Privadesa

Alexandre Viejo Galicia
Jordi Castellà-Roca

PID_00183943

UOC
Universitat Oberta de Catalunya

Índex

Introducció

La tecnologia és una peça clau en l'engranatge, que fa avançar la societat en conjunt. No obstant això, la tecnologia aplicada a la informació aconsegueix potenciar els individus en particular.

Els qui en el passat passaven una quantitat ingent d'hores adquirint coneixement de llibres i altres documents, amb l'inici de l'era d'Internet van conèixer un món nou en què es multiplicava la quantitat d'informació disponible i la facilitat d'accedir-hi arribava a cotes difícils d'imaginar abans. En aquest escenari nou, els ordinadors d'arreu del món estan connectats entre si. Ordinadors de persones anònimes, de petites empreses, grans corporacions i governs es comuniquen els uns amb els altres i comparteixen els coneixements. Internet permet a individus que no es coneixerien mai en persona compartir el que veuen i el que saben.

Internet s'ha convertit en el gran aparador per a qualsevol que es vulgui donar a conèixer al món. És el lloc idoni per a les empreses a la recerca de publicitat directa, per als exhibicionistes i també per als *voyeurs* que entenen la Xarxa com la seva televisió a la carta per a passar les estones d'oci.

La digitalització massiva de documents antics, la utilització d'aquest format per a generar els documents actuals i la interconnexió dels ordinadors ha creat informació inalterable i disponible per a tothom. Històricament, la privadesa de les persones s'ha basat en el fet que aquestes persones controlaven la seva pròpia informació. No obstant això, és pràcticament impossible mantenir aquest control en un món dominat pel format digital i els ordinadors connectats. Una vegada es fa pública una certa dada, pot ser replicada i recol·locada en qualsevol altre lloc de la Xarxa, sense que se n'alteri la qualitat o variï amb el temps.

Aquest escenari ha propiciat que els ordinadors es converteixin en una eina indispensable per a interactuar amb el món. Fòrums, blogs, xarxes socials, etc.: totes aquestes aplicacions permeten als individus sadollar la seva curiositat encara que aquest camí també els porta a compartir el seu propi coneixement.

Finalment, els qui es consideraven solament consumidors d'aquest gran teatre de varietats han descobert que ells també són actors principals en una gran obra global. Han comprovat que el seu rastre a Internet augmenta a mesura que interactuen amb la Xarxa i que les seves pròpies dades personals, interessos o opinions poden ser adquirits fàcilment, tan fàcilment com ells troben informació referent a altres individus.

Tanmateix, no solament els individus anònims poden accedir a totes aquestes dades; les empreses i els governs han après a explotar tota aquesta quantitat d'informació personal disponible i proporcionada per les persones mateixes. Així, han desenvolupat noves maneres d'obtenir beneficis i influir en les masses. En molts casos, la privadesa ja és cosa del passat. Recuperar-la, és una de les tasques del present.

Objectius

En els materials didàctics associats a aquest mòdul, l'estudiant hi trobarà els continguts necessaris per a assolir els objectius següents:

1. Entendre què és l'empremta digital d'un usuari, què conté i com es crea.

2. Conèixer els mecanismes utilitzats habitualment a Internet per a identificar i rastrejar un usuari sense que se n'adoni.

3. Conèixer els atributs i les dades personals que formen un perfil d'usuari i entendre els mètodes utilitzats per a generar-los.

4. Entendre com poden utilitzar els perfils d'usuari les empreses i les organitzacions per a augmentar els seus beneficis.

5. Conèixer la definició de *privadesa* en el marc de les tecnologies de la informació i entendre les implicacions de les polítiques de privadesa aplicades per les empreses d'Internet que tracten dades personals.

6. Tenir nocions bàsiques sobre les tècniques disponibles per a preservar la privadesa dels usuaris i conèixer les virtuts i els inconvenients que tenen i els diferents àmbits d'aplicació.

1. L'empremta digital

Actualment, la majoria dels habitants de les societats desenvolupades interactuen de manera habitual amb el món digital (per exemple, televisió, telèfons mòbils, Internet, sensors o RFID). L'empremta digital d'una certa persona està formada pel rastre que deixen aquestes interaccions.

L'empremta digital és essencial per a proporcionar personalització, màrqueting dirigit, reputació digital i diversos serveis basats en el medi social.

Més enllà de proporcionar serveis o anuncis personalitzats, no ens ha de passar per alt la importància de l'empremta digital respecte a la reputació dels individus. La utilització de les empremtes digitals en els processos de selecció de personal és una realitat. En aquest sentit, Coutu i altres (2007) expliquen en el seu estudi que, abans, el currículum d'una persona era completament controlat per aquesta persona i es basava únicament en el que aquesta persona comunicava a l'empresa que hi estava interessada. En canvi, actualment, el currículum d'un individu està format pels primers deu ítems que surten al Google.

La utilització de l'empremta digital no es limita als departaments de recursos humans de les empreses. Una sèrie d'enquestes i anàlisis fetes el 2007 per Pew Internet suggereixen que el 47% de les persones han buscat informació sobre elles mateixes a Internet. Aquest tipus de cerca es coneix com a *cerca de vanitat* o *d'ego*. Més enllà de la cerca d'informació pròpia, el mateix estudi indica que més de la meitat de tots els usuaris adults d'Internet han utilitzat un motor de cerca per a seguir les petjades d'una altra persona. Un 11% d'aquests usuaris buscaven informació sobre algú a qui estaven pensant contractar. No obstant això, crida l'atenció que un 19% dels usuaris buscava informació sobre companys de feina, col·legues o competidors. Respecte a la informació adquirida, un 72% se centrava en les dades de contacte, un 37% en assoliments professionals i interessos, un 33% en perfils de xarxes socials o professionals, un 31% en fotos i, finalment, un 28% buscava també informació respecte a antecedents personals.

Lectura recomanada

Per a consultar les enquestes i les anàlisis fetes el 2007 per Pew Internet, llegiu l'article següent:

M. Madden i altres (2007). "Digital footprints: Online identity management and search in the age of transparency".

Altres resultats interessants de l'enquesta de Pew Internet indiquen que el 10% dels usuaris d'Internet tenen una feina que els obliga a promoure la seva reputació digital o fer-ne publicitat. Addicionalment, un 20% dels adults nordamericans diuen que les seves empreses tenen polítiques especials per a pro-

moure que els treballadors es facin publicitat correctament a Internet. Finalment, s'ha de destacar que les persones conclouen que un 90% de la informació que localitzen sobre elles mateixes és precisa.

Davant aquestes dades, la importància de l'empremta digital en la societat actual no admet dubtes. A continuació, les preguntes que s'han de resoldre són les següents: què conté aquesta empremta?, com es crea?

1.1. El contingut de l'empremta digital

Habitualment, es considera que l'empremta digital d'una persona la formen les seves visites a certs llocs web, les seves cerques, els seus perfils en xarxes socials o en llocs similars, correus, blogs, apunts o *posts* en fòrums, etc. Tanmateix, l'empremta digital va molt més allà. En aquest sentit, el Discovery Channel ofereix una aplicació que il·lustra com una persona deixa un rastre digital en moltes de les seves accions quotidianes.

Com a exemples d'aquesta situació destaquem les interaccions següents i el rastre que deixen:

- **Llegir notícies per Internet**: el lloc web de notícies pot identificar l'usuari (per exemple, amb un identificador d'usuari i una contrasenya, galetes o *cookies*, o una adreça IP) i extreure'n certa informació, com els seus horaris, les notícies que l'interessen, els anuncis que selecciona, etc. Aquestes dades ajuden a proporcionar un servei personalitzat a l'usuari però també permeten conèixer els seus interessos i les seves activitats, cosa que pot ser un problema per a aquest usuari.

- **Anar amb cotxe a la feina**: els cotxes més moderns són capaços d'emmagatzemar dades relacionades amb la velocitat, els frens, la utilització de cinturons de seguretat, etc. Els cotxes equipats amb GPS són capaços d'emmagatzemar les rutes seguides. Aquesta informació és útil en cas d'accident, però també pot ser utilitzada en contra seva (per exemple, les asseguradores poden utilitzar aquest coneixement per a rebutjar certs conductors).

Venda de dades de conducció

Segons va recollir el diari alemany *AD*, el fabricant del GPS TomTom va vendre les dades de conducció dels seus usuaris a la policia dels Països Baixos.

- **Pagar amb targeta de crèdit**: els bancs guarden registres complets dels moviments fets amb les targetes que emeten. Les targetes ofereixen comoditat als compradors, però les dades de transaccions poden ser utilitzades per tercers per a saber els hàbits de compra de les persones. Aquestes dades també poden ser robades en atacs informàtics a les bases de dades dels bancs.

- **Pagar un peatge**: fins i tot pagant amb diners en metàl·lic (cosa que es considera un sistema de pagament anònim), les càmeres que hi ha als peatges guarden les imatges dels vehicles i les matrícules d'aquests vehicles. Aquests sistemes eviten que conductors deshonestos saltin els peat-

ges, però també permeten localitzar els conductors honestos i analitzar fins a cert punt els seus hàbits de conducció i les seves rutes.

- **Aparcar en un pàrquing públic de pagament**: la utilització de càmeres de seguretat per a garantir el pagament del servei implica una situació similar a la del pagament del peatge.

- **Utilitzar una etiqueta (*tag*) RFID per a identificar-se a la feina**: la utilització d'aquests dispositius evita que persones no autoritzades entrin en entorns de treball restringits. No obstant això, les etiquetes RFID poden contenir una gran quantitat d'informació personal del propietari. Aquesta informació la pot robar un atacant que sigui en el radi d'acció de l'etiqueta. Addicionalment, aquests dispositius permeten a l'empresa saber els hàbits del treballador en el lloc de treball.

- **Consultar el correu electrònic**: molts sistemes de correu utilitzats habitualment (per exemple, Gmail/Googlemail o Hotmail) emmagatzemen i analitzen els missatges enviats i rebuts. A partir d'aquesta informació les empreses proporcionen anuncis personalitzats a l'usuari. No obstant això, indirectament també poden tenir accés a una gran quantitat d'informació personal dels usuaris, i comprometre així la confidencialitat dels seus missatges.

- **Accedir a una xarxa sense fil d'un lloc públic**: la informació transmesa durant la utilització d'Internet amb aquest sistema pot ser emmagatzemada pel proveïdor de serveis. Addicionalment, aquests sistemes són més susceptibles de ser atacats per pirates o *hackers* que podrien accedir a aquesta informació.

- **Trucar per telèfon**: les companyies telefòniques emmagatzemen l'hora i el número de cada trucada. L'agència de seguretat nacional dels Estats Units (NSA) pot monitorar les trucades internacionals sense informar els usuaris. Addicionalment, certs pirates poden accedir a la base de dades de la companyia telefònica i accedir a tota aquesta informació.

- **Utilitzar una xarxa social**: la informació publicada en aquests llocs web és emmagatzemada i analitzada per les empreses que gestionen aquests sistemes. Aquestes empreses utilitzen les dades adquirides amb finalitats comercials. La informació que els usuaris publiquen en aquestes xarxes és generalment molt personal i inclou números de telèfon, adreces, estat civil, opinions sobre productes, activitats que han dut a terme, viatges, etc.

- **Utilitzar un sistema de missatgeria instantània**: les empreses que proporcionen aquests serveis emmagatzemen i analitzen les converses que s'hi tenen. Generalment, les dades adquirides són utilitzades amb finalitats comercials.

Escoltes telefòniques

El març del 2005 el primer ministre de Grècia va confirmar que el seu telèfon havia estat punxat, igual que el de l'alcalde d'Atenes i cent dignataris més d'alt rang. Les escoltes també es van fer a treballadors de l'Ambaixada americana. Vassilis Prevelakis i Diomidis Spinellis van pubicar una detallada descripció del cas a la revista *IEEE Spectrum* (vol. 44, núm. 7) el juliol del 2007 amb el títol "The Athens Affair".

- **Enviar una consulta a un cercador d'Internet**: el motor de cerca pot identificar l'usuari (per exemple, mitjançant usuari i contrasenya, galetes o adreça IP) i extreure els seus interessos a partir de les consultes que fa l'usuari. La informació que se n'extreu generalment s'utilitza amb finalitats comercials.

Configuració dels navegadors

La configuració dels navegadors es pot utilitzar per a identificar usuaris de manera gairebé unívoca (Eckersley, 2010).

Com s'observa, la majoria d'interaccions que fa una persona cada dia deixa un cert rastre que forma la seva empremta digital. La grandària d'aquesta empremta varia segons el nombre i la importància de les seves interaccions; no obstant això, es pot inferir la facilitat amb què es pot extreure informació personal dels individus. En aquest mòdul ens centrarem en els problemes de privadesa associats a les interaccions que es produeixen a Internet.

1.2. Com es crea l'empremta digital a Internet

L'empremta digital d'una persona es pot construir de manera activa o passiva. En qualsevol cas, s'ha de destacar que són les accions de l'individu el que deixa en descobert les seves traces. La diferència entre construcció activa i passiva es basa en si aquests rastres es deixen de manera conscient (de manera activa) o inconscient (passiva).

La **creació activa d'una empremta digital** es fonamenta en accions que l'usuari fa deliberadament. Des de l'arribada del Web 2.0, la presència a la Xarxa d'aplicacions socials en què els usuaris generen un perfil amb fotos, opinions i tota mena de dades personals (per exemple, feina, situació civil i número de telèfon) ha crescut de manera molt significativa. En qualsevol cas, l'empremta digital creada de manera activa és responsabilitat de l'usuari mateix, que pot aprendre a controlar-la en benefici propi.

Lectura recomanada

L'article següent analitza la utilització dels perfils públics que fan les empreses per a seleccionar els treballadors:

J. Terry (2008, 7 de febrer). "Leaving a digital footprint: Online activities follow students to job interviews, professional world". *The State News*.

La **creació passiva d'una empremta digital** es fonamenta en elements pràcticament invisibles per a l'usuari com el *web caching* i les galetes. Aquests elements són tan transparents per a l'usuari que fins i tot pot ser que no sàpiga que existeixen i, per tant, es poden considerar més perillosos des del punt de vista de la privadesa.

