"Lei sa già chi sono io"

Privacy nell' era della riproducibilità digitale dei dati personali

Alessandro Barenghi

barenghi - at - elet.polimi.it

20 gennaio 2012

Introduzione

Scenario

- Security $\neq$ privacy!
  - Security: "The state of being free from danger or threat" [a]
  - Privacy: "A state in which one is not observed or disturbed by other people" [b]
- Garantire entrambe richiede approcci diversi, anche se integrabili
- La prima è un punto studiato da tempo, la seconda è una preoccupazione più giovane

[a] The Oxford Dictionary, 2011
[b] The Oxford Dictionary, 2011

Problem?

What's this privacy stuff anyway?

- I dati personali possono essere diffusi per varie ragioni :
  - Volontaria: "Mi interessa che il mondo lo sappia" (eg. CV)
  - Involontaria: "Oh beh, ma tanto non lo legge nessuno"
  - Forzata: se i muri hanno orecchie, figuriamoci i cavi....
- Il primo punto non è tecnicamente un problema, ma gli altri due...
- Disclaimer: non sono un sociologo/psicologo quindi il resto del talk sarà il più tecnico possibile :)

????

Un megafono invisibile

- Gestire grandi quantità di dati in formato digitale è molto veloce
  - Un essere umano legge circa una pagina al minuto
  - Il vostro cellulare "legge" 200 pagine al secondo
- Riprodurre i dati produce una copia perfetta
- "The World Wide Web Never Forgets"
- Risultato: i dati si diffondono molto più di quello che è intuibile
- Esempio: un post su una bacheca di Facebook può essere letto in media da 130 persone

PROFIT!

Stakeholders

- Chi è interessato ai nostri dati ?
  1. stalkers: "il mio tessooro...."
  2. criminali: "Aah, identità fresche!"
  3. aziende: Data is money (Amazon, Facebook)
  4. enti governativi: Knowledge is power
- Nei primi 2 casi, interessarsi delle informazioni altrui è illegale: tutti sono interessati a evitarlo
- Negli altri 2 , è banalmente "poco gentile": non necessariamente tutti vogliono evitarlo

Questi dati? Oh, erano sparsi in giro...

Frammenti perfettamente significativi

- Non si dà molto peso a fornire piccole quantità di dati personali a enti diversi
- Fino al secolo scorso, ricostruire il profilo di una persona richiedeva giorni (mesi?) di indagini
- Ora sono presenti aggregatori di informazioni (e.g. 123people.com) in grado di farlo in minuti
- Il trucco: aggregare dati provenienti da molte fonti!

Quasi identificatori

I dettagli contano

- Aggregando piccoli quantitativi di informazioni è possibile effettuare deduzioni sull' identità di una persona
- Esempio :
  - Ho dati di statistiche mediche con solo zip code, sesso ed età del malato
  - Non conosco nome, nè indirizzo del malato
  - Ho accesso a una raccolta di dati anagrafici (es. rastrello dati da Facebook)
- Incrociando i dati anagrafici con le statistiche mediche riesco a identificare il malato!

Aggregazione di massa

Solve et coagula

- Cosa succede se la cosa viene fatta su scala industriale?
  - L' aggregazione su vasta scala è stata la sorgente di maggior profitto di Amazon (recommendation engine)
- Cosa succede se la cosa viene fatta su scala statale?
  - Iniziativa di IARPA[a]: predizione dei crimini da dati pubblici raccolti (ha funzionato su rapine in banca a Santa Clara)
- Nessuno di questi casi usa tecniche di intercettazione/spionaggio sono tutti dati pubblici!

[a] http://www.iarpa.gov

Forzare la mano

Origliare 2.0

- Forzare la privacy altrui è una pratica attiva da prima della digitalizzazione dei dati
- L' infrastruttura di Internet era stata concepita per condividere al meglio le informazioni
- Risultato: ci sono mezzi molto efficienti per fare spionaggio
- Wikileaks: ha pubblicato di recente "The Spy Files" raccogliendo documentazione su chi produce microspie digitali.[a]

[a] Anche in Italia abbiamo produttori.

Eavesdropping

the Cookie Monster

- Origliare è la forma di spionaggio con risorse minime
- "Se è talmente semplice, se ne saranno presi cura tutti, no?"
- No. HTTP e FTP passano tutte le informazioni come le forniamo...
- Quindi chiunque abbia accesso alla linea di comunicazione le può ascoltare!
- Esempio pratico: Firesheep automatizza la raccolta di cookies di sessione di Facebook :)

Encrypt the internet!

