonimato.

MÉTRICA DE LA PRIVACIDAD

- Para cada sujeto se puede definir un *estado de privacidad* (describe una obtención de datos) como un vector o tupla: (a_1, a_2, a_3) donde:

  a_1 representa la probabilidad de que se conozca el *identificador del peticionario*.

  a_2 denota la probabilidad de que se conozca el *data-handle* (donde se almacena un dato; se usa para cargar y sacar datos; referencia un fichero en disco, un trozo de memoria o un fichero dentro de un fichero más grande).

  a_3 representa la probabilidad de que se conozca el *contenido de los datos*.
- Para cada elemento de la tupla el valor "0" significa que no se sabe nada y "1" que se sabe todo.
- Un estado en el que la *privacidad ha sido comprometida* se representa como *(1, 1, x)* donde x pertenece al intervalo cerrado [0, 1]. Aquí se puede vincular el identificador del solicitante al dato que se esta interesado.
- Si dos sujetos conocen una parte de un secreto y sus tuplas son:

  (a_1, a_2, a_3) y (b_1, b_2, b_3) *respectivamente, entonces después de confabularse la información revelada será:* $(c_1, c_2, c_3) = (a_1, a_2, a_3) * (b_1, b_2, b_3)$ donde:

  $c_i = \text{máx}(a_i, b_i)$ *si* $a_i \neq 0$ y $b_i \neq 0$; *en caso contrario* $c_i = 0$. *Para poder evaluar la privacidad obtenida se puede utilizar el número de confabulaciones requeridas para comprometer el secreto.*
- Representación gráfica:

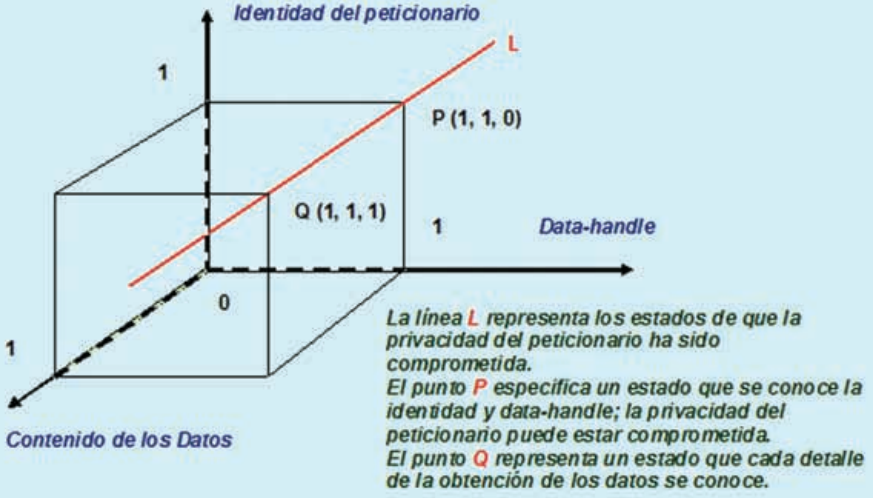


Fig. 3.- Medida de la privacidad.

Location-Based Services) se pueden utilizar técnicas de anonimato de comunicaciones que utilizan datos de posición falsos (denominados dummies) mezclados con datos de posición verdaderos. Cada usuario envía varios dummies con los datos de posición verdaderos GPS/Galileo.

El anonimato de localización se puede mejorar con dos requisitos:

(a) Ubicuidad. Los sujetos pueden estar en un área entera, los observadores deben comprobar muchas regiones de esa área para encontrar sujetos concretos. En relación a la métrica del nivel de anonimato se puede definir un indicador F (en notación porcentual) definido como el cociente de dividir el número de regiones en las que el sujeto o dato existe dividido entre el número total de regiones del área.

(b) Congestión. Si existe un gran número de sujetos en una región, es difícil distinguir un sujeto de entre muchos en la misma región. En relación a la métrica del nivel de anonimato se puede definir un indicador P definido como el número de sujetos en una región específica. Esto se puede extender a sujetos que se mueven. Se puede definir H(P) como la diferencia de P en cada región desde el instante de tiempo t al instante de tiempo siguiente (t+1). Si H(P) en cada región es bajo se mejora el nivel de anonimato de localización.