Web 2.0

El terme *Web 2.0* està habitualment associat amb aplicacions web que faciliten compartir informació, la interoperabilitat i la col·laboració en el Web.

Alguns exemples del Web 2.0 són les comunitats web, els serveis web, les xarxes socials, els wikis i els blogs.

El *web caching* es basa en la utilització de servidors cau. Aquest tipus de memòria cau emmagatzema tota mena d'informació accessible mitjançant el navegador o *browser*. La motivació que tenen és reduir l'amplada de banda consumida, la càrrega dels servidors i el retard en la baixada. Un servidor cau emmagatzema còpies dels documents que hi passen, de manera que les peticions subsegüents poden ser respostes pel servidor cau mateix. El servidor web es pot situar en diversos punts de l'arquitectura de comunicacions: es pot localitzar en el navegador mateix de l'usuari (anomenat *browser cache*) o en un servidor intermediari o *proxy* a càrrec del mateix servidor (o servidors) web al

qual intenta accedir l'usuari. En aquest últim cas parlem d'un *proxy cache* capaç d'emmagatzemar el trànsit web de l'usuari i posar a la disposició del propietari del servidor una gran quantitat d'informació de l'individu en qüestió.

Les galetes guarden informació dels usuaris en el seu propi ordinador. Les pàgines web utilitzen aquests contenidors de dades per a reconèixer l'usuari que es connecta freqüentment a aquesta pàgina i saber-ne les preferències i els requisits.

Ús de galetes

Un exemple senzill de l'ús de galetes ocorre quan l'usuari intenta connectar-se a certa pàgina de notícies en què està registrat, i no n'ha de recordar l'identificador d'usuari ni la contrasenya ja que el servidor accedeix a la galeta que hi ha guardada a l'ordinador i obté aquestes dades sense la interactuació de la persona.

Aquest sistema de reconeixement automàtic facilita la interacció dels usuaris; no obstant això, les galetes poden emmagatzemar informació personal que pot ser perillosa des del punt de vista de la privadesa. En aquest sentit, aquests contenidors poden emmagatzemar qualsevol dada que l'usuari hagi introduït en un formulari (per exemple, l'adreça de casa seva) i, evidentment, poden guardar dades més comunes com l'adreça de correu electrònic o el proveïdor de serveis d'Internet. L'extensa implantació d'aquesta tecnologia, la senzillesa d'execució i la capacitat d'acció que té sense requerir l'atenció de l'usuari converteix les galetes en un punt controvertit pel que fa als perills de privadesa i l'empremta digital dels usuaris.

Vegeu també

La generació i la utilització de galetes les comentarem amb detall en el subapartat 1.2.1. Més endavant, dedicarem el subapartat 1.2 a detallar algunes tecnologies amb característiques similars a les galetes que es poden usar com a substitut.

1.2.1. Galetes HTTP

Les galetes són utilitzades com un mecanisme de comunicació persistent entre els llocs web que les originen i els navegadors dels visitants. La persistència de les galetes es tradueix en el fet que són un simple fragment de text que s'emmagatzema en el disc dur de l'ordinador de l'usuari i permet al servidor web identificar l'usuari cada vegada que es connecta, saber les seves preferències d'usuari, dur a terme funcions de "cistella de la compra" i qualsevol activitat que es pugui fer emmagatzemant un fragment de text en un ordinador. S'ha de destacar que el sistema d'identificació per galetes és la manera més senzilla i efectiva d'identificar un mateix usuari que torna a un lloc web perquè l'estesa utilització d'adreces IP dinàmiques descarta aquest mètode per a aquesta comesa, i el fet de requerir a l'usuari que introdueixi l'identificador d'usuari i la contransenya pot ser un inconvenient per a la persona.

Bases de dades sobre individus

La companyia RapLeaf construeix bases de dades sobre individus utilitzant les seves activitats en xarxes socials, l'historial de compres i altres interaccions amb el Web.

Les dades recollides s'utilitzen per a campanyes de publicitat dirigida.

Com ja hem dit, les galetes són simplement fragments de text. No són programari, cosa que implica que no poden ser programades, no poden portar virus, programes espies (*spyware*) o programes maliciosos (*malware*) en general. No obstant això, sí que poden ser utilitzades per programari nociu per a saber les activitats que fa l'usuari a Internet. S'ha de recordar que les galetes indiquen les pàgines a les quals accedeix l'usuari, l'identificador d'usuari i la contrasenya que utilitza, les seves preferències i certes dades personals que hagi introduït en algun moment. De la mateixa manera, les galetes poden ser robades per un pirata que accedeixi a l'ordinador de l'usuari.

En resum, les galetes no perjudiquen la privadesa de l'usuari directament, però poden ser utilitzades amb aquest fi.

Funcionament bàsic de les galetes

Com ja hem indicat, les galetes són generades pel servidor. Això es pot fer mitjançant CGI *script*. Després, la galeta és col·locada al costat del client. En una connexió posterior entre el client i el servidor, el servidor web és capaç d'obtenir totes les galetes que hi ha al costat del client.

Atributs de les galetes

Les galetes estan formades per certs atributs. Destaquem els més rellevants des del punt de vista de la privadesa:

1) ***Domain***: un servidor solament pot accedir a galetes generades per un altre servidor dins del mateix domini.

Exemple

Els servidors de bali.vacation.com i mexico.vacation.com poden accedir a les galetes generades per servidors de vacation.com.

2) ***Path***: aquest atribut funciona de manera similar al domini ja que limita la visibilitat de les galetes basant-se en el *path* de l'URL.

Exemple

En aquest sentit, una galeta en què s'especifica *domain=bali.vacation.com* i *path=account* fa que solament s'accedeixi a la galeta des de pàgines que estan situades a l'URL o que en pengen: bali.vacation.com/account. Aquest atribut es pot deixar inactiu per a evitar aquesta restricció.

3) ***Expires*** i ***Max-Age***: aquest atribut indica al navegador quan s'ha d'eliminar una certa galeta. La data marca exactament el dia i l'hora d'expiració. Hi ha una variació (RFC 2965) en què es pot especificar el temps de vida de la galeta en segons des que va ser rebuda del servidor. Les galetes marcades per a eliminar-les es destrueixen en tancar el navegador.

4) ***Secure***: aquest atribut indica al navegador que la galeta en qüestió només pot ser utilitzada en connexions encriptades. El servidor ha de situar una galeta d'aquest tipus mitjançant un canal segur.

5) ***HttpOnly***: aquest atribut indica al navegador que la galeta en qüestió només pot ser utilitzada pel protocol HTTP. Això evita que *scripts* situats al costat del client accedeixin a les galetes i, per tant, que puguin ser robades per atacs basats en *cross-site scripting*.

Cross-site scripting

El *cross-site scripting* generalment inclou qualsevol atac que permeti executar codi de *script*, com VBScript o Javascript, en el context d'un altre lloc web.

Aquests errors es poden trobar en qualsevol aplicació que presenti informació en un navegador web. El problema és que generalment no es validen correctament les dades d'entrada.

Creació de l'empremta digital mitjançant galetes

A continuació mostrem una sèrie de passos que indiquen com poden utilitzar les galetes els servidors web per a guardar un registre de les pàgines que ha visitat l'usuari:

1) L'usuari visita portal.com, un web gestionat per advts.com, i selecciona amb el ratolí un bàner corresponent a shoe.com, també gestionat per advts.com.

2) El servidor advts.com col·loca una galeta en el navegador, això és, portal.com::zapatos.com, i encamina l'usuari a shoe.com, web al qual passa també la informació corresponent.

3) L'usuari visita un bàner de vacaciones.com situat a zapatos.com, i el lloc vacaciones.com també és gestionat per advts.com.

4) El servidor advts.com obté la galeta del navegador i l'actualitza a portal.com::zapatos.com::vacaciones.com; llavors encamina l'usuari a vacaciones.com.

Lectura recomanada

En la pàgina web següent trobareu més informació respecte a atacs basats en *cross-site scripting*: "The Cross-Site Scripting (XSS) FAQ".

Problemes de privadesa associats a l'ús de galetes

Exemple

Seguint l'exemple del subapartat anterior, imaginem-nos que advts.com introdueix a la galeta un identificador únic (per exemple, 1234) per a l'usuari que accedeix a zapatos.com. Dins d'aquest lloc web, l'usuari compra unes sabatilles de tresc. Després, clica a un bàner de vacaciones.com. Aquest últim lloc és dins del domini d'advts.com i, per tant, pot accedir a la galeta del navegador de l'usuari. A partir de la galeta, s'obté l'identificador 1234 assignat i pot contactar amb zapatos.com per saber quines sabates ha comprat l'usuari 1234 en el lloc web. Després de saber que l'usuari ha comprat unes sabatilles de tresc, vacaciones.com pot oferir a l'usuari paquets de vacances enfocats al muntanyisme.

En aquest petit exemple hem comprovat la importància de la informació que desprenem en interactuar amb el Web i com les galetes ajuden a fer que els servidors segueixin el nostre rastre i n'obtinguin beneficis. En aquest cas, el problema de privadesa analitzat està associat al fet que tots els webs consultats eren dins del mateix domini i, per tant, tots els servidors que hi estaven involucrats podien recuperar la galeta de l'usuari. Si l'usuari accedeix a un servidor extern a aquest domini, aquest servidor no podrà accedir a la galeta ni al rastre de l'usuari.

Per tant, la utilització de dominis es veu com una mesura de control d'accés a la informació de l'usuari. No obstant això, aquesta mesura de seguretat pot ser fàcil d'evitar: encara que les galetes només s'envien al servidor que les va definir o a un altre de dins del mateix domini, una pàgina web pot contenir imatges o altres components emmagatzemats en servidors d'altres dominis. De les galetes que es creen durant les peticions d'aquests components se'n diu **galetes de tercers**.

Analitzadors de trànsit

La utilització d'analitzadors de trànsit ha demostrat que és efectiva per a obtenir l'identificador d'usuari i la contrasenya dels usuaris del Facebook mitjançant la captura de les seves galetes que es transmeten per mitjà de connexions HTTP no encriptades.

Amb la utilització de galetes de tercers, una entitat pot fer un seguiment dels usuaris per totes les pàgines en què ha col·locat imatges o components similars.

Finalment, s'ha de destacar que, en una connexió HTTP no encriptada, les galetes viatgen en clar juntament amb la resta d'informació transmesa. Un atacant que intercepti mitjançant un detector o *sniffer* aquesta connexió pot tenir accés a la informació que contenen les galetes. Això representa un problema de privadesa important que pot ser resolt utilitzant seguretat a escala de capa de transport (TLS) per a encriptar la connexió entre navegador i servidor. El protocol HTTPS és un exemple del tipus de connexió que soluciona aquesta situació.

1.2.2. Alternatives a les galetes

Hi ha alternatives vàlides a les galetes a l'hora d'identificar un usuari (o navegador) en particular. No obstant això, aquestes alternatives, generalment, no són tan fiables i, per tant, acaben convertint les galetes en l'opció preferida a la pràctica.

Flash cookies o *local shared objects*

Les *flash cookies* són porcions de dades que els servidors web que utilitzen Adobe Flash guarden en els ordinadors dels usuaris. Generalment, aquests servidors utilitzen les *flash cookies* per a emmagatzemar preferències d'usuari, però també són utilitzades per a obtenir la informació de navegació dels usuaris i evitar així els controls que hagin aplicat aquests usuaris sobre les clàssiques galetes HTTP. Singel (2009) explica que, en aquelles dates, més de la meitat dels llocs web més rellevants utilitzaven *flash cookies* per a seguir usuaris i emmagatzemar informació sobre aquests usuaris. De tots aquests llocs web, solament n'hi havia quatre que n'esmentaven la utilització en les seves polítiques de privadesa.

Amb la configuració per defecte, el client *flash* no sol·licita a l'usuari cap permís per a emmagatzemar *flash cookies* en el disc dur. A més, la configuració inicial de l'aplicació pot emmagatzemar fins a 100 kb d'informació en el dispositiu. De manera similar a les galetes HTTP, una *flash cookie* no pot ser llegida per servidors fora del domini de l'entitat que la va crear.

Adreça IP

Més amunt, hem fet una referència breu a la utilització d'adreces IP com a tècnica poc fiable per a fer un seguiment d'usuaris. Aquest mètode es basa principalment a emmagatzemar l'adreça IP de l'ordinador que sol·licita les pàgines. Això és així a causa del mateix protocol IP. Cada sol·licitud que rep el servidor té l'adreça IP de l'ordinador en el qual s'executa el navegador (o del seu servidor intermediari en cas que se'n faci servir). El servidor pot guardar aquesta informació, independentment de l'ús de galetes.

El problema principal d'aquesta mesura és l'escassa fiabilitat que té a l'hora d'identificar unívocament un determinat usuari. Els ordinadors i els servidors intermediaris utilitzats poden ser compartits per diversos usuaris o el mateix ordinador pot tenir assignades diferents adreces IP en diferents sessions (com el cas típic d'assignació dinàmica d'IP). A més, s'ha de destacar que aquesta tècnica solament proporciona identificació; no pot substituir l'ús de galetes per a emmagatzemar preferències i aplicacions similars.

Identificació mitjançant navegador

Eckersley (2010) ha investigat el nivell de singularitat de les petjades dels navegadors a partir de les dades de configuració que transmeten aquests navegadors per petició als diferents llocs web. En particular, ha analitzat les petjades d'una gran quantitat de navegadors diferents que es van connectar a un determinat lloc web (http://panopticlick.eff.org). Dels resultats obtinguts es va concloure que la petjada d'un navegador (la configuració d'aquest navegador) conté almenys 18,1 bits d'entropia. Això implica que, en triar un navegador a l'atzar, en les millors circumstàncies solament n'hi haurà un entre 286.777 que compartirà la mateixa configuració. En el cas de navegadors que utilitzen Flash o Java, augmenta el grau de singularitat. En aquests casos, els navegadors contenen 18,8 bits d'informació que pot ser utilitzada per a identificar-los unívocament. Com a resultat d'això, en l'estudi esmentat el 94,2% dels navegadors amb Flash o Java utilitzaven una configuració diferent de la resta.