Soluzione

- Dal talk precedente, sappiamo che SSL è in grado di garantire la confidenzialità dei dati trasmessi
- Per anni SSL ha rappresentato un collo di bottiglia prestazionale, rallentando l' adozione
- Ottobre 2010: Google pubblica i dati di carico dei propri server facendo notare che SSL è $<1\%$
- "Encrypt the internet"! HTTPSEverywhere è un plugin di firefox che sfrutta HTTPS ovunque possibile[a]

[a] fornito da www.eff.org

Se telefonando...

Intercettazioni telefoniche

- Per anni, intercettare comunicazioni telefoniche implicava essere collegati al cavo
- ... ma ora le chiamate vengono trasmesse via IP sulle dorsali principali!
- Se si potesse inserirsi e fingersi una cella per cellulari, si finirebbe ad ascoltare tutto...
- Nel 2010 è stato mostrato come emulare completamente una cella con equipaggiamento reperibile comunemente
- Avvertenza:serve leggersi mezza specifica del GSM
- L' anno scorso Vodafone ha iniziato a vendere mini-celle da casa....

Soluzione

Cellulari cifrati?

- Esistono applicazioni in grado di effettuare chiamate cifrate
- Come evitare il man in the middle?
  - consegno la chiave a mano alla persona con cui voglio parlare
  - sistema a certificati simile a SSL
- Skype usa nativamente un sistema a certificati per cifrare le vostre chiamate e chat [a]

[a] chiaramente, Skype può decifrarle...

Too smart metering

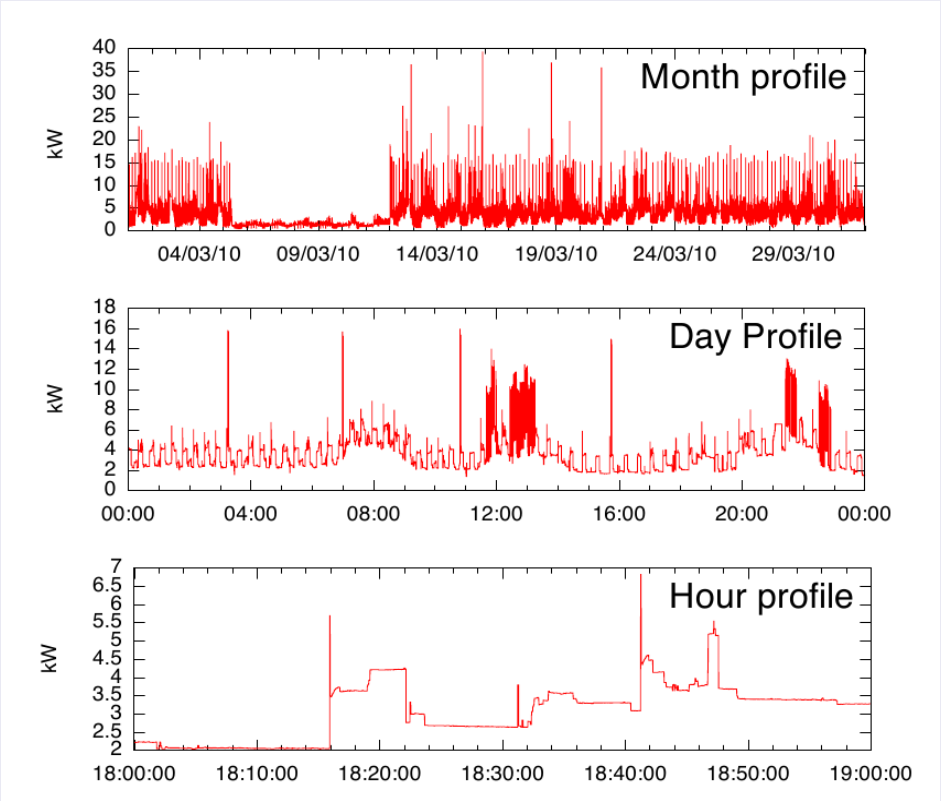
Quando il contatore diventa troppo intelligente

- La telelettura dei contatori dell' energia elettrica è realtà da un po'
- E' comoda^{TM}, precisa, e molto frequente....
- Talmente tanto frequente da rivelare le abitudini casalinghe!
- E' possibile identificare gli elettrodomestici dai picchi di consumo
- Alcuni contatori trasmettono queste informazioni in chiaro sulla rete elettrica [a]

