Il problema della privatezza

Memorizzazione e processamento di una grande quantità di dati

Informazione disponibile sulla rete accessibile da qualsiasi locazione e in qualsiasi momento

Tecniche efficienti per raccogliere informazioni e creare nuove basi di dati

Facilità di raccolta, scambio e trasmissione di informazione

Facilità di accesso e collegamento fra informazioni

$\Longrightarrow$ Perdita di controllo su:

- Cosa è raccolto
- Come è utilizzato
- Eventuali rilasci successivi

Mentre prima dovevamo fidarci di un insieme ristretto di persone dobbiamo ora fidarci di una intera rete.

Il problema della privatezza – 2

Località, frammentazione e difficoltà di accesso ai dati hanno rappresentato in passato una forma di protezione che oggi non esiste più

Es. dati medici: il medico deve assicurare la confidenzalità di ciò che vede o viene a conoscere riguardo ai pazienti (giuramento di Ippocrate)

- Non c'è più un medico ..... c'è un sistema complesso
  - medico curante
  - medico specialista
  - ospedale
  - laboratorio di analisi
  - farmacista
  - datore di lavoro
  - assicurazione .....

Il problema della privatezza – 2

Il record medico elettronico rappresenta informazioni sempre più complete (decine di campi) e si va sempre più estendendo (supporto e collegamento a immagini - es. raggi X)

Anche all'interno di una singola struttura ospedaliera diverse persone hanno accesso ai dati relativi ai pazienti

Diverse informazioni mediche vengono memorizzate in diversi sistemi e utilizzate per diversi scopi (compagnie di assicurazione, aziende farmaceutiche, centri di ricerca, scuole, datori di lavoro, ...)

Dati medici vengono raccolti e aggregati

- necessario e vitale per la ricerca
- può compromettere la privatezza dei pazienti

Il problema della privatezza

Gli abusi possono essere molti. Alcuni esempi da recenti studi negli USA:

- il 40% delle società di assicurazione rilascia dati medici a terze parti (datori di lavoro, società finanziarie) senza il consenso del cliente
- più della metà delle 500 maggiori aziende statunitensi ha ammesso di utilizzare informazioni mediche nelle decisioni di assunzione e in altre decisioni sul personale.
- un funzionario di banca, in servizio in un comitato statale relativo alla sanità, ebbe accesso alla lista dei pazienti diagnosticati di tumore. Collegò i dati alla lista dei suoi clienti e annullò i prestiti.
- una industria farmaceutica ha rilevato una compagnia di servizi sanitari acquisendo accesso a un database relativo alle prescrizioni di 56 milioni di persone. Ha poi contattato i medici curanti dei pazienti con particolari sintomi per convincerli a prescrivere un antidepressivo da loro commercializzato.

Garante della Privacy e Legge 675/96

In Italia, il Garante della Privacy regolamenta la gestione dei dati personali.

La legge 675/96 regola raccolta, mantenimento, e divulgazione di informazioni personali:

- Richiede consenso dell'interessato riguardo alla raccolta e al trattamento dei dati personali nonchè alla loro diffusione.
- Richiede a chi raccoglie i dato di applicare misure di sicurezza.
- **Art. 22, comma 1. – Dati sensibili**: I dati personali idonei a rivelare l'origine razziale ed etnica, le convinzioni religiose, filosofiche o di altro genere, le opinioni politiche, l'adesione a partiti, sindacati, associazioni o organizzazioni a carattere religioso, filosofico, politico o sindacale, nonché i dati personali idonei a rivelare lo stato di salute e la vita sessuale, possono essere oggetto di trattamento solo con il consenso scritto dell'interessato e previa autorizzazione del Garante.

Raccolta di informazione

Gestori di siti Web spesso raccolgono informazioni personali sugli utenti stessi, attraverso pagine di registrazione, survey form, ordini, competizioni online, e utilizzando software in modo a volte non noto agli utenti.

Queste informazioni sono spesso passate a terze parti all'insaputa dell'utente.

In alcuni casi sappiamo che dati su di noi sono raccolti .....