Mitjançant l'observació dels visitants que tornaven més d'una vegada a la pàgina web, l'autor va ser capaç d'estimar la rapidesa amb què canviaven les petjades dels navegadors. Es va concloure que les configuracions dels navegadors canvien ràpidament, però també es va demostrar que fins i tot les heurístiques més simples eren capaces d'identificar un mateix navegador que havia estat modificat (generalment actualitzat). En aquest sentit, els resultats de l'estudi indiquen que les heurístiques utilitzades identificaven correctament un navegador actualitzat amb una probabilitat del 99,1%. El fals positiu ocorria amb una probabilitat del 0,86%.

URL (*query string*)

Una altra tècnica per a seguir la navegació d'un determinat usuari que no accepta galetes consisteix a incrustar informació en l'URL. Normalment s'usa per a aquest fi la *query string*, que és part de l'URL, però també es poden utilitzar altres parts.

Format de la *query string*

La *query string* és la part d'un URL que conté les dades que s'han de passar al servidor web perquè generi la pàgina que el navegador sol·licita.

Un URL típic que contingui una *query string* té el format següent: http://server/path/program?query_string.

Aquest mètode consisteix en el fet que el servidor web afegeix *query strings* als enllaços de la pàgina web que conté, a l'hora de servir-la al navegador. Quan l'usuari segueix un d'aquests enllaços modificats, el navegador retorna al servidor la *query string* afegida.

Les *query strings* utilitzades d'aquesta manera són molt similars a les galetes: bàsicament, les dues tecnologies es basen en l'ús de porcions d'informació definides pel servidor i retornades pel navegador de l'usuari posteriorment. No obstant això, hi ha diferències, ja que una *query string* és part d'un URL. Si l'URL és reutilitzat posteriorment, s'envia al servidor la mateixa porció d'informació. Si, per exemple, les preferències d'un usuari estan codificades en la *query string* d'un URL, i l'usuari envia aquest URL a un altre usuari per algun mitjà, aquestes preferències seran utilitzades també per aquest altre usuari.

A més, fins i tot si l'usuari mateix accedeix a la mateixa pàgina dues vegades, no hi ha garantia que s'utilitzi la mateixa *query string* les dues vegades. Un exemple d'aquesta situació es dóna quan el mateix usuari arriba a la mateixa pàgina partint de dos orígens diferents: un de provinent d'una altra pàgina del mateix servidor web, i l'altre, d'un cercador. En aquest cas, les respectives *query strings* són normalment diferents, mentre que en cas d'haver utilitzat galetes HTTP aquestes galetes haurien estat idèntiques.

Un altre desavantatge de les *query strings* està relacionat amb la seguretat: emmagatzemar en una *query string* informació que identifica una sessió permet o simplifica diversos atacs contra la seguretat dels usuaris. Per exemple, permet a un atacant fer una fixació de sessió, i forçar així l'usuari a treballar amb una identificació o sessió en particular triada per aquest atacant.

Autenticació HTTP

El protocol HTTP inclou mecanismes d'autenticació (per exemple, *digest access authentication*) que permeten accedir a una pàgina web només quan l'usuari ha facilitat un nom d'usuari i una contrasenya correctes. Una vegada s'han introduït les credencials, el navegador les emmagatzema i les utilitza per a accedir a les pàgines següents, sense tornar-les a demanar a l'usuari.

Des del punt de vista de l'usuari, l'aplicació d'aquestes tècniques tenen un efecte similar al de la utilització de galetes HTTP: el nom d'usuari i la paraula clau només es demanen una vegada, i a partir de llavors l'usuari obté accés a les

pàgines del servidor. Internament, el nom d'usuari i la contrasenya s'envien al servidor amb cada petició del navegador. Aquesta informació pot ser utilitzada per a seguir l'usuari durant la seva navegació de manera efectiva.

S'ha de destacar que s'apliquen mecanismes d'expiració de sessió a aquests mètodes. Per exemple, una sessió en particular normalment expira després d'un període d'inactivitat determinat, de manera que queda invalidada per a una recuperació posterior.

Addicionalment a la mesura anterior, s'ha de considerar l'aplicació de tècniques criptogràfiques a l'enviament del nom d'usuari i la contrasenya ja que, si són enviats en clar, poden ser capturats per un detector.

2. Perfils d'usuari

En aquest apartat primer resumirem quins elements formen un perfil d'usuari. Després explicarem diversos entorns en què poden ser adquirits aquests elements a causa de la publicació activa o passiva que en fan els propietaris. Finalment, introduirem diversos escenaris en què es mostra la importància dels perfils d'usuaris i com els poden explotar les empreses.

2.1. Atributs dels perfils d'usuari

A continuació resumirem els diferents elements que formen un perfil d'usuari:

- **Atributs personals**: són les dades personals típiques: nom, edat, sexe, adreça, etc.

- **Interessos**: es refereix als temes preferits o d'interès de l'usuari, com per exemple futbol o cotxes.

- **Opinions**: són els punts de vista que té l'usuari sobre certs temes, com per exemple cinema o música.

- **Topologia d'amistats**: són els amics de l'usuari i les seves identitats.

- **Localització**: són els llocs que freqüenta l'usuari, les rutes habituals que segueix per a anar a la feina, a casa, etc.

2.2. Creació dels perfils d'usuari

Els elements que formen part d'un perfil d'usuari poden ser obtinguts en les diferents interaccions actives o passives d'aquest usuari amb diverses aplicacions disponibles a Internet. A continuació, analitzarem a escala general les aplicacions més comunes i la informació personal que se'n pot extreure.

2.2.1. Creació activa: aplicacions socials

Aquest subapartat s'ocupa de les aplicacions que generen una empremta digital de l'usuari de manera activa, és a dir, l'usuari mateix amb les seves interaccions publica de manera conscient la seva informació personal i la identifica com a seva. Hi ha una gran quantitat d'aplicacions socials que exploten el Web 2.0. A continuació analitzarem les categories més rellevants.

Xarxes socials

Krishnamurthy i Wills (2010) han fet una anàlisi de l'accessibilitat i la disponibilitat de certs atributs personals en dotze xarxes socials (Facebook, LinkedIn, etc.). Específicament, els elements que buscaven eren els següents:

- Foto personal
- Localització
- Sexe
- Nom
- Amics
- Activitats
- Edat
- Escoles
- Lloc de treball
- Data de naixement
- Codi postal
- Adreça de correu electrònic
- Número de telèfon
- Adreça física

En la taula es mostra el grau de disponibilitat de cada atribut (situat a cada fila de la taula). La primera columna indica el nombre de xarxes socials en què l'atribut en qüestió està disponible per a tots els usuaris de la xarxa i el propietari no pot restringir l'accés. L'atribut en qüestió pot ser accessible fins i tot per a individus externs a la xarxa social. La segona columna mostra el nombre de xarxes socials en què l'atribut està disponible als usuaris mitjançant la configuració de privadesa per defecte, però en aquest cas l'usuari pot restringir el seu accés a voluntat. La tercera columna mostra el nombre de xarxes socials en què l'atribut en qüestió pot ser emplenat pels usuaris, però per defecte el valor que té no es mostra a tothom. La quarta columna mostra el recompte de les xarxes socials en què l'atribut en qüestió no forma part del perfil de l'usuari i, per tant, la seva informació no està disponible.

Grau de disponibilitat d'atributs personals a les xarxes socials

Atributs	Nivell de disponibilitat			
	Disponible sempre	Disponible per defecte	No disponible per defecte	No disponible mai
Foto personal	9	2	1	0
Localització	5	7	0	0
Sexe	4	6	1	0
Nom	5	6	1	0
Amics	1	10	1	0
Activitats	2	8	0	2

Atributs	Nivell de disponibilitat			
	Disponible sempre	Disponible per defecte	No disponible per defecte	No disponible mai
Edat	2	5	4	1
Escoles	0	8	1	3
Feina	0	6	1	5
Aniversari	0	4	7	1
Codi postal	0	0	10	2
Correu	0	0	12	0
Telèfon	0	0	6	6
Adreça	0	0	4	8

Consumer Reports, el 2010, va publicar una enquesta feta en dues mil llars nord-americanes respecte a quina informació personal publicaven al Facebook. En aquesta enquesta s'ofereix un punt de vista complementari del que mostra la taula anterior. En la taula següent s'indica el tant per cent d'individus sobre el total de llars enquestades que publiquen obertament un cert atribut a la xarxa social Facebook.

Lectura complementària

Consumer Reports National Research Center (2010). "Annual State of the Net Survey". *Consumer Reports* (vol. 75, núm. 6).

Usuaris que publiquen certs atributs a les xarxes socials

Atribut	Percentatge
Nom complet	84%
Fotos personals	63%
Correu	51%
Data de naixement	42%
Fotos de familiars	24%
Noms de família i amics	19%
Feina	17%
Noms de familiars	16%
Adreça de casa	7%
Número de mòbil	7%
Número de telèfon fix	4%

A més d'atributs personals, les xarxes socials també emmagatzemen les relacions d'amistat entre usuaris i els seus interessos pel que fa a música, pel·lícules, esports, etc.

Finalment, hi ha algunes xarxes socials que també inclouen informació sobre la ubicació dels usuaris en temps real.

Blogs, microblogs i fòrums

Aquest tipus d'aplicacions (per exemple, Twitter i Google Blog) proporcionen un conjunt d'atributs personals que és similar al que ofereixen les xarxes socials. No obstant això, en aquest cas, el fet d'obtenir els interessos dels usuaris pot ser una tasca més complexa, ja que generalment l'usuari no els pot seleccionar de manera explícita.

En els blogs i microblogs (i aplicacions similars), una gran part de la informació personal (com interessos i opinions) surt en els missatges dels usuaris: si un determinat usuari parla en el seu blog sobre la navegació, altres usuaris entenen que està interessat en vaixells, regates i, possiblement, altres esports que es desenvolupen al mar.

No obstant això, per a un ordinador pot ser més difícil entendre els interessos i les opinions que s'amaguen en el text d'un blog. El tractament de dades de text i la interpretació de la semàntica d'aquestes dades és una tasca complexa. De fet, la semàntica és un tret inherentment humà que es defineix mitjançant un consens social (Sánchez, Isern i Millán, 2010). En conseqüència, la interpretació semàntica de dades en format text es basa en evidències oposades en una font o diverses fonts de coneixement construïdes de manera manual. La idea que hi ha al darrere d'aquest procés és imitar el raonament humà usant el coneixement implícit o explícit.

Un possible mètode per a aconseguir això és l'ús del coneixement estructurat modelitzat en forma de taxonomies o més generalment ontologies.

> Les ontologies fan referència a la formulació d'un esquema conceptual exhaustiu i rigorós dins d'un domini o diversos dominis concrets, amb la finalitat de facilitar la comunicació i l'intercanvi d'informació entre diferents sistemes i entitats. Les ontologies es construeixen perquè siguin processades per les màquines i, per tant, poden ser aplicades en aquest camp que ens ocupa.

En aquest sentit, les ontologies s'han utilitzat amb èxit en àrees relacionades amb l'extracció d'informació de recursos textuals (Sánchez, Isern i Millán, 2010).

Iniciatives com el web semàntic (Berners-Lee, Hendler i Lassila, 2001) han propiciat la creació de moltes ontologies que abracen des d'un àmbit general fins a temes concrets. Gràcies a la utilització d'aquestes fonts de coneixement, és possible mapar certes paraules oposades en textos generats per usuaris (per

exemple, "esports aquàtics") als conceptes ontològics que hi estan relacionats (per exemple, per a esports aquàtics, "els esports que impliquen una activitat física a l'aigua").

Addicionalment, les eines d'extracció d'informació de recursos textuals són capaces d'obtenir l'opinió dels usuaris respecte a certs temes. Per exemple, si un usuari publica la seva opinió sobre una certa pel·lícula al Twitter, aquest tipus d'esquemes poden analitzar si la qualificació de l'usuari és positiva o negativa.

Compartició de continguts multimèdia

Aquest tipus d'aplicacions (YouTube, Picasa, etc.) es basen en l'intercanvi de continguts multimèdia (vídeos, fotos, etc.). En aquest escenari, la informació personal és revelada principalment a causa de l'ús de metadades o *tags* vinculades a cada arxiu multimèdia.

L'etiquetatge social és una manera informal d'assignar etiquetes definides per l'usuari als diferents elements compartits. En lloc de classificar el contingut publicat d'acord amb les directrius de classificació bibliogràfiques, els usuaris defineixen els seus propis termes de manera informal basant-se únicament en les associacions que evoca l'element que volen classificar.

Grootveld i altres (2008) van elaborar un estudi sobre la contribució de les etiquetes socials, les metadades de professionals i les metadades generades automàticament en el procés de recuperació d'informació de vídeos. En aquest treball, hi va haver 194 participants que van etiquetar 115 vídeos, mentre que n'hi va haver 140 més que van fer cerques en la col·lecció de vídeos per a obtenir respostes a vuit preguntes. Els resultats obtinguts mostren que en el context actual les etiquetes socials proporcionen un procés de recuperació efectiu, mentre que les metadades generades de manera automàtica no ho aconsegueixen.