Existen diferencias notables entre los conceptos de seguridad de la información denominados anonimato y confidencialidad. La confidencialidad sólo esta relacionada con ocultar el significado de la información transmitida, se resuelve utilizando tecnología de cifrado (simétrico de clave privada, de bloque o flujo o asimétrico o de clave pública, con PKI). El cifrado ayuda a ocultar el significado de los datos transferidos pero si alguien intercepta dicho tráfico puede obtener relaciones de comunicación valiosas como quién se comunica con quién, durante cuanto tiempo y dónde. Así mismo quién esta interesado en qué contenido y qué información se da a un servicio. En algunos escenarios de aplicación, como por ejemplo elecciones, sondeos de opinión o consultas anónimas, es esencial la privacidad y el anonimato (confidencialidad ocultando los datos de tráfico). Se debe tener en cuenta que la privacidad y anonimato no restringido en todas las comunicaciones no esta permitido. Los criminales y delincuentes pueden utilizar los sistemas de comunicaciones modernos para su propio beneficio contra la ley. El gobierno esta autorizado a interceptar las comunicaciones de un sujeto criminal o delincuente. El cifrado extremo a extremo no ayuda al anonimato ya que sólo oculta el contenido de un mensaje pero la información de direccionamiento IP (L3) y MAC (L2) es visible.

El anonimato no se resuelve con cifrado y persigue entre otros los siguientes objetivos:
(i) La protección de la identidad.
(ii) La protección de la localización.
(iii) La denegación de acciones (acceder, transmitir, recibir, leer, etc.). El ámbito de protección del anonimato abarca muchos aspectos. Son cuestiones de anonimato: ¿quién es el emisor y el receptor?, ¿quién eres tu?, ¿con quién te comunicas?, ¿dónde estas localizado?, ¿dónde esta localizado el receptor/servidor con el que te comunicas?. La mayor parte de las redes que permiten el anonimato sólo protegen al emisor. Una buena red que permita el anonimato de las comunicaciones requiere protección mutua. Ninguna organización puede por sí sola estar anónima; uno sólo puede obtener confidencialidad; las redes que se encargan del anonima-

to requieren cooperación, es decir el anonimato requiere cooperación. Implantando el anonimato los adversarios no podrán saber quién se comunica con quién

El anonimato es cada vez más necesario por un número creciente de razones:
(i) El navegar por Internet deja trazas como: entradas de log en servidores Web, entradas de log en MTAs (Agentes de Transferencia de Correo Electrónico), monitorización local por parte del ISP, a menudo mandado por la Ley, monitorización del tráfico de tránsito por routers y pasarelas.
(ii) Privacidad.
(iii) Libertad de expresión.
(iv) Elecciones y sondeos electrónicos.
(v) Buscar un nuevo trabajo.
(vi) Buscar un socio corporativo.
(vii) Anonimato con vistas a recoger información.
(viii) Anonimato para la investigación de mercado.
(ix) Anonimato para dar pistas a la policía.
(x) Anonimato para exalcohólicos, exdrogatictos, exenfermos, expresidiarios, para personas incluidas en planes de protección de testigos para juicios, etc.

El anonimato presenta pros y contras; como pros, facilita la comunicación de disidentes políticos/religiosos, personas amenazadas por terroristas/mafia/delincuentes, permite comunicarse con privacidad sobre redes vehiculares VANETs, etc. Como contra inhibe los mecanismos de auditoría, monitorización y accountability que se encargan de responsabilizar a los sujetos de lo que hacen o intentan hacer, registrando quién hace qué cosa y cuando y donde en una red.

Controles técnicos de privacidad

Los controles técnicos para la privacidad (PETs, Pricacy-Enhancing Technologies) pueden clasificarse en tres categorías:

(1) Protección de las identidades de los usuarios, para ello se utiliza:
(i) Anonimato. Un usuario puede utilizar un recurso o servicio sin revelar su identidad.
(ii) Pseudónimos. Un usuario que actúa bajo un pseudónimo puede utilizar un recurso o servicio sin revelar su identidad. Los pseudónimos se pueden clasificar según su función en dos categorías: personales (pseudónimos personales públicos, pseudónimos personales no públicos y pseudónimos personales privados) y de rol (pseudónimos de negocios y pseudónimos de transacción). Los pseudónimos se pueden clasificar según su generación en cuatro categorías: pseudónimos auto-generados, pseudónimos de referencia, pseudónimos criptográficos y pseudónimos unidireccionales.
(iii) No observabilidad. Un usuario puede utilizar un recurso o servicio sin que otros puedan observar que el recurso o servicio esta siendo utilizado.
(iv) No relación/vinculación. El emisor y receptor no pueden ser identificados comunicándose entre sí.

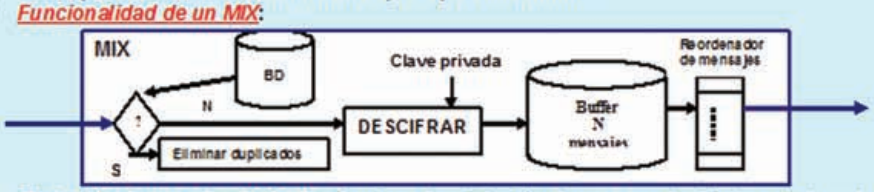
(2) Protección de las identidades de los sujetos de los datos, para ello se utiliza:

La despersonalización o anonimato de los sujetos de los datos. Existen dos tipos:
(i) Despersonalización perfecta. Los datos se hacen anónimos de modo que el sujeto de los datos ya no es identificable.
(ii) Despersonalización práctica. Se

Fig. 4.- Red de anonimato DC-Network.