... ma non abbiamo alcun controllo sul suo uso e disseminazione

Spesso non sappiamo neppure che dati che ci riguardano sono raccolti:

- Microsoft passport in Windows XP
- Link referenziali
- File di log
- Cookies

Link referenziali

Quando accediamo ad una pagina Web, il client può passare al server contattato l'URL della pagina dalla quale siamo partiti (refer link) e dalla quale abbiamo seguito il link.

Ragioni:

- controllare l'efficacia di banner pubblicitari
- tracciare come gli utenti si muovono lungo un sito

Link referenziali

.... il refer link rivela però informazioni private

- Motori di ricerca (es., Altavista, Lycos, ...) incorporano la query dell'utente nell'URL. Questa informazione è quindi comunicata con il refer link.
- URL passate da un sito ad un altro utilizzando protocolli crittografici (es., SSL) possono essere mandati ad un sito successivo in chiaro (attraverso un link non crittato).
- A volte anche numeri di carte di credito (passati con compilazioni di form sul Web) sono incorporate nell'URL.

File di log

Ogni server Web mantiene un log degli accessi ai file (pagine) gestite. Il file di log è sotto il controllo del server, che registra:

- Nome/indirizzo IP della macchina che ha fatto la connessione
- Tempo della richiesta
- URL richiesta
- Tempo necessario per scaricare il file
- Username della persona che ha scaricato il file (se è stata usata autenticazione HTTP)
- Eventuali errori originati
- Il tipo di browser utilizzato
- La precedente pagina che era stata scaricata dal browser (refer link)

File di log – 2

I log del Web possono essere combinati con altri file di log (es. informazione di login/logout mantenuta da un provider, o log di mail server) per risalire all'identità di chi ha acceduto la pagina.

Generalmente, ma non sempre, questo tipo di correlazione richiede la collaborazione di altre organizzazioni.

I log possono essere confusi attraverso l'utilizzo di

- server proxy. Quando un utente accede ad una pagina web attraverso un indirizzo proxy, anzichè l'indirizzo della macchina utente, viene passato al server l'indirizzo del proxy.
- anonymous surfers funzionano da proxy per gli utenti che vogliono mantenere anonimi i loro accessi web (es., www.anonymizer.com, www.freedom.net, www.the-cloak.com,....).

Anonymous surfers

1. L'utente manda la URL richiesta all'anonimizzatore, che sottopone poi la richiesta direttamente.
2. Il server di destinazione vede quindi le richieste come provenienti dall'anonimizzatore.
3. Il server manda la propria risposta all'anonimizzatore, che la passa poi all'utente.

Anche con i proxy non si ha vero anonimato:

- I gestori degli anonimizzatori hanno tutte le informazioni sugli accessi che i loro utenti effettuano....
- Chi utilizza proxy si deve quindi fidare del proxy sul fatto che questi non passerà informazioni ad altri.

Cookie

Un cookie è un pezzo di testo ASCII che un server Web passa a una istanza utente del browser. Una volta ricevuto il cookie, il browser lo rispedirà con ogni richiesta successiva.

Lo scopo originale dei cookie (introdotti con Netscape 2.0) era di dare ad un server Web la possibilità di riconoscere uno stesso cliente che sottopone più richieste HTTP. Il riconoscimento del cliente era utilizzato poi nelle applicazioni.

Ad esempio, un catalogo on line può memorizzare la sessione ID nel cookie così che il server possa tenere traccia di quali articoli ci sono nel "carrello" del cliente.

Cookie

I cookie sono gestiti nella memoria del browser e, se persistenti, sono memorizzati dal browser.

- .netscape/cookies (Unix)
- Preferences/Netscape/MagicCookie (Mac)
- ../windows/cookies (Windows)

Cookie persistenti possono essere utilizzati per memorizzare preferenze dell'utente (es., colore della schermata) così che l'utente non debba rispecificarle ogni volta.

I cookie sono stati utilizzati per tracciare i movimenti degli utenti sulla rete e esaminare come gli utenti si muovono da sito a sito. Viene quindi compromesso l'anonimato.

Cookie – 2

Inizialmente i cookie originati da un sito potevano essere passati ad altri siti.