Kakogianni i Soderberg (2010) han proporcionat una categorització de les variables utilitzades pels usuaris en el procés d'etiquetatge. Amb la finalitat d'investigar quins elements utilitzen els usuaris per a etiquetar, els autors van consultar les etiquetes més populars del Flickr. Totes les etiquetes consultades es poden dividir en categories d'acord amb l'anàlisi de la imatge en qüestió. La taula mostra aquesta categorització:

Categories més utilitzades per a etiquetar els continguts multimèdia

	Categoria	Etiqueta
Qui	Persones	família, amics, jo, nadó, noia
	Animals	gat, gos, animals, ocell, ocells

	Categoria	Etiqueta
	Edificis	casa, església, museu
	Objectes	flors, aigua, flor, menjar, cotxe
	Colors	verd, negre, blau, color, vermell
	Paisatge	sol, cel, posta de sol, núvols, platja
Què	Activitats	viatge, vacances, excursió, travessa
	Esdeveniments	noces, festa, aniversari
Quan	Temporada	estiu, hivern, primavera, tardor
	Vacances	Nadal, Halloween
	Moment del dia	nit, dia, vespre
On	Continent	Europa, Austràlia, Àsia
	País	Japó, Itàlia, França, Estats Units, Xina
	Lloc específic	parc, jardí, zoo, casa
	Característiques geogràfiques	platja, neu, carrer, ciutat, mar
Pel que fa a	Abstracte	arquitectura, art, moda, amor

Els autors conclouen que hi ha una forma de pensament similar entre les persones en descriure una fotografia. Per descomptat, el nombre d'etiquetes que es poden utilitzar són infinites, però, a la pràctica, els elements utilitzats tendeixen a centrar-se en aquestes categories.

De manera addicional, les aplicacions d'intercanvi de fotos i vídeos, generalment, permeten als usuaris compartir grans quantitats de dades de localització per a indicar en quin lloc es van generar aquests elements.

Missatgeria instantània

Els usuaris d'aquestes aplicacions, generalment, emplenen un formulari amb alguns atributs personals. Aquests atributs són similars als utilitzats en les xarxes socials i, per tant, els resultats analitzats més amunt són aplicables en aquest subapartat.

Aquest tipus d'aplicacions, en general, ofereixen una llista de contactes a l'usuari i mostren els seus respectius estats (en línia, fora de línia, ocupat, etc.) amb la finalitat de facilitar la comunicació. Aquesta llista de contactes és una representació directa de les relacions d'amistat de cada usuari i funciona de manera similar a la subxarxa d'amics que s'utilitza en les xarxes socials.

Addicionalment, avui dia les aplicacions de missatgeria instantània s'integren als mòbils i als dispositius i són capaces de proporcionar dades de localització.

També s'ha de destacar que amb aquestes eines és possible extreure els interessos i les opinions dels usuaris de les converses en format text. Aquestes eines utilitzen mètodes similars als explicats en el subapartat dedicat a blogs i microblogs. No obstant això, la complexitat del procés depèn de la manera com es gestionen aquests missatges.

Gestió de missatges

La complexitat del procés depèn de la manera com es gestionen els missatges. Per exemple, si són emmagatzemats en el servidor durant molt temps (per exemple, Skype), el procés d'extracció d'informació es pot fer de manera més senzilla que si els missatges només s'emmagatzemen en els ordinadors dels usuaris.

Mons virtuals i jocs en xarxa massius

Els usuaris d'aquest tipus de jocs emplenen un formulari amb alguns atributs personals. Com que és un procés similar al de les xarxes socials, s'esperen uns resultats similars pel que fa a la informació que contenen aquests serveis.

Els mons virtuals i els jocs en xarxa massius, fins a cert punt, es poden considerar representacions gràfiques d'una xarxa social i, per tant, poden contenir dades molt similars a les d'una xarxa social estàndard (interessos, topologia d'amistats, etc.). És molt freqüent que aquestes aplicacions continguin comunicació textual entre els usuaris, de la qual es pot extreure també opinions o interessos.

Per tant, el problema d'aquests entorns no és la inexistència d'informació, sinó la manera d'adquirir-la. Per exemple, en un món virtual els contactes fets entre usuaris poden ser considerats com les relacions d'amistat de les xarxes socials. No obstant això, obtenir aquestes relacions pot ser molt complex depenent de l'estructura del joc analitzat. En canvi, en les xarxes socials aquesta informació hi surt ben reflectida i gestionada. Respecte a opinions i interessos, adquirir-los depèn també de la capacitat d'emmagatzemar i recuperar els missatges generats pels usuaris.

2.2.2. Creació passiva: navegació i motors de cerca

Dins dels mètodes per a crear de manera passiva la petjada dels usuaris ens centrarem en dos de molt comuns i pràcticament inherents a l'ús d'Internet: la navegació entre pàgines web i la utilització de motors de cerca.

Historial de navegació

En l'obtenció dels patrons de navegació dels usuaris, aquests usuaris no hi intervenen directament. Per tant, considerem que aquesta informació s'obté dels usuaris de manera passiva i moltes vegades sense el seu propi coneixement.

A partir de les pàgines visitades per un usuari, la informació bàsica del seu perfil que s'obté són els seus interessos. El procés d'adquirir aquestes dades es fonamenta en l'ús de mecanismes per a identificar unívocament l'usuari que accedeix a diferents pàgines web. Una vegada és identificat l'usuari i són emmagatzemades les pàgines consultades, es pot associar la temàtica d'aquestes pàgines amb els interessos de l'usuari analitzat.

Vegeu també

Els diferents mecanismes que hi ha per a adquirir dades (per exemple, galetes HTTP) han estat detallats a fons en el subapartat 1.2 d'aquest mòdul.

Motors de cerca

Quan un usuari vol buscar un cert terme en un motor de cerca (per exemple, a Google), tecleja les paraules clau de la consulta en una barra de cerca. Llavors, el motor aplica tècniques de recuperació per seleccionar i classificar els resultats. Després d'això, l'usuari avalua la llista de pàgines mostrada i obté la informació.

Juntament amb aquest procés, el motor de cerca construeix un perfil de l'usuari segons les consultes que fa.

Exemple

Google afirma que els seus servidors registren automàticament les sol·licituds formulades pels usuaris incloent-hi la consulta, l'adreça IP, el tipus de navegador, l'idioma del navegador, la data i l'hora de la sol·licitud i una galeta o més que poden identificar unívocament el navegador de l'usuari.

Una vegada és identificat l'usuari i queden registrades totes les consultes que fa, és possible obtenir la temàtica de les consultes que ha fet perquè generalment ja es tracta de paraules clau que es refereixen a algun tema en concret. Si són consultes més elaborades, és possible extreure'n la temàtica aplicant mesures similars a les explicades en el subapartat 2.2.1, dedicat a blogs i microblogs. En qualsevol cas, amb aquest procés és possible saber els interessos dels usuaris i, de fet, els motors de cerca l'utilitzen per a proporcionar cerques personalitzades.

Un exemple d'això ocorre quan un usuari ha fet consultes sobre temes relacionats amb la química i fa una nova consulta sobre el *mercuri*. El motor de cerca pot utilitzar els interessos de l'usuari extrets de consultes anteriors per a mostrar respostes relacionades amb l'element químic abans que respostes relacionades amb el planeta.

Com s'observa, aquest procés no té l'aprovació directa de l'usuari i, per tant, aquesta aplicació participa en la creació passiva de l'empremta digital de l'usuari.

2.3. Explotació dels perfils d'usuari

A continuació introduirem i comentarem diversos escenaris en què els perfils d'usuari són (o poden ser) utilitzats per diferents empreses per a obtenir beneficis.

2.3.1. Explotació dels interessos

Els interessos dels usuaris són molt valuosos. Es poden utilitzar per a proporcionar **publicitat dirigida**. Per exemple, les companyies d'automòbils mostren publicitat relacionada amb les persones que tenen interès en el món de l'automòbil.

Un altre exemple en aquest sentit és el Facebook. Aquesta xarxa social obté beneficis mitjançant la personalització dels anuncis que mostra als usuaris. Com més informació mostren els usuaris en els perfils, més personalitzada és la publicitat que reben (Wortham, 2010).

Google AdWords

Un exemple de l'ús dels interessos per a millorar la publicitat és Google AdWords. Aquesta tecnologia és un mètode que utilitza Google per a fer publicitat patrocinada. Els anuncis que mostra el motor de cerca estan relacionats directament amb les consultes fetes per l'usuari. Google cobra a cada empresa per cada clic fet sobre el seu anunci.

2.3.2. Explotació de les opinions

L'aplicació de models per a agregar les opinions dels diferents col·lectius permet obtenir informació extremament important sobre els seus comportaments i proporciona a les empreses la capacitat de preveure tendències futures. A més, recollir grans quantitats d'opinions sobre certs productes en particular ajuda a dissenyar campanyes de màrqueting i publicitat (Adamic, Leskovec i Huberman, 2006).

Com a demostració d'aquesta situació, Asur i Huberman (2010) van analitzar la capacitat de preveure beneficis respecte a les diferents estrenes cinematogràfiques utilitzant els missatges publicats al Twitter. La primera part de l'estudi es va basar a entendre com es construeixen les expectatives i l'atenció sobre una pel·lícula en particular. La segona part es va centrar en la manera com es propagaven les opinions tant positives com negatives de les persones i com influenciaven els altres. Les conclusions d'aquest treball van ser les següents:

- Les opinions adquirides en entorns socials són indicadors efectius del comportament de les masses en el món real.

- La ràtio de missatges relacionats amb una certa pel·lícula pot ser utilitzada per a construir un model de predicció dels beneficis que obtindrà aquesta pel·lícula. A més, el model generat és més efectiu que l'estàndard utilitzat generalment per la indústria del cinema (Hollywood Stock Exchange) per a aquest propòsit.

Els mercats financers són un altre escenari en què les opinions dels usuaris d'aplicacions socials són utilitzades per a preveure els beneficis obtinguts per les diferents empreses i l'èxit dels seus productes.

El programari dissenyat per a l'anàlisi lingüística és utilitzat per a extreure els sentiments del mercat. En aquest sentit, l'agència Dow Jones va crear un diccionari de 3.700 paraules que indiquen canvis en el sentiment. Uns exemples d'aquestes paraules són *fortalesa*, *guanyador*, *risc* i *conspiració*. Aquests programes analitzen el context de les frases i en detecten la tendència per a avisar posteriorment les empreses sobre comportaments massius.

Bloomberg

L'agència Bloomberg monitora notícies i publicacions al Twitter i avisa els usuaris si hi ha una gran quantitat d'individus que envien missatges respecte a un cert tema (per exemple, Apple).

Un altre escenari en què les opinions dels usuaris han demostrat que són eficaces és en l'organització de viatges i vacances. La firma PhocusWright (2011), encarregada de fer investigacions de màrqueting orientat a viatges, ha confirmat que els entorns socials i les opinions que generen tenen una influència enorme en els hàbits de compra d'aquest tipus de productes.

En aquest sentit, els investigadors conclouen que els usuaris, generalment, atorguen molta importància a les valoracions proporcionades per amics o individus similars a ells, mentre que la publicitat proporcionada pels mitjans habituals no rep tanta atenció.

2.3.3. Explotació de la localització

Els serveis de localització s'han convertit en una eina de màrqueting molt destacada per a petites empreses que depenen del trànsit de clients, com restaurants o bars (Pattison, 2010). El creixement d'aquests serveis es fonamenta en la capacitat que tenen d'explotar els nous dispositius mòbils que tenen una connectivitat excel·lent i que han estat adoptats per una gran quantitat de persones.

Els serveis de localització poden tenir molts papers. Ofereixen al client eines per a relacionar-se, programes de punts, guies de ciutats o valoracions de llocs en particular. Una de les aplicacions més interessants que tenen per a les petites empreses esmentades més amunt és que permeten adquirir dades dels clients que són dins de la seva zona d'influència i presentar-los ofertes, vals de descompte, etc. Per exemple, si un usuari és a pocs metres de cert restaurant, aquest usuari pot rebre un avís al dispositiu mòbil que l'avisi d'aquest restaurant i d'una certa oferta adaptada als seus gustos personals.

Bàsicament, aquestes aplicacions permeten als negocis connectar amb les persones i fidelitzar els clients.

2.3.4. Explotació dels perfils de manera global

En els subapartats anteriors hem introduït casos centrats en l'explotació d'una part del perfil. No obstant això, hi ha escenaris en què els beneficis s'obtenen d'adquirir tot el perfil i utilitzar els diversos elements que el componen depenent de l'ús que hi vulgui donar l'empresa adquirent.

La utilització de perfils per a millorar les estratègies de contractació de treballadors és una realitat avui dia. A més, es constata que l'ús d'aquestes noves tècniques ajuda a reduir els costos en el procés de contractació.

Lectura recomanada

"Social savvy recruiters utilising social media in their recruitment strategy" (2010, maig). *Personnel Today*.

Les companyies d'assegurances també utilitzen els perfils dels usuaris per a obtenir evidències de frau i reduir possibles pèrdues econòmiques en aquest sentit (Li, 2011). La tendència, fins i tot, va una mica més lluny i les companyies del sector estan estudiant la utilització dels perfils per a fixar preus personalitzats per a cada client (Beattie i Stagg-Macey, 2010). En aquest cas, les dades més rellevants que s'han d'extreure dels perfils són estils de vida i problemes mèdics.

La compra de perfils d'usuari a les empreses que ofereixen serveis socials és significativament habitual: per exemple, AOL rep mil peticions de dades per a utilitzar-les en delictes penals o civils (Hansell, 2006). Facebook rep de deu a vint peticions al dia. S'ha de destacar que els perfils els poden sol·licitar tant empreses com governs interessats en qüestions de seguretat nacional.

3. Definició i polítiques de privadesa

En aquest apartat introduïm el concepte de *privadesa* i les implicacions que té respecte a l'adquisició i l'explotació de dades personals mitjançant els mètodes que hem explicat més amunt. També detallem els dos nivells de privadesa que es tenen en compte en el disseny d'eines per a la protecció de la privadesa. Finalment, analitzem de manera general les polítiques de privadesa que s'apliquen a les dades personals dels usuaris recollides per les empreses.

3.1. Definició de *privadesa*

La privadesa és un dret humà fonamental. Aquest concepte engloba altres idees que hi estan relacionades, com la dignitat humana i la llibertat d'expressió. La privadesa s'ha convertit en un dels drets humans més importants de l'edat moderna (Rotenberg, 2000).