TECNOLOGÍA PARA EL ANONIMATO DE LAS COMUNICACIONES: CASCADA DE MIXES DE DAVID CHAUM

- *Red de remailers* para el *anonimato de emisor y receptor*, consta de un suficiente número de Mixes independientes (por ejemplo de 20 a 60 nodos globales).
- La *efectividad o rendimiento* de un sistema donde los mensajes atraviesan N Mixes vale: $r = [1 / (3.N - 1)]$, la *eficiencia temporal* medida como el cociente entre el tiempo necesario para enviar un mensaje de forma inmediata y el tiempo necesario para enviar un mensaje utilizando el mecanismo de anonimato vale: $t = [1 / (3.N - 1 + (N - 1) / 2)]$ ya que todos los mensajes son reales (a diferencia del mecanismo de anonimato *DC-Network* donde sólo existe *un mensaje real en cada sesión* y el resto son mensajes de valor todo a ceros) y la estación Mix debe esperar hasta que lleguen N mensajes.
- Los puntos de entrada y salida, así como los nodos internos de la cascada se eligen de forma arbitraria, la comunicación entre el usuario anónimo, los Mixes y preferiblemente el destino (servidor) se encuentra cifrada. Para prevenir ingerencias por parte de las entidades que aplican las leyes que desean conocer el camino de vuelta a un usuario anónimo, los Mixes crean una cascada localizada en el mayor número posible de zonas independientes desde el punto de vista jurisdiccional.
- Cuanto mayor sea el número de usuarios y de tráfico se mejora el nivel de anonimato. Un elevado volumen de tráfico originado por parte de muchos usuarios es un pre-requisito para un nivel de anonimato fuerte. Si el tráfico a través de todo o parte de la red de Mixes falla por debajo de un nivel crítico se hace posible trazar los paquetes de usuario individuales. La inyección de paquetes falsos por parte de los usuarios o nodos Mix ayudan a mantener el volumen de tráfico mínimo necesario para un nivel suficiente de anonimato pero también supone una carga significativa en la red de transporte. Con K Mixes pueden darse un máximo de K.(K – 1) /2 conexiones.
- *Funcionalidad de un MIX*:



- *Eliminación de mensajes duplicados*. Gracias a la eliminación de una capa de cifrado en cada nodo, el paquete que entra al Mix y el paquete descifrado que sale difieren completamente y no pueden correlacionarse. Si un atacante registra todos los paquetes que entran y salen de un Mix entonces si retransmite un paquete entrante revelará el destino del correspondiente paquete saliente debido a que se reconoce su patrón de bits. Por tanto un Mix evita reenviar cualquier paquete previamente recibido. Esto se hace manteniendo una BD que guarda los hash de mensajes de todos los paquetes re-enviados
- *Descifrado*. El Mix utiliza su clave privada (utilizando criptografía asimétrica y PKI) para descifrar cada paquete recibido, de este modo se revela la dirección del siguiente destino y también se cambia el patrón de bits del paquete de salida.
- *Buffer de re-ordenación de mensajes*. Para prevenir un *análisis de timing* de paquetes entrantes/salientes un buffer de mensajes recoge al menos N mensajes que se originaron de varios usuarios antes de re-enviarlos en orden aleatorio. Durante períodos de bajo volumen de tráfico puede necesitarse un cierto tiempo hasta que el buffer de mensajes se llene lo cual puede causar una gran latencia.

Fig. 5.- Esquema de la tecnología para anonimato de las comunicaciones: *Cascada de Mixes* de David Chaum.

modifican los datos personales para que la información que concierne a circunstancias personales o materiales sólo pueda ser atribuida a un individuo identificado o identificable utilizando una desproporcionada cantidad de tiempo, dinero y trabajo.

Controles para llevar a cabo la despersonalización son: los controles de inferencia para bases de datos estadísticas y los métodos de preservación de la privacidad en minería de datos.

Una amenaza al anonimato es el riesgo de la reidentificación. Posibles tipos de datos en registros estadísticos son: Datos de identidad (como por ejemplo, nombre, dirección, número personal, etc.), datos demográficos (como sexo, edad, nacionalidad, etc.) y datos de análisis (como hábitos, enfermedades, etc.). El grado de anonimato de los datos estadísticos depende del: tamaño de la base de datos, de la entropía de los atributos de datos demográficos que pueden servir como conocimiento suplementario para un atacante. La entropía de los datos demográficos depende del: número de atributos, del número de posibles valores de cada atributo, de la distribución de frecuencia de los valores y de las dependencias entre atributos.