In un secondo tempo Netscape ha modificato i suoi browser in modo che cookie possano essere passati solo al sito che li ha originati.

Gli sviluppatori Web hanno comunque trovato un modo per raggirare l'ostacolo aggiungendo cookie a immagini GIF che erano caricate su siti di terze parti.

Doubleclick Network, una azienda di pubblicità via Internet, è stata fra le prima ad utilizzare i cookie per correlare le attività degli utenti attraverso differenti siti web.

Cookie – 3

Doubleclick pagava siti Web per inserire un “<IMG SRC> tag” sulla pagina HTML del sito che causa il caricamente di un file GIF e un cookie.

Doubleclick sosteneva in questo modo di tenere traccia di quali utenti Internet hanno visto quale avviso pubblicitario in modo da non presentare lo stesso avviso due volte ad uno stesso utente (a meno che colui di cui viene fatta pubblicità paghi allo scopo).

Avvisi pubblicitari di Doubleclick erano anche presenti sul sito di Altavista, permettendo quindi a Doubleclick di mantenere traccia di tutte le richieste di ricerca fatte dagli utenti ad Altavista.

Doubleclick ha annunciato il 31 Dicembre 2001 la decisione di non continuare l’attività di profiling degli utenti (nel Gennaio 2001 era stato reso noto il suo piano di correlare i profili anonimi raccolti a informazione di identità).

Disabilitare i cookie

Netscape e Explorer hanno opzioni che permettono agli utenti di specificare quando i cookie possono essere ricevuti. Possiamo specificare che non accettiamo cookie o che vogliamo che la loro ricezione sia comunicata con la possibilità di decidere se accettarli o rifiutarli.

Nessun browser oggi permette di disabilitare la spedizione di cookie gia ricevuti (possiamo però operare a basso livello sul file) o di rifiutare cookie da alcuni siti ma accettarli da altri.

Ci sono servizi di cookie buster che operano da proxy filtrando cookie e banner pubblicitari.

Disabilitare i cookie – 2

Operando direttamente a livello di file system possiamo però eliminare i cookie:

- Ad esempio in Unix possiamo cancellare il file dei cookie e sostituirlo con un link a “/dev/null”. In Windows possiamo sostituire il file con un file vuoto su cui non sono possibili nè letture nè scritture.
- Possiamo accettare i cookie che vogliamo e poi dichiarare il file “read-only”, così che nessun cookie possa più essere accettato.
- Modificando direttamente il codice del browser.

Alcune funzioni (es., gestione “basket” in acquisti on-line) possono non essere possibili senza cookie.

Cookie per la protezione della privacy

I cookie violano la privacy quando sono utilizzati per tracciare le attività degli utenti nel Web, creando una mappa elettronica che segnala i percorsi di ogni singolo untente.

Usati propriamente i cookie possono in realtà aumentare la privacy. Es., permettono personalizzazione di interfacce Web senza il bisogno di raccogliere e mantenere informazioni personali sugli utenti.

Alcuni esempi di violazioni alla privacy...

- Alcuni programmi software che implementano servizi del tipo “shopping cart” permettevano agli utenti di vedere il contenuto dei “carrelli” di altri utenti
- AOL ha rilasciato informazioni sui propri iscritti a loro insaputa, in violazione dell’ “Electronic Communications Privacy Act” e in violazione all’agreement con i propri iscritti.
- Liberty Financial Companies, Inc., affermava di mantenere informazioni raccolte da bambini, comprendenti anche stato economico della famiglia, in forma anonima. Le informazioni erano invece mantenute in forma completamente identificabile.
- Internet Explorer 5.0 informava siti web quando utenti mettevano un bookmark sulle loro pagine a insaputa degli utenti.
- Intel ha prodotto un chip Pentium III contenente un unico numero seriale che permetteva di rintracciare transazioni degli utenti in rete.

Raccolta di dati sul Web

Primi studi ufficiali che riportavano la mancanza di privacy nel Web fatti dall’EPIC (Electronic Privacy Information Center) nel 1997 (Surfers Beware) e 1998 (Surfers Beware II: Notice is not enough) e 1999 (Surfers Beware III: Privacy Policies Without Privacy Protection).