La privadesa és reconeguda en diverses regions i cultures d'arreu del món. És un dret protegit per la Declaració Universal dels Drets Humans, el Pacte Internacional de Drets Civils i Polítics, i per moltes altres organitzacions internacionals i tractats regionals de drets humans. Gairebé tots els països del món inclouen el dret a la privadesa en la seva constitució. Com a mínim, aquestes lleis preveuen els drets d'inviolabilitat del domicili i el secret de les comunicacions. A més, les constitucions redactades recentment inclouen drets específics per a l'accés i el control de la informació personal per part de l'individu mateix. En el cas dels països en què la constitució no esmenta explícitament el dret a la privadesa, els tribunals han trobat reflectit aquest dret en altres disposicions (Laurant, 2003).

Es poden diferenciar quatre aspectes diferents respecte a la privadesa:

- **Privadesa de la informació**. Aquest aspecte implica establir normes que regeixen el recolliment i el tractament de dades personals com dades econòmiques, mèdiques i registres governamentals.

- **Privadesa corporal**. Aquest aspecte abraça la protecció física de les persones contra procediments invasius com anàlisis genètiques o proves de drogues.

- **Privadesa de les comunicacions**. Aquest aspecte inclou la seguretat i la privadesa del correu, telèfon, correu electrònic i qualsevol altra forma de comunicació.

- **Privadesa territorial**. Aquest aspecte implica establir límits a la intrusió en els entorns domèstics i altres escenaris, com el lloc de treball o l'espai públic. Això inclou cerques, videovigilància i controls d'identitat.

L'empremta digital tractada en aquest mòdul s'enfoca, principalment, a la creació, la utilització i el control d'aquesta empremta en l'àmbit d'Internet. Per tant, solament en els escenaris considerats s'hi apliquen els aspectes relacionats amb la privadesa de la informació i la privadesa de les comunicacions.

La privadesa de les comunicacions es proporciona, la major part, aplicant mesures de seguretat que limitin els atacs a les transmissions. Concretament, les solucions associades a aquest aspecte estan relacionades amb la utilització del maquinari adequat, les actualitzacions de programari i les eines específiques de seguretat informàtica.

En el cas de la privadesa de la informació, aquest aspecte està clarament relacionat amb el problema del maneig de dades de caràcter personal facilitades tant pels usuaris mateixos (creació activa de l'empremta digital) com obtingudes de manera invisible per a l'usuari (creació passiva de l'empremta digital). Per tant, aquest apartat se centra en aquest últim aspecte de la privadesa.

Vegeu també

En l'apartat 4 oferim possibles solucions per a mitigar el problema de l'explotació de dades personals.

Pel que fa al disseny d'esquemes que proporcionin privadesa de la informació, es poden definir dos nivells de privadesa: **anonimat** i **no-enllaçabilitat**. Un sistema manté l'anonimat dels usuaris quan impedeix que la seva identitat es faci pública. La no-enllaçabilitat és un nivell més fort que l'anonimat i es refereix al fet que les diferents interaccions d'un mateix usuari amb un cert sistema no puguin ser relacionades entre si. La no-enllaçabilitat impedeix seguir usuaris i crear perfils.

Per a mostrar la diferència entre aquests dos nivells, es poden introduir dues mesures de privadesa senzilles, una per cada nivell. Per a proporcionar anonimitat, n'hi pot haver prou amb l'ús de pseudònims, que permet a l'usuari que les seves activitats digitals s'associïn a una identitat falsa, de manera que proporciona privadesa a la seva identitat real. No obstant això, l'ús de pseudònims no proporciona no-enllaçabilitat, ja que totes les activitats que es fan amb aquest pseudònim es poden relacionar entre si. Si en alguna circumstància, una tercera entitat és capaç de descobrir la identitat real que hi ha rere el pseudònim, aquesta entitat també serà capaç d'enllaçar totes les activitats que s'han fet rere aquest pseudònim amb la identitat real de l'usuari. Per a evitar aquesta situació, un sistema que proporcioni no-enllaçabilitat ha de canviar freqüentment els pseudònims utilitzats per l'usuari seguint algun mecanisme que impedeixi que puguin ser relacionats entre si.

3.2. Polítiques de privadesa: qui és el propietari de la informació personal

En aquest subapartat tractarem de les polítiques de privadesa de les empreses que recopilen dades personals dels usuaris i que construeixen la seva empremta digital tant de manera activa com passiva. Aquestes polítiques són molt rellevants per a decidir la propietat dels usuaris sobre les seves pròpies dades i la manera en què poden ser utilitzades.

3.2.1. Polítiques respecte a la creació activa de l'empremta digital

Respecte a la creació activa de l'empremta digital, els usuaris de xarxes socials, generalment, consideren que la informació que publiquen (el contingut del seu perfil, les fotos, etc.) és de la seva propietat i estan sota el seu control. Això no sempre és així, però.

El 2009, Facebook va fer un canvi en els termes d'ús. Un dels nous paràgrafs afegits deia literalment el següent:

> "Les seccions següents sobreviuran al final del seu compte en els serveis del Facebook: conducta prohibida, contingut d'usuari, la seva política de privadesa, crèdits de regals, propietat, drets de propietat, llicències, peticions, disputes d'usuari, queixes, indemnitzacions, renúncia general, limitació de responsabilitat, acabament i canvis en el servei del Facebook, servei d'arbitratge, llei d'administració, lloc i jurisdicció i altres."

Abans d'aquest canvi, si l'usuari decidia tancar el compte, també desapareixia el contingut que hi havia publicat, però, després d'aquest canvi, el contingut continuava pertanyent a Facebook fins i tot després de tancar el compte, de manera que l'empresa podia vendre les fotos de l'usuari o utilitzar-les per a publicitat sense que aquest usuari en rebés diners a canvi. L'onada de crítiques que va comportar la modificació va obligar l'empresa a restaurar els termes originals; no obstant això, aquesta situació fa patent la importància dels termes d'ús de les aplicacions socials que generalment els usuaris accepten sense llegir.

Les condicions d'ús de les diferents xarxes socials varien de manera significativa entre si.

LinkedIn

LinkedIn, per exemple, exigeix als seus usuaris que concedeixin a l'empresa una "llicència no exclusiva, irrevocable, mundial, perpètua, il·limitada, transferible, transmissible i que proporciona a l'empresa dret de copiar, modificar parcialment, millorar, distribuir, publicar, eliminar, retenir, agregar, usar i comercialitzar de qualsevol manera coneguda ara o en el futur sense cap mena de consentiment, d'avís previ o de compensació per a l'usuari o per a tercers".

Twitter

Twitter, per la seva banda, "no reclama drets de propietat intel·lectual sobre el material proporcionat pels usuaris", i afegeix que els usuaris "poden eliminar el perfil en qualsevol moment mitjançant la supressió del compte i que aquesta acció també elimina qualsevol text i imatges que els usuaris hagin emmagatzemat en el sistema".

Com a conclusió, els usuaris d'aquesta mena d'aplicacions han de revisar els termes d'ús i les polítiques de privadesa de les empreses que ofereixen aquests serveis amb l'objectiu de saber com serà tractada la seva privadesa en el futur i les implicacions que tindrà en la seva empremta digital el contingut que facin públic.

3.2.2. Polítiques respecte a la creació passiva de l'empremta digital

Respecte a la **creació passiva de l'empremta digital**, de manera general es pot estudiar la política de privadesa aplicada pel Google als seus serveis (motor de cerca, calendari, blog, correu electrònic, localització, mapes, etc.), els quals utilitzen diferents mètodes d'identificació d'usuaris (per exemple, galetes HTTP i credencials d'accés) i adquireixen moltes dades personals en certs casos sense la col·laboració activa dels usuaris.

En les condicions de servei del Google, concretament en l'apartat dedicat a informació personal i privadesa, s'explica que "l'usuari dels serveis oferts per l'empresa accepta l'ús de les seves dades d'acord amb les polítiques de privadesa del Google".

Les polítiques de privadesa del Google es basen en cinc principis que descriuen la manera en què Google tracta la privadesa i la informació dels usuaris en tots els seus productes:

1) Utilitzar la informació per a oferir als usuaris productes i serveis valuosos.

2) Desenvolupar productes que reflecteixin pràctiques i estàndards de privadesa ferms.

3) Recopilar informació personal de manera transparent.

4) Oferir als usuaris alternatives significatives per a protegir la seva privadesa.

5) Supervisar de manera responsable la informació que emmagatzemem.

S'observa que els punts 1 i 3 fan referència explícita a l'adquisició de dades personals i la utilització d'aquestes dades per a augmentar el valor dels productes i serveis aplicats. L'ambigüitat d'aquests principis permet al Google explotar les dades personals per obtenir beneficis tant per a l'usuari com per a Google mateix.

Respecte al tipus d'informació recollida per Google, la seva política de privadesa explicita els elements següents:

- **Informació que proporciona l'usuari**: en registrar-se per a obtenir un compte de Google, el sol·licitant ha de proporcionar informació personal. És possible que es combinin les dades que proporciona l'usuari mitjançant el seu compte amb la informació procedent d'altres serveis de Google o de tercers per a oferir-li una òptima experiència i millorar la qualitat dels serveis proporcionats. Per a determinats serveis, es pot oferir a l'usuari l'oportunitat de decidir si vol, o no, que es dugui a terme aquesta combinació de dades.

- **Galetes**: en accedir al Google, s'envia una galeta o diverses galetes al seu equip o a un altre dispositiu. Les galetes s'utilitzen per a millorar la qualitat del servei, inclosos l'emmagatzematge de les preferències de l'usuari, la millora dels resultats de cerca i de la selecció d'anuncis i el seguiment de les tendències de l'usuari, com per exemple el tipus de cerques que fa. Google també utilitza galetes en els serveis publicitaris perquè anunciants i editors puguin oferir i administrar anuncis a Internet i en els serveis del Google.

- **Informació de registre**: quan l'usuari accedeix als serveis de Google mitjançant un navegador, una aplicació o un altre client, els servidors de l'empresa registren automàticament certa informació. Aquesta informació pot contenir la sol·licitud web, la interacció amb un servei, l'adreça IP, el tipus i l'idioma del navegador, la data i l'hora de la sol·licitud i una galeta o diverses galetes que permeten una identificació exclusiva del navegador o del compte.

- **Comunicacions d'usuaris**: quan l'usuari envia missatges de correu electrònic o altres comunicacions a Google, l'empresa pot conservar aquesta informació per processar les seves consultes, respondre a les seves peticions i millorar els nostres serveis. Si l'usuari envia o rep missatges SMS d'algun dels serveis que tenen aquesta funció, Google pot recopilar i conservar la informació associada a aquests missatges, com el número de telèfon o el contingut del missatge.

- **Llocs del Google afiliats en altres llocs**: Google ofereix alguns dels serveis en altres llocs web o mitjançant aquests altres llocs. És possible que la informació personal que facilita l'usuari mitjançant aquests llocs web s'enviï al Google per poder prestar el servei.

- **Aplicacions externes**: Google pot posar a la disposició dels usuaris aplicacions externes mitjançant els seus serveis. La informació recopilada per Google en habilitar una aplicació externa es processa d'acord amb el que estipula aquesta política de privadesa. La informació recopilada pel pro-

veïdor de l'aplicació externa es regeix per les seves pròpies polítiques de privadesa.

- **Dades d'ubicació**: Google ofereix serveis que tenen registrada la ubicació de l'usuari, com Google Maps i Latitude. Si s'utilitzen aquests serveis, Google pot rebre informació sobre la ubicació real de l'usuari o informació que es podria utilitzar per a determinar la seva ubicació aproximada.

- **Número d'aplicació exclusiu**: alguns serveis, com la barra Google, inclouen un número d'aplicació exclusiu que no està associat a l'usuari ni al seu compte. Aquest número i la informació sobre la instal·lació (per exemple, el tipus de sistema operatiu o el número de versió) es poden enviar a Google en instal·lar o desinstal·lar aquest servei, o quan aquest servei estableix contacte amb els servidors de Google de manera periòdica (per exemple, per a sol·licitar actualitzacions automàtiques del programari).

- **Altres llocs**: aquesta política de privadesa s'aplica únicament als serveis de Google, que no exerceix cap control sobre els llocs que surten en els resultats de cerca, els llocs que inclouen aplicacions, productes o serveis de Google ni els enllaços inclosos en els seus serveis. És possible que aquests altres llocs enviïn les seves pròpies galetes o altres arxius a l'equip de l'usuari, recopilin dades o sol·licitin l'enviament d'informació personal.

Respecte a l'ús que es dóna a la informació recopilada, Google la utilitza per a les finalitats següents:

- Proporcionar, mantenir, protegir i millorar els seus serveis (inclosos els serveis publicitaris) i desenvolupar nous serveis.
- Protegir els drets o la propietat del Google o dels usuaris.

Si aquesta informació es vol utilitzar amb finalitats diferents de l'objectiu pel qual s'ha recopilat, se sol·licita el consentiment previ de l'usuari.

Localització de les dades

Google processa la informació personal en els servidors dels Estats Units d'Amèrica i d'altres països. En alguns casos, la informació personal es processa fora del país de l'usuari.

4. Tècniques per a proporcionar privadesa

Els mecanismes que s'utilitzen per a controlar el contingut de l'empremta digital varien sensiblement segons el procés de creació sobre el qual s'apliquen. En aquest sentit, hem de distingir les mesures dissenyades per a controlar la creació activa de l'empremta digital (aquest procés es dóna principalment amb l'ús d'aplicacions socials) de les mesures que eviten la creació passiva de la petjada (aquest procés es fonamenta principalment en la identificació del navegador de l'usuari i l'anàlisi del comportament d'aquest navegador).

4.1. Control de la creació activa de l'empremta digital

En general, els mecanismes que s'utilitzen actualment per a preservar la privadesa dels usuaris, que de manera activa publiquen recursos i dades personals a Internet, es fonamenten en tres models bàsics:

- Aplicació del sentit comú
- Aplicació de mesures basades en el control d'accés a recursos
- Utilització de tècniques de pertorbació de dades

4.1.1. Sentit comú

Aquest model es basa a no generar contingut susceptible de ser explotat per tercers. Seguint aquesta idea, *Consumer Reports* presenta set consells bàsics per a evitar als usuaris problemes futurs de privadesa i seguretat. Aquests consells els enumerem a continuació:

Lectura recomanada

"7 things to stop doing now on Facebook" (2010). *Consumer Reports* (vol. 75, núm. 6).