(3) Protección de la confidencialidad e integridad de los datos personales, para ello se utiliza:

(i) Gestión de identidades con privacidad mejorada.

(ii) Limitación del control de acceso, incluyendo modelos de privacidad formal para el control de acceso.

(iii) Políticas de privacidad empresariales.

(iv) Esteganografía y para evaluar esteganoanálisis.

(v) Herramientas específicas como P3P (Platform for Privacy Preferences).

Dimensiones y principios de la privacidad.

La privacidad es el derecho de los individuos, grupos e instituciones a determinar por ellos mismos cuando, cómo y qué extensión de la información de ellos se comunica a otros. Se pueden identificar tres dimensiones en la privacidad:

(1) Privacidad personal. Proteger las personas contra todo tipo de interferencia indebida (como búsqueda física) e información que viole su sentido moral.

(2) Privacidad territorial. Proteger un área física alrededor de una persona que no pueda ser violada sin el consentimiento de la persona. Como salvaguardas las leyes se refieren a órdenes de registro para poder entrar en la propiedad privada.

(3) Privacidad de la información. Trata de la recogida, recopilación y diseminación selectiva de la información.

Los principios de la privacidad básicos son:

(i) Basados en Leyes y equidad.

(ii) Necesidad de recoger y procesar los datos.

(iii) Especificación del propósito y vincular el dicho propósito. No existen datos "no-sensibles".

(iv) Transparencia. Derecho del sujeto de los datos a corregir la información, borrar o bloquear datos almacenados incorrectamente/ilegalmente.

(v) Supervisión (significa control por parte de una autoridad de protección de datos independiente) y sanciones en su caso (a empresas que incumplan las leyes de la Agencia de Protección de Datos a nivel internacional, nacional y autonómico).

La protección de la privacidad puede ser emprendida por:

(i) Leyes de protección de datos y privacidad, promovidas por el gobierno.

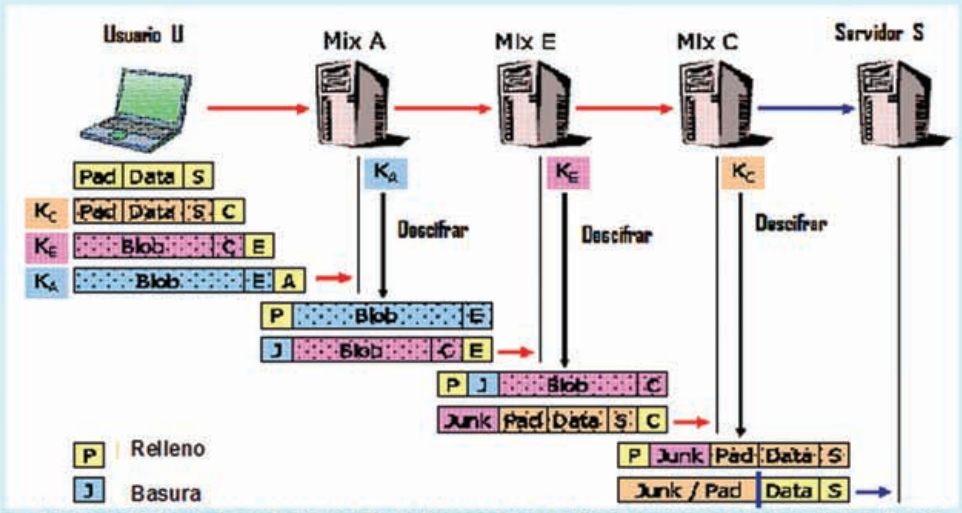
(ii) Autorregulación para prácticas

equitativas de información mediante códigos de conducta promovidas por negocios.
(iii) Tecnologías que mejoren la privacidad (o PETs, Privacy-Enhancing Technologies) adoptadas por los individuos.
(iv) Educación en privacidad de los consumidores y profesionales de las Tecnologías de la Información y las Comunicaciones.

Clasificación e identificación de redes para el anonimato.