Controlli sono anche stati effettuati dalla Federal Trade Commission (USA).

Recentemente diversi survey sono stati svolti su “policy practice” da parte di siti Web.

Dai diversi dati si evince che la applicazione di misure di privacy sta crescendo (si va verso una autoregolamentazione?) ma non è ancora soddisfacente.

Raccolta di dati sul Web

Dal rapporto della FTC del 1998 su un survey su siti Web commerciali:

- 92% raccoglieva informazioni sui consumatori, ma....
- solo il 14% riportava sul sito la pratica di gestione delle informazioni
- solo il 2% riportava una politica di privacy completa.

I dati sui siti più visitati erano più positivi:

- 97% raccoglieva informazioni personali
- 71% riportava una politica di gestione delle informazioni
- 44% riportava una politica di privacy completa

Per quanto riguarda i siti rivolti ai bambini

- 89% raccoglieva informazioni personali; 24% riportava una privacy policy
- solo l' 1% richiedeva il consenso dei genitori per raccolta/divulgazione

Raccolta di dati sul Web

Un più recente studio sui 162 siti più popolari per utenti Internet sotto i 13 anni mostrava che

- solo 114 siti riportavano una privacy policy
- dei 114 che avevano una privacy policy 90 siti raccoglievano informazioni personali dai ragazzi
- 14 siti raccoglievano informazioni personali ma non riportavano alcuna privacy policy
- solo il 55% delle privacy policy affermava di non raccogliere più dati di quanto necessario
- solo il 62% permetteva ai genitori di vedere i dati raccolti dai loro figli

Raccolta di dati sul Web

Da uno studio recente su 1001 genitori e 304 bambini:

- i ragazzi più grandi (13-17 anni) sono molto più disponibili a rilasciare informazione di quelli più piccoli (10-12 anni)
- il 57% dei ragazzi ed il 37% delle ragazze rilascerebbero il tipo di macchina dei genitori
- il 65% dei ragazzi rilascerebbe il nome del negozio preferito
- il 57% rilascerebbe il nome del negozio preferito dai genitori
- il 41% dei ragazzi più grandi e il 27% dei ragazzi più piccoli direbbe la mancia settimanale che riceve
- il 39% dei ragazzi più grandi fornirebbe informazioni private sulla famiglia
- il 16% dei ragazzi più piccoli ha dichiarato di aver rilasciato informazioni private sulla famiglia

Raccolta di dati sul Web

Nello studio del 1999 l'Electronic Privacy Information center riporta che

- Nessuno dei 100 siti più utilizzato per il commercio elettronico soddisfa i requisiti del FTC.

Un ulteriore survey condotto da Georgetown University afferma che meno del 10% dei siti visitati soddisfa i requisiti della FTC.

Da un recente survey della FTC su un campione di 300 siti web che colleziona informazioni personali risulta che solo il 20% soddisfa i quattro requisiti richiesti di notifica, consenso, accesso e sicurezza

Aspetti della privacy

Nel 1998 la FTC identificava cinque punti principali:

- Notifica: Gli utenti devono essere informati della politica adottata per la gestione dei dati personali prima che i dati siano raccolti.
- Scelta/Consenso: Gli utenti devono poter decidere se e come informazioni personali che li riguardano possono essere utilizzate.
- Accesso/partecipazione: Gli utenti devono avere accesso alle informazioni raccolte su di loro e devono essere in grado di contestare l'accuratezza o la completezza delle stesse.
- Integrità/Sicurezza: Devono garantire misure di protezione che garantiscano che utenti non autorizzati non abbiano accesso alle informazioni.
- Implementazione: Meccanismi di autoregolazione o governativi dovrebbero imporre sanzioni per chi non rispetta i principi.

Seal program

Un meccanismo emergente di autoregolamentazione è rappresentato dai "seal program" (es., TRUSTe, BBBOnLine, WebTrust).