1) Els usuaris no han d'utilitzar contrasenyes senzilles (febles).

2) Els perfils d'usuari no han de contenir dates de naixement completes.

3) Els usuaris han d'utilitzar els mecanismes de privadesa proporcionats per les aplicacions basades en el Web 2.0 (per exemple, xarxes socials).

4) Els usuaris no han d'incloure noms o dades de familiars menors d'edat en el contingut multimèdia publicat en el Web 2.0.

5) Els usuaris no han d'indicar l'absència o presència a casa.

6) Els usuaris han d'impedir que els motors de cerca (per exemple, Google) els trobin.

7) Els usuaris menors d'edat no han d'utilitzar les xarxes socials sense la supervisió d'un adult.

4.1.2. Control d'accés a les dades personals

Aquest model de mesures de privadesa permet als usuaris seleccionar les persones que poden accedir a una dada personal en particular. D'aquesta manera, no es modifiquen els recursos publicats per l'usuari i, per tant, un atacant amb prou drets d'accés seria capaç d'obtenir un perfil complet i real de l'usuari.

Les mesures de control d'accés són bàsiques en el procés de creació activa de l'empremta digital. Concretament, són utilitzades generalment en aplicacions del Web 2.0 com les xarxes socials. S'ha de destacar que, com que són mesures aplicades conscientment per l'usuari, no poden ser utilitzades en el procés de creació passiva de l'empremta digital.

Les mesures de control d'accés es fonamenten en tres tecnologies diferents:

- **Configuració individual de privadesa**. Aquest mecanisme és el més senzill d'implementar i, de fet, és l'utilitzat per defecte en les xarxes socials estàndard i les aplicacions que hi estan relacionades. Es basa a seleccionar entre moltes opcions de privadesa proporcionades per l'aplicació mateixa la configuració idònia per a cada persona. Assumint que l'amfitrió (*host*) de l'aplicació web sigui honest, el sistema és efectiu. No obstant això, un estudi elaborat per Bilton (2010) posa en relleu la complexitat dels controls de privadesa proporcionats per Facebook. Bilton indica que un usuari que vulgui evitar la disseminació de la majoria de la seva informació personal ha d'utilitzar més de cinquanta botons i seleccionar entre més de cent setanta opcions diferents. Aquesta complexitat facilita l'error i augmenta la possibilitat d'assignar drets erronis a entitats deshonestes. També hi ha xarxes socials basades en una arquitectura distribuïda (per exemple, Diáspora) en què cada usuari gestiona el seu propi servidor web, que conté tots els seus recursos i als quals pot aplicar les mesures d'accés que consideri necessàries. En aquest cas, la inexistència d'un servidor central proporcionat per una empresa en particular facilita que l'usuari en faci un control més bo.

- **Utilització de criptografia**. En aquest cas, els continguts que es publiquen s'encripten utilitzant una determinada clau secreta que solament tenen les persones autoritzades a accedir a aquest recurs (Graffi i altres, 2009). La criptografia en ús pot ser tant simètrica com asimètrica. Aquesta mesura de privadesa pot ser aplicada als continguts independentment de l'aplicació web; per tant, la seguretat proporcionada és més gran que en el cas anterior. Addicionalment, la complexitat del sistema es basa a triar qui té accés a cert recurs; per tant, la complexitat en la configuració no és especialment rellevant. Com a inconvenient greu d'aquesta mena de mesures, indiquem

que no són compatibles amb totes les aplicacions del Web 2.0 i, per tant, la implantació que tenen està molt limitada.

- **Qualitat de les relacions entre l'usuari i la resta d'entitats**. En aquest cas, l'accés als recursos s'aconsegueix mitjançant l'anàlisi de la qualitat de la relació entre l'usuari que publica el recurs i l'entitat que intenta accedir a aquest element (Banks i Wu, 2009). Depenent de la qualitat d'aquesta relació, es denega o s'accepta l'accés. Aquesta solució es basa en la utilització d'un sistema d'anàlisi i comptabilització de confiances. L'objectiu que té és automatitzar el procés de configuració de la privadesa.

4.1.3. Pertorbació de les dades personals

Les mesures basades en la pertorbació se centren a modificar les dades personals per a augmentar-ne l'ambigüitat i, per tant, la privadesa dels usuaris que les han generades. Aquest procés es coneix com a *generalització*.

La generalització d'atributs personals és utilitzada habitualment pels mètodes de pertorbació de bases de dades (Domingo-Ferrer i altres, 2008). Encara que aquestes tècniques no s'utilitzen generalment per al control de l'empremta digital a Internet, hi ha alguns treballs en la literatura que utilitzen aquest tipus de mesures fins a un cert nivell.

En aquest sentit, Hay i altres (2008) presenten un treball basat en l'anonimització de xarxes socials, i se centren específicament en la pertorbació de la xarxa d'amics de cada usuari. Addicionalment, aquest treball proposa la generalització, l'eliminació i l'aleatorització dels atributs personals de l'usuari.

També és habitual aplicar tècniques de pertorbació a les dades referents a la localització dels usuaris. Aquestes dades s'associen als recursos publicats pels usuaris. Les mesures proposades en aquest àmbit se centren a reduir el nivell de detall de la localització: parlem de proximitat en comptes de localització precisa (Glassman, 2010).

4.2. Control de la creació passiva de l'empremta digital

La generació passiva de l'empremta digital d'un usuari depèn principalment de la capacitat d'identificar correctament les seves interaccions mitjançant l'ús de galetes o altres mètodes. Per tant, les mesures per a evitar o controlar aquest procés es basen a impedir la identificació correcta per part de l'entitat que fa el seguiment de l'usuari.

Els sistemes d'identificació demostren que mesures senzilles com l'ús de pseudònims per part dels usuaris són completament ineficaces perquè la identificació i l'anàlisi del rastre de l'usuari, generalment, ocorre sense que la víctima s'adoni de la situació. És la mateixa màquina de l'usuari la que delata la identitat real malgrat que la persona s'amagui darrere d'un pseudònim. Per això, les eines per a preservar la privadesa han de ser més sofisticades i s'han de focalitzar a fer que l'entitat que interactua amb la màquina de l'usuari no pugui identificar aquesta màquina. Aquest procés, per extensió, proporciona anonimitat al propietari.

Vegeu també

Els sistemes d'identificació els estudiarem en el subapartat 1.2 d'aquest mòdul.

En la literatura, les eines que proporcionen aquestes característiques s'emmarquen dins dels esquemes basats en canals anònims.

Al començament, els canals anònims es van centrar en el correu electrònic, per la gran importància que tenia aquest sistema de comunicació. La primera aplicació pràctica va sorgir gràcies a les investigacions de Baran (1964), que va crear un sistema amb el qual es podien comunicar dues persones mitjançant la participació d'una tercera part de confiança. Aquesta entitat s'encarregava de rebre les comunicacions i reenviar-les al destinatari sense que aquest destinatari (ni ningú que escoltés el canal) sabés que havia enviat al principi el missatge. Això s'aconseguia eliminant la capçalera d'identificació i introduint les dades de l'entitat de confiança. Aquesta tecnologia va evolucionar fins a arribar a l'actual, coneguda com a *reenviador de correu* o *remailer*.

El 1981, David Chaum va presentar en el seu treball "Untraceable electronic mail, return addresses, and digital pseudonymous" una solució per a enviar informació de manera segura i anònima, i va iniciar així la investigació sobre els canals anònims. Aquesta solució es basava en la utilització de criptografia asimètrica o criptografia de clau pública per a crear un sistema de nodes (*mixes*). Des de llavors s'ha avançat molt en aquesta àrea de la privadesa a Internet i han estat moltes les solucions plantejades tant des d'un punt de vista teòric com pràctic.

Més endavant, gràcies a l'aparició de noves eines i aplicacions d'Internet, va sorgir la necessitat de garantir la privadesa durant la utilització d'aquestes eines i aplicacions. Actualment, els canals anònims permeten fer de manera anònima tota mena d'activitats a Internet, com ara accedir a pàgines web, utilitzar motors de cerca o enviar correus electrònics.

De manera general, classifiquem els diferents tipus de canals anònims en tres grups depenent de la tecnologia en què es basen:

- Node central de confiança
- Xarxes de nodes
- *Onion routing*

4.2.1. Node central de confiança

Dins d'aquesta categoria trobem la utilització de servidors intermediaris per a suplantar l'usuari real davant un proveïdor de serveis web.

Un servidor intermediari consisteix en un servidor que s'encarrega d'acceptar connexions d'un grup d'usuaris i reenviar-les a la destinació sol·licitada. El servidor intermediari amaga les adreces IP dels usuaris i les substitueix per la seva. D'aquesta manera totes les connexions dels usuaris estan associades a una adreça que no és la seva, i mantenen la privadesa. Els servidors intermediaris s'utilitzen generalment per a aconseguir una navegació web anònima (per exemple, accés a pàgines web o utilització de motors de cerca).

Unes innovacions recents en aquest camp han permès als servidors intermediaris oferir serveis addicionals com són autoritzar filtracions de continguts (bloqueig de galetes, bloqueig de contingut per a adults, etc.) o control d'accés (s'ha d'iniciar la sessió per a utilitzar el servidor intermediari).

S'ha de destacar que un dels problemes associats als mètodes que depenen d'un servidor o uns servidors de confiança és que l'amenaça contra la privadesa de l'usuari es trasllada del proveïdor de serveis a aquestes entitats que saben totes les interaccions dels usuaris i poden generar els seus perfils.

En el camp dels reenviadors de correu trobem diverses propostes. Anon.penet.fi va ser el primer reenviador de correu honest. El sistema es basava en un servidor que contenia una taula de correspondència entre adreces de correu i pseudònims perquè els usuaris poguessin enviar i rebre missatges sense haver-se d'identificar. Anon.penet.fi garantia l'anonimat eliminant qualsevol informació que pogués identificar l'usuari emissor (capçaleres, dades dins del missatge, etc.).

Reenviador de correu

Un reenviador de correu o *remailer* és un servidor que rep missatges, els processa i n'elimina les capçaleres, i els encamina fins al destinatari i proporciona anonimat als participants.

Generalment s'apliquen tècniques criptogràfiques i es combinen diversos reenviadors de correu per a aconseguir un grau més alt d'anonimat.

Aquest sistema presentava importants vulnerabilitats, la més destacada de les quals era la necessitat de guardar aquesta taula que comprometia clarament la privadesa dels usuaris si hi havia algun subjecte que hi tenia accés. Per això, es va dissenyar un nou reenviador de correu anomenat *Cypherpunk* (Parekh, 1996) que introduïa canvis importants respecte a Anon.penet.fi. Les característiques principals són les següents:

- Elimina la taula de correspondència.
- El procés de retransmissió accepta l'ús d'un node o diversos nodes reenviadors de correu encadenats.

- S'eliminen els elements identificatius i s'utilitza criptografia asimètrica per a fer arribar el missatge al node reenviador de correu volgut.

- S'estableix la possibilitat d'introduir ordres per a cada node reenviador de correu de la cadena en la seva capa respectiva d'encriptació. D'aquesta manera, en encriptar el missatge amb la clau pública d'un reenviador de correu concret, es pot indicar alguna funció dins del missatge (esperar-se una estona per a reenviar el missatge, afegir-hi noves capçaleres, etc.).

Cypherpunk és un mètode més robust que l'anterior; no obstant això, també té algunes vulnerabilitats. Per exemple, és feble contra un atacant passiu que escolti l'entrada i la sortida dels reenviadors de correu. Si un node envia els missatges a mesura que els rep, és senzill enllaçar l'entrada amb la sortida i descobrir el contingut dels missatges. Un altre problema greu d'aquest esquema és que no té en compte la mida dels missatges i és possible associar missatges d'entrada i de sortida pel nombre de bits que ocupen.

Els servidors Nym (Kaashoek i Mazieres, 1998) són un esquema amb certes semblances amb el reenviador de correu Cypherpunk. La característica principal d'aquests servidors és que no mantenen un registre de les persones que utilitzen el seu servei, i per tant, en cap moment no s'auditen les comunicacions, de manera que es manté l'anonimat dels usuaris. El funcionament és el següent:

- El primer pas és crear una adreça *nym* i configurar-la creant un parell de claus PGP.

- Aquest parell de claus són enviades al servidor Nym juntament amb instruccions o *reply blocks* per a informar-lo d'on pot ser l'usuari per a reenviar els missatges de resposta. El servidor respon que ha rebut la informació.

- A partir d'aquest moment, el servidor guarda la instrucció relacionada amb l'adreça de correu anònima corresponent.

- Quan arriba un missatge destinat a una adreça en concret, aquest missatge no es guarda sinó que s'envia directament a l'usuari corresponent utilitzant les instruccions emmagatzemades. Aquest comportament provoca una debilitat similar a la del Cypherpunk.

Crowds (Reiter i Rubin, 1998) és un sistema col·laboratiu que pretén proporcionar privadesa als usuaris que accedeixen al Web. En aquest esquema, cada node contacta amb un servidor central i rep una llista de participants. L'usuari, llavors, envia la seva petició de web mitjançant un altre usuari triat a l'atzar. Cada node que rep una petició decideix a l'atzar si la reenvia a un altre node de la xarxa o si envia la petició directament al lloc web. La resposta del lloc web s'envia a l'usuari iniciador seguint el mateix camí però en sentit invers. Totes

les peticions que travessen la xarxa s'envien encriptades utilitzant criptografia simètrica entre parelles de nodes. Els problemes associats a aquesta proposta són els següents:

- El servidor central actua com a coll d'ampolla.

- La utilització de parelles de claus entre usuaris i el fet que el tipus de graf que formen els nodes sigui complet implica que cada node ha de guardar un gran nombre de claus criptogràfiques.