Atendiendo al criterio de la latencia que introducen las redes que proporcionan anonimato, éstas se pueden clasificar en dos grupos:
(1) Con comunicación de baja latencia. Se basan en una cascada de mixes. Pequeños buffers o la no existencia de buffers de reordenación da lugar a una pequeña latencia. Son adecuadas para acciones interactivas como navegación Web y Chat. Algunas implementaciones son:
(i) Onion Routing. Utilizan el mecanismo del onion router. La mayor red de anonimato conocida del tipo Onion Router de la Segunda Generación es TOR (desarrollada por el DoD, Departamento de Defensa USA y soportada por EFF (Electronic Frontier Foundation) se trata de un ejemplo de Onion Router de la Segunda Generación). Véase http://tor.eff.org/.
(ii) JAP (Java Anon Proxy). Es un proxy Web basado en Java, originalmente desarrollado por TU Dresden. Es fácil de utilizar, proporciona navegación Web de baja latencia. La cascada fija de mixes introduce el inconveniente de la existencia de únicos puntos de entrada y de salida a monitorizar. Incorpora una función de path tracking que puede activarse bajo petición judicial, pero todos los operadores de los mix deben estar de acuerdo. Véase http://anon.inf.tu-dresden.de/.
(iii) I2P (Java). Mixnet gratuita, totalmente distribuida (P2P). Presenta una latencia variable a través del API I2P.
(iv) FreeNET (Java). Permiten un almacenamiento P2P para acceso y publicación de contenido anónimo. Es demasiado lenta.
(v) Anonymizer.net. Es un proxy de anonimato. Ver URL: http://www.anonymizer.com. La principal ventaja de las redes de anonimato de baja latencia es que la baja latencia permite servicios interactivos como navegación Web, mensajería instantánea (IM) o incluso conexiones SSH (Secure SHell).

Los principales inconvenientes de estas redes son:
(a) La fuerza del anonimato cae precariamente durante los períodos de volumen de bajo tráfico.
(b) Los esquemas de baja latencia son vulnerables a un observador global que monitoriza todos los nodos debido a que el timing de los paquetes que atraviesan la red puede correlacionarse.
(2) Con comunicación de elevada latencia. Son adecuadas para acciones de almacenamiento y re-envío típicas del correo electrónico. Se basan en una conexión en cascada de Mixes, utilizan buffers grandes de reordenación de mensajes con alta latencia. Algunas implementaciones son:
(i)CypherPunk (Remailer de anonimato tipo I): Utilizan PGP para cifrar. Ver URL: http://www.csua.berkeley.edu/cypherpunks/Home.html. Tecnología antigua y red vieja.
(ii) Mixmaster (Remailer de anonimato tipo II). Requieren un cliente especial. Ver URL: http://www.mixmaster.sourceforge.net. Existen cerca de treinta y cinco servidores, red muy estable. Un Frontend Mixmaster basada en Web es http://anonymouse.org.



- *No trazabilidad del camino a través de la cascada de Mixes de una red de remailer de David Chaum.* Si cada Mix sólo conoce su predecesor y sucesor inmediato de la *cascada* entonces no es posible trazar el camino completo desde el usuario anónimo U al servidor S destino. La no trazabilidad se mantiene incluso aunque varios de los Mixes colaboran (se confabulan) siempre que al menos un Mix sea independiente en la cascada.
- *No trazabilidad gracias al cifrado por capas utilizando criptografía asimétrica de clave pública.* Cada Mix publica su clave pública K_x. La confianza se establece utilizando certificados digitales o listando la clave pública en un directorio de confianza. Para enviar datos de forma anónima al servidor destino S, el usuario U elige un camino arbitrario a través de la *red remailer* formada por los Mixes A, B, C, D E y F, seleccionando por ejemplo la cascada *U – A – E – C – S*. En el funcionamiento hacia atrás desde el punto de salida C al punto de entrada A de la cascada, el usuario U cifra de forma recursiva el mensaje de datos con la clave pública K_x del Mix receptor x y añade la dirección de x por delante del mensaje cifrado. Debido a que el tamaño del bloque de datos puede variar así como el número total de Mixes de la cascada seleccionada, el mensaje original al servidor S se rellena con bytes aleatorios de modo que *el paquete cifrado* enviado por el usuario U al punto de entrada tiene la *misma longitud que todos los otros paquetes* transferidos por la red de Mixes. En cada nodo, la dirección del receptor corriente se elimina de la delantera del paquete y el mismo número de bytes e añaden en forma de relleno aleatorio al final del paquete. Este esquema garantiza que el tamaño del paquete permanezca el mismo y previene que los nodos internos aprendan su posición dentro de la cascada. El Mix entonces utiliza su clave privada para descifrar todo el paquete revelando la dirección del siguiente salto en la cascada. El *punto de salida* descarta todos los rellenos y envía el mensaje original al servidor destino S.

Fig. 6.- No trazabilidad utilizando criptografía de clave pública en una *cascada de Mixes* de una *red de remailer*.