Chi sottoscrive un seal program dichiara di obbedire un determinato codice riguardo alla gestione di informazioni e deve sottoporsi a controlli di adeguatezza allo scopo di mostrare poi sul sito Web un "privacy seal".

I criteri di adeguatezza includono:

- controlli su raccolta e uso di informazioni "personally identifiable"
- controlli per rintracciare identificatori nella base di dati e verificare, ad esempio, che richieste di cancellazione dagli utenti siano state rispettate.

TRUSTe include anche monitoraggio da parte di terzi e verifiche periodiche sui database gestiti dai sottoscrittori.

Seal program

Costituiscono un modo conveniente per

- le aziende: per dimostrare che soddisfano certi requisiti
- gli utenti: per identificare i siti Web che seguono specifici principi di gestione delle informazioni.

.... però.....

recentemente Geocities (membro del programma TRUSTe), un provider per pagine web gratuite, ha rilasciato a terze parti informazioni circa i propri iscritti, molti dei quali bambini.

Platform for Privacy Preferences (P3P)

Proposta del World-Wide Web Consitium (W3C) di un meccanismo per regolare scambio di informazioni nel Web.

Permette a siti Web di comunicare la pratica di privacy seguita e agli utenti di specificare richieste di privacy su informazioni che li riguardano.

Lo scopo del P3P è di permettere agli utenti di specificare preferenze riguardo alla gestione delle informazioni personali che li riguardano.

Dopo aver configurato le proprie preferenze, gli utenti possono navigare nella rete senza più preoccuparsi circa il rilascio di informazione.

Il browser con P3P al momento del contatto con il server riceve una “privacy policy proposal” dal server che specifica quali dati il server richiede e come verranno gestiti.

Se la proposta del server è in accordo con le preferenze dell’utente le informazioni vengono rilasciate e l’accesso al server consentito.

Platform for Privacy Preferences (P3P) – 2

P3P propone un “vocabolario” con cui i siti Web possono comunicare la pratica seguita nella gestione dei dati.

Distingue

- categorie di dati
- scopo di utilizzo
- riceventi

P3P – categorie di dati

Le categorie di dati includono:

- informazione fisica di contatto (es., indirizzo postale e numero di telefono)
- Informazione di contatto online (e-mail)
- Identificatori (es. carta di identità)
- Identificatori finanziari che collegano un individuo a uno strumento finanziario (es., carta di credito)
- Informazioni su computer (es. browser e IP)
- Dati navigazionali (es., siti visitati)
- Dati demografici (es., età, data di nascita, sesso, stipendio).
- Contenuto (es. contenuto di messaggi e-mail).

P3P – scopi

Lo scopo permette di specificare se i dati possono essere utilizzati per

- completare l'attività (0)
- scopi di amministrazione del sito (1)
- personalizzazione del sito all'utente (2)
- ricerca e sviluppo (5)
- contattare l'utente per offrire servizi e prodotti
- altri scopi

P3P – riceventi

Permette di specificare preferenze su rilasci successivi, comprende quattro categorie:

- il sito e i suoi agenti (agente è una terza parte che processa dati per conto del sito).
- organizzazioni che seguono le stesse "privacy practice" del sito
- organizzazioni che seguono "privacy practice" differenti
- altre terze parti o pubblico rilascio

Platform for Privacy Preferences (P3P)

Ha incontrato obiezioni:

- La comunicazione in realtà è attraverso “agenti” che operano per mezzo dell’utente, quindi l’utente può non essere informato (contrario a “informed consent”)
- Siti Web potrebbero impropriamente utilizzare il vocabolario P3P per acquisire informazioni dall’utente
- Raramente gli utenti cambieranno le specifiche date di “default” quindi potrebbero rilasciare informazione inavvertitamente

Protezione della privacy

Il problema della privacy è complesso e richiede la applicazione non soltanto di misure tecniche ma anche legislative e, in generale etiche.

- Leggi e “public policies”
- Misure tecniche
- Politiche e pratiche di organizzazioni e individui (autoregolamentazione)

È importante notare che privacy sulla rete molto spesso è in conflitto con il requisito di “accountability”. Esistono quindi pro e contro che vanno considerati.