- És vulnerable a l'**atac del predecessor**: un grup d'atacants pot entrar a la xarxa i esperar que es formin les cadenes de nodes per a l'enviament de missatges. Si un cert usuari envia molts missatges, surt en la majoria de les cadenes d'enviament i augmenten les possibilitats que els atacants el detectin. Aquest atac depèn principalment del nombre d'atacants que cooperen i del nombre d'usuaris totals del sistema.

Dins d'aquesta línia, s'ha de destacar que hi ha un esquema també basat en l'ús d'un servidor central de confiança, però que se centra específicament en l'enviament de consultes a motors de cerca (per exemple, Google). El protocol *useless user profile* (UUP) (Castellà-Roca i altres, 2009) segueix una idea similar a Crowds, ja que un cert usuari interessat a enviar una consulta no l'enviarà per si mateix sinó que un altre usuari ho farà per ell. No obstant això, en aquest cas, la xarxa és completament dinàmica: quan hi ha diversos usuaris que volen consultar alguna cosa al motor de cerca, es posen en contacte mitjançant el servidor central, que els ajuda a formar un grup. Una vegada format, els membres del grup utilitzen un protocol criptogràfic per a intercanviar les consultes entre si sense saber quina consulta correspon a cada individu. Finalment, cada usuari envia al motor de cerca la consulta que li ha correspost i fa difusió (*broadcast*) de la resposta a la resta de membres del grup. Cada usuari rep solament la resposta que correspon a la seva pregunta i descarta la resta. En acabar aquest procés, desapareix el grup. Si un dels usuaris vol fer una altra consulta ha de sol·licitar al servidor central un nou grup.

Aquest sistema proposat és molt interessant perquè elimina els problemes de Crowds respecte a l'atac del predecessor i la gran quantitat de claus que fan falta. No obstant això, aquest esquema no considera atacants interns i requereix investigació addicional per a cobrir aquest tipus d'atacants mantenint un temps d'execució raonable.

4.2.2. Les xarxes de nodes

Les xarxes de nodes o *mix networks* es fonamenten en l'ús d'un grup de nodes interconnectats entre si i que formen una xarxa en què cada node amaga la correspondència d'entrada i sortida dels seus missatges mitjançant criptografia. L'objectiu és que els missatges de sortida de la xarxa no es puguin relacionar amb els d'entrada i, per tant, que el proveïdor de serveis no sigui capaç d'identificar l'emissor original del missatge. El comportament general de les xarxes de nodes amb la presència de nodes deshonestos es basa en el fet que, mentre hi hagi una certa quantitat de nodes honestos a la xarxa, l'anonimitat dels usuaris queda garantida.

Chaum (1981) va ser el primer a introduir aquest concepte que va acabar essent adoptat i millorat per altres investigadors.

El primer esquema que s'ha de considerar el va proposar Cottrell i altres (2003) amb el sistema anomenat *Mixmaster*. Aquesta proposta preveu solucionar les vulnerabilitats de Cypherpunk:

- Mixmaster no reenvia automàticament els missatges que rep, sinó que els posa en cua fins que en té un nombre determinat. Una vegada aconsegueix la quantitat adequada, els reenvia al node següent de manera aleatòria.

- Un altre canvi important és que s'especifica una mida uniforme per als missatges que s'envien. Si el missatge és més petit, hi afegeix bits de farciment, i si és massa gran, el divideix en blocs de la mida corresponent. D'aquesta manera s'evita que es puguin associar missatges d'entrada amb els de sortida per la mida que fan.

- Cada node només coneix el node següent del camí. D'aquesta manera, si un dels nodes és maliciós, no podrà trobar ni la ruta fins a l'origen del missatge ni la destinació del missatge.

Aquest sistema cobreix els atacs de repetició de missatges assignant identificadors als missatges, i obligant cada node a comprovar en una taula interna si aquest identificador ha estat rebut abans. Si està repetit, es descarta el missatge.

La vulnerabilitat principal del sistema és que solament proporciona anonimitat al qui envia el missatge; es pot dir que només cobreix el camí d'anada del missatge i no una possible resposta.

El sistema Mixminion (Danezis, 2003) es va dissenyar per a solucionar la inexistència d'anonimitat en el camí de resposta de Mixmaster. Les característiques principals del sistema són les següents:

- Mixminion introdueix un sistema de resposta a missatges que garanteix l'anonimat de l'emissor i del receptor establint un canal bidireccional de comunicació.

- Les transmissions en Mixminion es basen en el protocol TCP i la seguretat en els enllaços s'aconsegueix mitjançant el protocol TLS.

- En Mixmaster, s'evitava l'atac de repetició guardant els identificadors dels missatges rebuts. Com que els identificadors no es guarden indefinidament, persisteix la possibilitat d'aquesta mena d'atacs. Mixminion soluciona aquesta situació amb l'associació de claus a missatges i un procés d'actualització d'aquestes claus. Es descarten automàticament els missatges que arriben amb una clau antiga.

Les xarxes de nodes funcionen com una capa d'ofuscació entre l'usuari que envia el missatge i el que el rep. Un dels problemes que es presenten en aquest tipus d'aplicacions és proporcionar garanties als usuaris que els seus missatges han estat processats correctament per la xarxa. En un escenari en què tots els nodes fossin honestos no faria falta, això; aquest requisit, però, no és gaire realista i, per tant, és important presentar proves que el procés s'ha seguit correctament. En aquest sentit, la xarxa de nodes proposada per Chaum incloïa un sistema de rebuts signats pels nodes dissenyat per a demostrar que tenien un comportament honest. Aquest camp de treball s'anomena *robust and verifiable mix constructions*.

Lectura recomanada

G. Danezis; C. Diaz; P. Syverson (2010). "Systems for Anonymous Communication". *CRC Handbook of Financial Cryptography and Security* (pàg. 341-390).

4.2.3. *Onion routing*

Les xarxes de nodes, generalment, incorporen un cost en temps elevat perquè cada node espera un cert temps per a obtenir diversos missatges, i una vegada en té el nombre volgut els processa i reenvia en l'ordre decidit. Els sistemes basats en *onion routing* eliminen aquest retard per augmentar la velocitat de transmissió: els nodes d'aquests sistemes no ofusquen l'ordre d'entrada i sortida dels missatges.

En resum, l'*onion routing* es basa en l'ús de circuits bidireccionals i de baixa latència per a proporcionar anonimitat als usuaris. El que ofusquen aquests esquemes és la ruta de nodes seguida pel missatge. S'utilitza criptografia de clau pública per a establir el circuit entre tots els nodes que hi ha, mentre que les dades es transmeten utilitzant criptografia simètrica.

La primera proposta basada en *onion routing* va sortir a l'obra de Goldschlag, Reed i Syverson (1996). En aquest disseny inicial, hi ha un primer missatge que obre el circuit mitjançant la xarxa. Per a generar el circuit a cada node s'hi assigna un identificador. Aquest primer missatge s'encripta utilitzant diverses capes de criptografia pública. Cada capa pot ser desencriptada solament pel node corresponent utilitzant la clau privada d'aquest node. Aquest primer missatge porta material criptogràfic compartit entre l'emissor i cada node del circuit, que serà utilitzat per a enviar les dades en la fase següent del protocol. Els missatges enviats sempre estan encriptats per capes: en el primer missatge, les capes es construeixen amb claus públiques i, en els missatges següents d'enviament de dades, les capes es generen mitjançant les claus simètriques obtingudes abans. Aquest tipus de transmissions encriptades per capes és el que dóna el nom d'*onion* ('ceba', en anglès) al protocol.

L'objectiu de l'*onion routing* és dificultar l'anàlisi de trànsit feta per un adversari, i protegir així l'anonimitat de l'emissor i receptor del missatge. No obstant això, aquesta tècnica va demostrar ser feble en situacions de poc trànsit en què un atacant passiu podia descobrir la coincidència entre el missatge d'entrada a la xarxa i la sortida d'aquest missatge. D'aquest tipus d'atacs se n'ha dit *timing attacks* i *end-to-end correlation attacks* (Raymond, 2000).

Tor, possiblement, és el disseny més rellevant basat en *onion routing* (Dingledine i altres, 2004). Tor va ser publicat el 2004 i utilitza una xarxa d'encaminadors o *routers* per a reenviar trànsit TCP. Aquesta aplicació està dissenyada específicament per a tractar trànsit web i s'usa generalment juntament amb l'eina Privoxy, encarregada d'eliminar qualsevol component actiu que hi hagi a les pàgines web transmeses, gestionar les galetes, etc.

FoxTor i TorButton

FoxTor i TorButton són dos connectors o *plug-ins* per al navegador Firefox que combinen la xarxa Tor amb l'eina Privoxy.

La xarxa Tor està formada per una llista de servidors voluntaris que actuen com els nodes encaminadors del sistema. Els usuaris que es connecten a la xarxa creen circuits de tres encaminadors triant-los a l'atzar per a fer les transmissions anònimes. Tor no utilitza el primer missatge per a repartir el material criptogràfic simètric entre els nodes del circuit; en comptes d'això segueix un sistema interactiu en què l'usuari es connecta al primer node i li requereix que es connecti al node següent. S'estableix així un canal bidireccional utilitzant el protocol *Diffie-Hellman autenticat* (Diffie i altres, 1992) per a compartir claus simètriques. Una vegada establert el circuit i repartides les claus, el protocol envia les dades seguint el procediment habitual de l'*onion routing*. Igual que la proposta inicial, Tor té dificultats contra un adversari passiu amb una visió global de la xarxa i situacions de poc trànsit.

Encara que l'objectiu de les tècniques basades en *onion* és reduir el cost en temps respecte a les mesures basades en xarxes de nodes, a la pràctica, aquests mecanismes alenteixen considerablement la navegació web. Prenent com a exemple l'enviament d'una consulta a un motor de cerca, Boneh i altres (2007)

van demostrar que fer aquesta consulta amb un circuit Tor de dos nodes (per defecte se n'utilitzen tres) requeria vint-i-cinc vegades més de temps que no pas fer una consulta directa sense anonimat.

Resum

En aquest mòdul ens hem centrat a explicar la situació de la privadesa en l'era de les tecnologies de la informació. Concretament, hem mostrat les implicacions de l'existència d'una empremta digital i les mesures que s'han de prendre per a controlar-la.

En el primer apartat hem tractat del concepte *d'empremta digital* i hem explicat què conté i com es crea. Hem posat un èmfasi especial en els mecanismes utilitzats per a identificar els usuaris cada vegada que utilitzen serveis o aplicacions a Internet. La identificació és el pas inicial per a fer el seguiment i l'estudi dels usuaris.

En el segon apartat ens hem centrat en els perfils d'usuaris. Hem descrit quins atributs formen un perfil d'usuari i hem explicat amb detall els diferents escenaris en què s'obtenen aquestes dades dels usuaris mateixos. Finalment, hem presentat diversos exemples en què es mostra com les empreses exploten els perfils d'usuari per a millorar els seus resultats.

En el tercer apartat hem definit el concepte de privadesa i hem explicat el funcionament de les polítiques de privadesa que apliquen a les dades adquirides les empreses d'Internet. En aquest apartat hem estudiat qui és el propietari de les dades personals una vegada les empreses les han obtingudes.

Finalment, en el quart apartat hem presentat diferents mètodes que hi ha en la literatura, que han estat dissenyats per a preservar la privadesa dels usuaris que utilitzen els variats serveis presents a Internet. En aquest apartat també hem introduït els avantatges i els inconvenients de les diferents propostes i també l'àmbit d'aplicació de cadascuna.

Activitats

1. En aquesta activitat estudiarem quina informació referent a nosaltres es pot trobar a Internet. L'objectiu és avaluar el nivell de privadesa que hem perdut i quina és la tendència. Per a fer l'estudi heu de triar dos companys de classe i presentar una memòria amb els apartats següents:

- Informació oposada. Si trobeu informació sensible l'heu de notificar a la persona interessada i en l'informe heu de comentar el tipus d'informació oposada sense detallar quina és aquesta informació.
- Fonts d'informació consultades.
- Mesures o recomanacions per a disminuir la quantitat d'informació personal accessible a Internet.

A continuació presentem un exemple de la informació que podeu buscar. Aquest exemple solament és una guia i, per tant, el podeu modificar de la manera que considereu convenient.

- Dades bàsiques (nom, àlies, edat, telèfons, fotografia, data i lloc de naixement, alçada, pes, estat civil, família, etc.).
- Dades mèdiques (grup sanguini, operacions, fumador, bevedor, exercici físic, etc.).
- Dades legals i financeres (delictes greus, multes de trànsit, targetes de crèdit, deutes, etc.).
- Aficions (llibres, música, televisió, pel·lícules, esports, etc.).
- Comportament en línia (notícies, blog, web, correu electrònic, xarxes socials, etc.).
- Personalitat (intel·ligència, *myers-briggs*, orientació política, etc.).
- Amenaces potencials (jocs en línia, interès pel terrorisme, etc.).

2. En aquesta activitat estudiarem quina informació es pot trobar en una xarxa social, concretament a la xarxa social Facebook. L'estudi es divideix en una sèrie de fases que es detallen a continuació; aquestes fases són els punts de la memòria que heu de lliurar:

a) Entrada al Facebook. Si teniu un compte al Facebook el podeu utilitzar en l'estudi. Si no n'hi teniu cap, en podeu crear un i afegir-hi alguns amics o companys de classe. En aquesta primera fase heu de descriure els punts següents:

- Compte escollit per a l'estudi.
- Tipus d'informació que es pot introduir al Facebook. Podeu seguir la guia orientativa utilitzada en la primera activitat.
- Eines disponibles al Facebook per a protegir la privadesa.
- Aplicacions que ens proporciona el Facebook o que hi ha disponibles al Facebook i que es poden utilitzar per a obtenir informació d'altres usuaris.

b) Topologia d'amistats de la vostra xarxa social. Com a mínim us demanem que mostreu en el graf tots els nodes que hi ha a distància 2 del compte de l'usuari escollit. La secció ha de contenir la informació següent:

- Distància escollida.
- Representació del graf.
- Si desenvolupeu una eina per a obtenir aquesta informació, heu de descriure el funcionament d'aquesta eina. El desenvolupament d'aquesta eina és opcional.

c) Privadesa dels nodes. Per a cada node del graf estudiarem el nivell de privadesa que té. La memòria ha d'incloure la informació següent:

- Per a cada tipus d'informació que s'ha definit en el primer punt heu d'indicar quants nodes la mostren de manera pública. Aquesta informació la podeu presentar de manera agregada (tant per cent).
- Informació obtinguda de manera implícita. Potser un usuari no aporta informació d'una certa classe (per exemple, opció política), però es pot deduir per comentaris o fotos que sí que són mostrats de manera pública.