TECNOLOGÍA DE ANONIMATO *TOR* (SEGUNDA GENERACIÓN DE *ONION ROUTER*)

- *Tor* fue diseñada y desarrollada como parte del programa *Onion Router* del Laboratorio de Investigación Naval USA (*Dpto. de Defensa USA*). Actualmente el *desarrollo Tor* esta soportado por *EFF* (*Electronic Frontier Foundation*).
- *Características principales:*
  - Consigue el anonimato bi-direccional de *streams TCP* sobre Internet.
  - Secreto de reenvio perfecto gracias a los intercambios de claves criptográficas *D-H (Diffie-Hellman)*.
  - Servidores de directorio confiables proporcionan la información corriente sobre los Onion Routers.
  - Políticas de salida definen los computadores y puertos al que se conectarán un nodo de salida.
  - A través de una topología de circuito *leaky-pipe* y tráfico de señalización dentro de banda dinámico se puede dejar la cascada en cualquier nodo intermedio.
  - Estableciendo puntos de encuentro se posibilita servicios ocultos.
  - Normalmente de 450 a 750 *Onion Routers* se activan a escala global.
  - Véase la URL: *http://tor.eff.org/*.
- *Componentes*:
  - *Onions Routers* (*ORs*). Se registra como un nodo oficial en la red de anonimato *Tor*. Su clave de identidad pública se publica en servidores de directorio confiables. El certificado digital TLS de los OR se firma digitalmente con la clave de identidad publicada, es decir no se utiliza una Autoridad de Certificación centralizada de raíz. Dependiendo de la política de salida publicada, un OR actúa bien sólo como nodo intermedio o bien también puede utilizarse como un nodo de salida.
  - *Onions Proxies* (*OPs*). Es un una interfaz de usuario a la red de anonimato Tor. No se autentica a la red.
  - *Conexiones cifradas TLS* entre *ORs/OPs* utilizando certificados digitales y claves efímeras D-H. Las conexiones OR-OR realizan autenticación mutua, las conexiones OP-OR sólo realizan autenticación del lado OR.
  - *Sitios servidores.* Son los destinos de las comunicaciones de los OPs de los Usuarios.
- *Células, circuitos y streams (corrientes TCP) en una red Tor*:

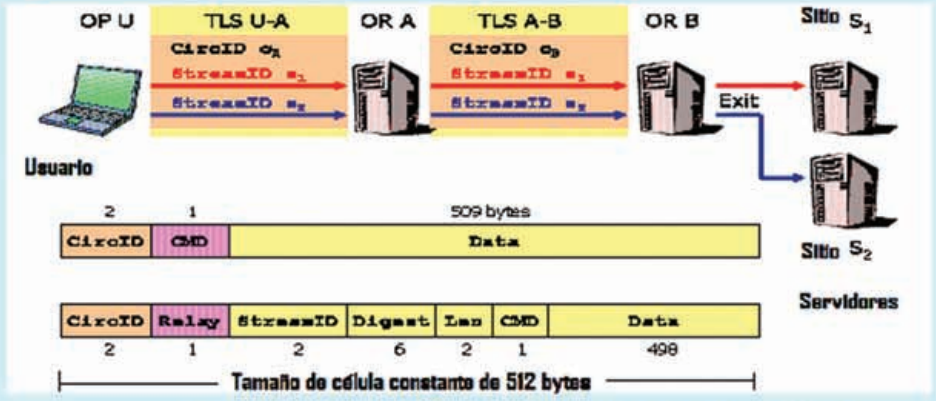


Fig. 7.- Tecnología de anonimato *Tor* (segunda generación de *Onion Router*).

(iii) Mixminion (Remailer de anonimato tipo III). Ver URL: http://www.mixminion.net. Requieren un cliente especial. Son redes experimentales. El proyecto Free Haven (almacenamiento de datos distribuido anónimo) ver URL: http://www.freehaven.net/anonbib.

(iv) Mym servers. Vieja escuela. Son remailers como por ejemplo nym.alias.net y anon.penet.fi. Estas redes no son sistemas de correo electrónico, por tanto necesitan de verdaderos sistemas de correo electrónico que son susceptibles de monitorizar trazas.

Como principal ventaja de las redes de anonimato de alta latencia: Se obtiene un alto grado de anonimato debido a que se impide las correlaciones temporales mediante el uso de grandes buffers de reordenación y un elevado número de mixes en la cascada. Como principales inconvenientes:

(a) La elevada latencia debido al uso de buffers de reordenación no permiten servicios interactivos casi en tiempo real como navegación Web o mensajería instantánea (IM).

(b) Algunos mixes de la cascada pueden no estar en línea en el instante en que llega un mensaje retardado, de este modo a menudo se pueden perder mensajes. Estadísticas publicadas por herramientas como echolot ayudan a seleccionar una cascada fiable.

Algunos ejemplos de redes de anonimato son: Mnet, Winny, GNUnet, Mute, Marabunta, Morphmix, Tarzan, AntsP2P, AnoNET, Crowds, WASTE, Invisible IRC, Entropy, Rodi, infraestructuras para sistemas internet con resiliencia, etc.

Identificación de amenazas a la privacidad.