3. En aquesta activitat farem una anàlisi de les polítiques de privadesa de Microsoft similar a la proposada en el subapartat 3.2.2, en el qual hem parlat de les polítiques del Google.

L'estudi es divideix en una sèrie de fases que detallem a continuació; aquestes fases són els punts de la memòria que heu de lliurar:

- Analitzeu quines mesures utilitza Microsoft per a identificar els usuaris i obtenir-ne les dades.
- Esbrineu quin tipus d'informació personal recull Microsoft.
- Estudieu de quina manera utilitza la informació recollida Microsoft.
- Esbrineu quines polítiques de seguretat s'apliquen a les dades emmagatzemades.

Glossari

adreça IP *f* Codi numèric que identifica un ordinador específic a Internet. Les adreces d'Internet són assignades per un organisme anomenat *InterNIC*.

broadcast *m* Vegeu **difusió**.

CGI *f* Vegeu **interfície comuna de passarel·la**.

common gateway interface *f* Vegeu **interfície comuna de passarel·la**.

cookie *f* Vegeu **galeta**.

criptografia asimètrica *f* Sistema d'encriptació que consisteix a utilitzar un sistema de doble clau: clau pública i clau privada. La clau pública és coneguda per tothom i s'utilitza per a convertir el text en clar que volem encriptar en un criptograma, que tan sols es podrà tornar a convertir en text en clar mitjançant la clau privada, coneguda solament per la persona a la qual va remesa la informació encriptada mitjançant la clau pública.
sin. **criptografia de clau pública**

criptografia de clau pública *f* Vegeu **criptografia asimètrica**.

criptografia simètrica *f* Classe de criptografia que utilitza una única clau per a encriptar i desencriptar la informació. Com que solament hi ha una clau per a convertir el text en clar en criptograma i viceversa, aquesta clau ha de ser coneguda per les dues parts que volen intercanviar la informació.

detector *m* Aplicació de monitoratge i d'anàlisi per al trànsit d'una xarxa per a detectar problemes buscant cadenes alfanumèriques o de caràcters en els paquets. Es pot usar il·legalment per a rebre dades privades en una xarxa i, a més, és difícil de detectar.
en sniffer

difusió *f* Modalitat de transmissió que preveu l'enviament d'un missatge a tots els ordinadors connectats a una mateixa xarxa.

galeta *f* Fitxer que s'envia a un navegador per mitjà d'un servidor web per a registrar les activitats d'un usuari en un lloc web.
en cookie

GPS *m* Vegeu **sistema de posicionament global**.

HTML *m* Vegeu **llenguatge d'etiquetatge d'hipertext**.

HTTP *m* Vegeu **protocol de transferència d'hipertext**.

identificació per radiofreqüència *f* Tecnologia que permet identificar un objecte per ràdio, mitjançant una etiqueta (*RFID tag*) que aquest objecte porta adherida o inserida.
en radio frequency identification
sigla **RFID**

interfície comuna de passarel·la *f* Programa d'interfície que permet al servidor d'Internet utilitzar programes externs per a fer una funció específica. Executa un programa i formata els resultats en HTML de manera que puguin ser presentats en el navegador. Els *scripts* d'interfície comuna de passarel·la també s'usen per a introduir una varietat de sistemes d'anàlisi i de trànsit de mesura d'audiència d'un lloc.
en common gateway interface
sigla **CGI**

llenguatge d'etiquetatge d'hipertext *m* Llenguatge informàtic utilitzat per a crear documents d'hipertext que fa servir una llista finita d'etiquetes que descriu l'estructura general de diversos tipus de documents enllaçats entre si al Web.

localitzador uniforme de recursos *m* Adreça d'un lloc o d'una font, normalment un directori o un fitxer, al Web i la convenció que utilitzen els navegadors per a trobar fitxers i altres serveis distants.
sigla **URL**

malware *m* Vegeu **programa maliciós**.

navegador *m* Programa informàtic que permet veure diversos tipus de recursos d'Internet disponibles al Web i interactuar-hi.

programa espia *m* Programa que recopila informació d'un ordinador i després transmet aquesta informació a una entitat externa sense el coneixement o el consentiment del propietari de l'ordinador.
en spyware

programa maliciós *m* Programa, document o missatge susceptible de causar perjudicis als usuaris de sistemes informàtics.
en malware

protocol de control de transmissió / protocol d'Internet *m* Conjunt d'instruccions que dicten com s'han d'enviar paquets d'informació per diferents xarxes. També té una funció de verificació d'errors per a assegurar-se que els paquets arriben a la destinació en l'ordre apropiat.
sigla **TCP/IP**

protocol de seguretat de nivell de transport *m* Protocol criptogràfic –successor del protocol de capa de sòcol segur– que proporciona comunicacions segures per una xarxa, freqüentment Internet.
sigla **TLS**

protocol de transferència d'hipertext *m* Mètode utilitzat per a transferir fitxers d'hipertext per Internet. Al Web, les pàgines escrites en aquest protocol utilitzen l'hipertext per a enllaçar amb altres documents.

radio frequency identification *f* Vegeu **identificació per radiofreqüència**.

RFID *f* Vegeu **identificació per radiofreqüència**.

sistema de posicionament global *m* Sistema global de navegació per satèl·lit que ens permet fixar a escala mundial la posició d'un objecte, una persona, un vehicle o una nau.

sniffer *m* Vegeu **detector**.

spyware *m* Vegeu **programa espia**.

TCP/IP *m* Vegeu **protocol de control de transmissió / protocol d'Internet**.

TLS *m* Vegeu **protocol de seguretat de nivell de transport**.

URL *m* Vegeu **localitzador uniforme de recursos**.

virus *m* Programa creat especialment per a envair ordinadors i xarxes i amb intenció destrutiva. El dany pot ser mínim, com per exemple que surti una imatge o un missatge a la pantalla, o pot fer molt dany alterant o fins i tot destruint fitxers.

Web *m* Conjunt de fitxers de text i multimèdia i altres serveis connectats entre si per mitjà d'un sistema de documents d'hipertext.

Bibliografia

Asur, S.; Huberman, B. A. (2010). "Predicting the Future with Social Media". A: *Proceedings of the 2010 International Conference on Web Intelligence and Intelligent Agent Technology* (pàg. 492-499).

Banks, L.; Wu, S. F. (2009). "All Friends are NOT Created Equal: An Interaction Intensity based Approach to Privacy in Online Social Networks". A:*Proceedings of the 2009 International Conference on Computational Science and Engineering* (pàg. 970-974).

Baran, P. (1964). "On distributed communications: IX security secrecy and tamper-free considerations".

Berners-Lee, T.; Hendler, J.; Lassila, O. (2001). "The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". *Scientific American* (vol. 284, núm. 5, pàg. 34-43).

Castellà-Roca, J.; Viejo, A.; Herrera-Joancomartí, J. (2009). "Preserving user's privacy in web search engines". *Computer Communications* (vol. 32, núm. 13-14, pàg. 1541-1551).

Chaum, D. (1981). "Untraceable electronic mail, return addresses, and digital pseudonyms". *Communications of the ACM* (vol. 4, núm. 2, pàg. 84-88).

"Condiciones de servicio de Google" (2011). Google. <http://www.google.com/accounts/TOS?hl=ES>

Consumer Reports National Research Center. "Annual State of the Net Survey" (2010). *Consumer Reports* (vol. 75, núm. 6).

Coutu, D. i altres (2007, juny). "We Googled You". *Harvard Business Review* (vol. 85, núm. 6).

Danezis G.; Diaz, C.; Syverson, P. (2010). "Systems for Anonymous Communication". *CRC Handbook of Financial Cryptography and Security* (pàg. 341-390).

Danezis, G.; Dingledine, R.; Mathewso, N. (2003). "Mixminion: Design of a type III anonymous remailer protocol". A: *Proceedings, IEE Symposium on Security and Privacy* (pàg. 2-15).

Diffie, W.; Oorschot, P. C. van; Wiener, M. J. (1992). "Authentication and authenticated key exchanges". *Designs, Codes and Cryptography* (vol. 2, pàg. 107-125).

Dingledine, R.; Mathewson, N.; Syverson, P. (2004). "Tor: The second generation onion router". A: *Proceedings of the 13th USENIX SecuritySymposium* (pàg. 303-319).

Domingo-Ferrer, J.; Sebé, F.; Solanas, A. (2008). "An Anonymity Model Achievable Via Microaggregation". *Lectures Notes in Computer Science* (vol. 5159, pàg. 209-218).

Eckersley, P. (2010). "How Unique Is Your Browser?". A: *Proceedings of the Privacy Enhancing Technologies Symposium*.

Glassman, N. (2010, agost). "3 Questions About Location-Based Social Networks with face2face". *SocialTimes*.

Goldschlag, D. M.; Reed, M. G.; Syverson, P. F. (1996). "Hiding routing information". *Lecture Notes in Computer Science* (vol. 1174, pàg. 137-150).

Graffi, K. i altres (2009). "Practical security in p2p-based social networks". A: *Proceedings of the IEEE 34th Conference on Local Computer Networks* (pàg. 269-272).

Greenberg, Andy (2009, 13 de gener). "Privacy Groups Target Android, Mobile Marketers". *Forbes*. <http://www.forbes.com/2009/01/12/mobile-marketing-privacy-tech-security-cx_ag_0113mobilemarket.html>

Hansell, S. (2006, febrer). "Increasingly, Internet's Data Trail Leads to Court". *The New York Times*.

Hay, M. i altres (2008). "Resisting Structural Identification in Anonymized Social Networks". A: *Proceedings of the 2008 Conference on Very Large Databases*.

Krishnamurthy, B.; Wills, C. E. (2010). "On the leakage of personally identifiable information via online social networks". *ACM SIGCOMM Computer Communication Review* (vol. 40, núm. 1, pàg. 112-117).

Laurant, C. (2003). "Privacy & Human Rights 2003 - An international survey of privacy laws and developments". Electronic Privacy Information Center.

Leskovec, J.; Adamic, L. A.; Huberman, B. A. (2006). "The dynamics of viral marketing". A: *Proceedings of the 7th ACM Conference on Electronic Commerce*.

Li, S. (2011, gener). "Insurers are scouring social media for evidence of fraud". *Los Angeles Times*.

Madden, M. i altres (2007). "Digital Footprints: Online identity management and search in the age of transparency". <http://pewresearch.org/pubs/663/digital-footprints>

Mazieres, D.; Frans Kaashoek, M. (1998). "The Design, Implementation and Operation of an Email Pseudonym Server". A: *5th ACM Conference on Computer and Communications Security* (pàg. 27-36).

Melenhorst, M.; Grootveld, M.; Setten, M. van (2008). "Tag-based information retrieval of video content". A: *Proceeding of the 1st International Conference on Designing Interactive User Experiences for TV and Video*.

Milton, N. (2010, maig). "Price of Facebook Privacy? Start Clicking". *The New York Times*.

Möller, U. i altres (2003). "Mixmaster protocol (version 3)". IETF Internet Draft.

Parekh, S. (1996, 5 d'agost). "Prospects for remailers: where is anonymity heading on the internet?". *Online Journal* (vol. 1, núm. 2). <http://freehaven.net/anonbib/cache/remailer-history.html>

Pattison, K. (2010, octubre). "Geolocation Services: Find a Smartphone, Find a Customer". *The New York Times*.

PhoCusWright (2011). "PhoCusWright's Social Media in Travel: Traffic & Activity".

"Políticas de privacidad de Google" (2011). Centro de Privacidad de Google. <http://www.google.es/intl/es/privacy/>

Raymond, J. F. (2000). "Traffic analysis: Protocols, attacks, design issues, and open problems". *Lecture Notes in Computer Science* (vol. 2009, pàg. 10-29).

Reiter, M.; Rubin, A. (1998). "Crowds: Anonymity for web transactions". *ACM Transactions on Information and System Security* (vol.1, núm. 1, pàg. 66-92).

Rotenberg, M. (2000). "Protecting Human Dignity in the Digital Age". Unesco.

Saint-Jean, F. i altres (2007). "Private web search". A: *Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society* (pàg. 84-90).

Sánchez, D.; Isern, D.; Millán, M. (2010). "Content Annotation for the Semantic Web: an AutomaticWeb-based Approach". *Knowledge and Information Systems*. En premsa.

"7 things to stop doing now on Facebook" (2010). *Consumer Reports* (vol. 75, núm. 6).

Singel, R. (agost del 2009). "You deleted your cookies?". *Wired*.

"Social savvy recruiters utilising social media in their recruitment strategy" (2010, maig). *Personnel Today*.

Soderberg, J.; Kakogianni, E. (2010). "Automatic tag generation for photos using contextual information and description logics". A: *2010 International Workshop on Content-based Multimedia Indexing* (pàg. 1-7).

Stagg-Macey, C.; Beattie, C. (2010, abril). "Leveraging social networks: an in-depth view for insurers". *Document*.

Terry, J. (2008, 7 de febrer). "Leaving a digital footprint: Online activities follow students to job interviews, professional world". *The State News*.

"The Cross-Site Scripting (XSS) FAQ". <http://www.cgisecurity.com/xss-faq.html>

Wortham, J. (maig del 2010). "Facebook Glitch Brings New Privacy Worries". *The New York Times*.

"Your digital footprint: how much information about your life gets recorded by big business and Big Brother" (2011). *Koppel on Discovery Channel*. <http://dsc.discovery.com/convergence/koppel/interactive/interactive.html>