Se pueden identificar entre otras las siguientes amenazas:

(1) Amenazas a la privacidad en el nivel de aplicación. Son amenazas a la recogida/transmisión de grandes cantidades de datos personales. Aquí se incluyen proyectos para nuevas aplicaciones sobre autopistas de la información como: redes de centros sanitarios, redes de las Administraciones Públicas, redes de investigación, comercio electrónico, teletrabajo, enseñanza a distancia, uso privado. Ejemplo de desarrollos en esta área es la creación de una infraestructura de información para un mejor cuidado de la salud donde diversas redes de las entidades de salud Europeas (Seguridad Social Española, etc.) intercambien información como ficheros de casos de pacientes, se establezcan sistemas de telediagnóstico y tratamiento clínico, etc.

(2) Amenazas a la privacidad a nivel de comunicación. Aquí se incluyen las amenazas al anonimato de emisor/re-enviador/receptor, las amenazas al anonimato del proveedor de servicios, las amenazas a la privacidad de la comunicación, por ejemplo monitorizando/registrando los datos transaccionales, extrayendo perfiles de usuario y almacenándolos a largo plazo.

(3) Amenazas a la privacidad a nivel de sistema. Por ejemplo amenazas a nivel de acceso al sistema.

(4) Amenazas a la privacidad de registros de auditoria. Como por ejemplo robo de identidad, este es el delito más grave contra la privacidad. Amenazas a la privacidad desde otra perspectiva son la acumulación y minería de datos, la seguridad deficiente del sistema, amenazas al Gobierno (el Gobierno tiene muchísimos datos privados de la mayoría de las personas: impuestos, etc.), Internet como amenaza a la privacidad (correo electrónico no cifrado, navegación Web, ataques, etc.), derechos corporativos y negocios privados (ciertas compañías pueden recoger datos que el Gobierno no se le esta permitido), privacidad se vende con muchas trampas (gratis no significa gratis, por ejemplo aceptar tarjetas de "comprador frecuente" reduce la privacidad del propietario).

Aplicaciones del anonimato

Las tecnologías de anonimato de las comunicaciones presentan un creciente número de usos y en cada uno de ellos se pueden identificar ciertos abusos:

(1) Utilización legítima a nivel personal. Los ciudadanos pueden necesitar mantener en anonimato su actitud sexual, sus creencias/visión religiosa, su genoma, sus tendencias sobre política, etc. Se deben evitar ciertos abusos como el traking (seguimiento de personas) y el profiling (generar perfiles sobre las tendencias, religión, preferencias-gustos, etc. de las personas) realizado por ISPs (Proveedores de Servicios Internet), diferentes tipos de empresas, gobiernos e incluso por buscadores Internet como Google, Yahoo, etc. Por ejemplo un usuario o un spammer pueden abrir una cuenta falsa en Hotmail y enviar correos electrónicos por Webmail sin que el receptor sepa la identidad real del remitente.

(2) Utilización legítima a nivel de empresa-corporación. Las empresas necesitan mantener en anonimato sus actividades de inteligencia de negocios, deben de tratar de detener el análisis de la competencia (I+D+i, adquisiciones, etc.), deliberaciones legales, comunicaciones de países no democráticos o de lugares en guerra, comunicaciones de periodistas. Se deben evitar ciertos abusos y prevenir la discriminación de información y económica.

(3) Utilización legítima a nivel de gobierno. Los gobiernos necesitan el anonimato en las comunicaciones diplomáticas (por ejemplo ocultando donde se encuentra el presidente o el embajador). Así mismo es necesario el anonimato en las peticiones anónimas de ciudadanos a la policía, en la investigación criminal (por ejemplo el CNI, la Policía Nacional-Guardia Civil o el FBI esta husmeando en el sitio Web de una persona física o jurídica) y en cualquier asunto que pueda ser de interés público.

MÉTRICA PARA MEDIR EL GRADO DE PRIVACIDAD BASADA EN LA ENTROPÍA

- La entropía permite medir la aleatoriedad o incertidumbre en los datos privados.
- Cuando un atacante gana más información de unos datos privados la entropía disminuye.
- La métrica consiste en comparar el valor de la entropía actual con su valor máximo, la diferencia muestra la información pedida/robada.
- *CASO PRÁCTICO*:

  Dado un número de teléfono privado de diez cifras: (a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10}), se sabe que cada dígito se almacena como un valor de un atributo separado. Suponer que intervalo de valores para cada atributo es del cero al nueve y que todos los atributos son igualmente importantes (es decir su peso en $w_j = 1$). Se pide determinar la pérdida/robo de información cuando un atacante averigua tres dígitos del número de teléfono privado.
  - La máxima entropía es cuando el atacante no posee información alguna acerca del valor de cada atributo o dígito del número de teléfono. El conjunto A de atributos tiene diez cifras.
  - El atacante asigna una distribución de probabilidad uniforme a los valores de cada atributo, por ejemplo $a_j = i$ con probabilidad de 0,10 para cada i en el intervalo de 0 a 9.
  - La máxima entropía vale: $H'(A) = \sum_{j=0}^{9}(w_j \sum_{i=1}^{10}(-0{,}1.\log_2(0{,}1))) = 33{,}3$
  - Supongamos que en el instante de tiempo t el atacante *averigua los tres dígitos* de más a la izquierda del número de teléfono privado.
  - La entropía en el instante de tiempo t vale: $H(A, t) = 0 + \sum_{j=4}^{10}(w_j \sum_{i=0}^{9}(-0{,}1.\log_2(0{,}1))) = 23{,}3$
  - Los atributos (a_1 a_2 a_3) contribuyen con el valor cero al valor de la entropía debido a que el atacante sabe sus valores correctos, por tanto
  - La *pérdida/robo de información* en el instante de tiempo t (cuando un subconjunto de valores de todos los atributos del conjunto A ha sido revelado) vale: $D(A, t) = H'(A) - H(A, t) = 10$.

CONCEPTO DE DATA HAVENS ANÓNIMOS

- Existe una *necesidad* de *Data Havens Anónimos* (zonas de almacenamiento anónimo de información) por un *conjunto creciente de razones*: las patentes de software pueden hacer peligrar el software *open-source*, la necesidad de descargas de audio/video, la existencia de leyes de copyright fuertes que prohiben la libre distribución de libros y otro tipo de materiales. Los *recursos necesarios* son: servicios de directorio y de búsqueda distribuidos, repositorios da datos anónimos y resistencia a ataques *DoS/DDoS* (*Denial of Service*). Como proyecto más relevante en *Data Havens Anónimos*: *freenetproject.org*.

Fig. 8.- Métrica para medir el grado de privacidad basada en la entropía. Concepto de *Data Havens Anónimos*.

Las tecnologías que permiten el anonimato pueden ser utilizadas para el bien de forma legítima pero también pueden utilizarse para el mal por parte de las mafias, terroristas que realizan ciber-ataques, pedófilos para pornografía infantil, delincuentes en general, entidades que fomenten y conciencien sobre bulimia, anorexia, drogas, secuestros, violaciones, terrorismo, etc.

Consideraciones finales.

Nuestro grupo de investigación lleva trabajado más de veinte años en el área de las tecnologías de privacidad para redes dentro de las cuales existen muchos me-

canismos como los diversos tipos de anonimato, las diferentes técnicas criptográficas/de criptoanálisis y esteganográficas/ de esteganoanálisis, DWM, criptografía basada en umbrales, los mecanismos basados en pseudónimos, alias nicks, etc.

Este artículo se enmarca en las actividades desarrolladas dentro del proyecto LEFIS-APTICE (financiado por Socrates. European Commission).

Bibliografia.

- Areitio, J. "Seguridad de la Información: Redes, Informática y Sistemas de Información". Cengage Learning-Paraninfo. 2009.

- Areitio, J. "Análisis en torno a los esquemas de compromiso digital y su aplicación en seguridad de red". Revista Española de Electrónica. Nº 644/645. Julio-Agosto 2008.

- Areitio, J. "Análisis en torno a la auditoria de seguridad en tecnologías de la información y las comunicaciones". Revista Española de Electrónica. Nº 625. Diciembre 2006.

- Areitio, J. "Test de penetración y gestión de vulnerabilidades, estrategias clave para evaluar la seguridad de red". Revista Española de Electrónica. Nº 653. Abril 2009.

- Senior, A. "Protecting Privacy in Video Surveillance". Springer. 2009.

- Gutwirth, S., Poullet, Y., De Hert, P., Terwangne, C. and Nouwt, S. "Reinventing Data Protection". Springer. 2009.

- Howard, R. "Cyber Fraud". Auerbach Publishers, Inc. 2009.

- Bettini, C., Jajodia, S., Samarati, P. and Wang, X.S. "Privacy in Location-Based Applications: Research Issues and Emerging Trends". Springer. 2009.

- Flegel, U. "Privacy Respecting Intrusion Detection". Springer. 2007.

- Vaidya, J., Clifton, C. and Zhu, M. "Privacy Preserving Data Mining". Springer. 2005.

- Dumas, M.B. and Schwartz, M. "Principles of Computer Networks and Communications". NJ. Prentice-Hall. 2008.

- Fischer-Hübner, S. "IT-Security and Privacy: Design and Use of Privacy-Enhancing Security Mechanisms". Springer. 2001.

- Comer, D.E. "Computer Networks and Internets". NJ. Prentice-Hall. 2008.

- Gritzalis, D., de Capitani di Vimercati, S., Samarati, P. and Katsikas, S. "Security and Privacy in the Age of Uncertainty". Springer. 2003.

- Solove, D.J. "The Digital Person: Technology and Privacy in the Information Age". New York University Press. 2004.