[a] Carluccio et al. al 28th CCC

Too smart metering

Tipico consumo da casa



S. McLaughlin, P. McDaniel, and W. Aiello. *Protecting consumer privacy from electric load monitoring* (CCS '11)

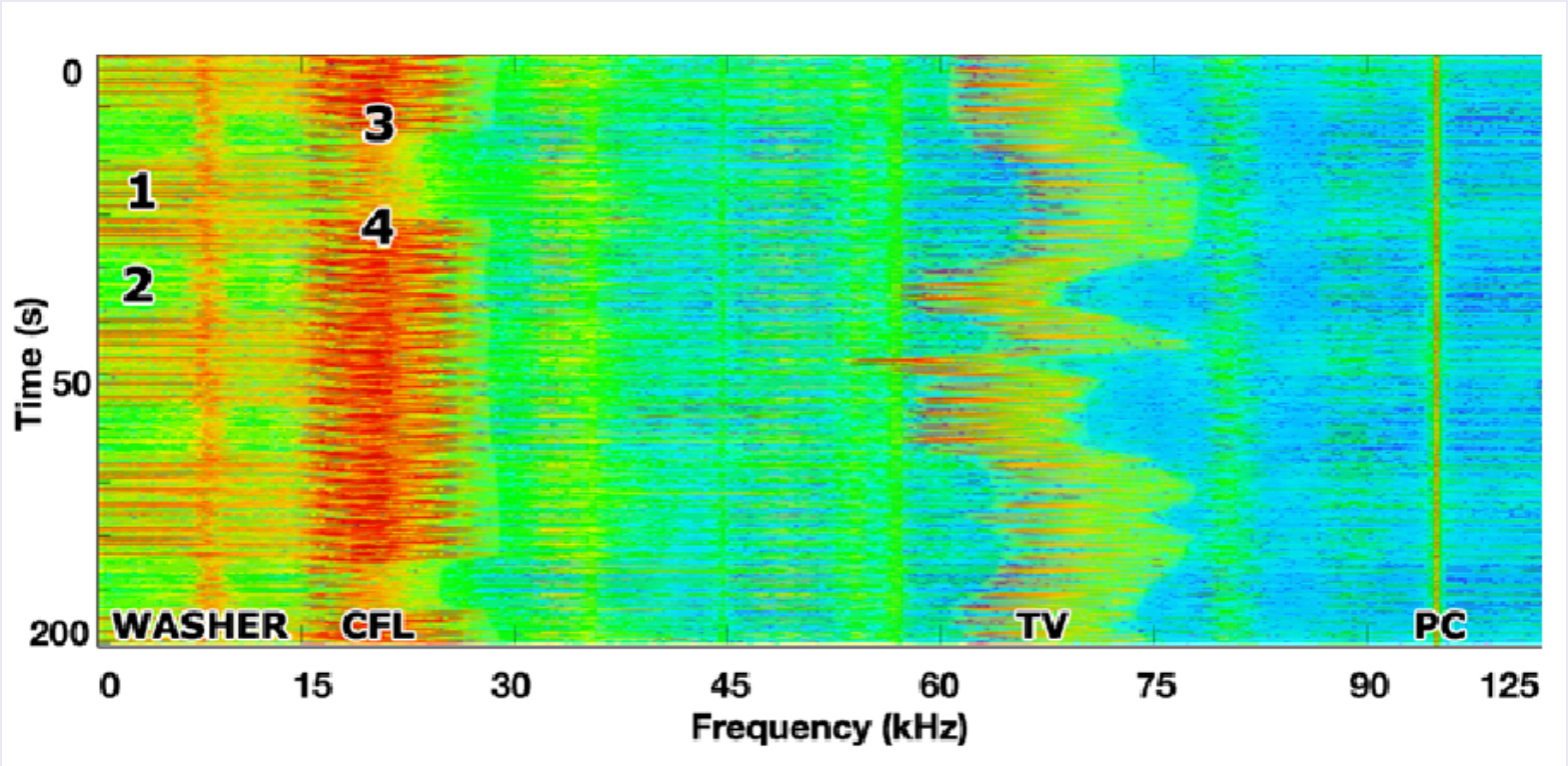
Auditel on steroids

Sempre più difficile....

- E' facile identificare elettrodomestici e dedurre abitudini
- Qual è il passo successivo?
  - Gli alimentatori switched mode "segnalano involontariamente" sul cavo quanto consumano
  - Buona parte degli elettrodomestici ne ha uno, inclusi, ad esempio, i televisori...
  - I televisori flat panel (LCD e Plasma) consumano in proporzione alla luminosità della scena...
- Cosa succede se mi faccio un database dei profili di consumo di un televisore a seconda del film?

Auditel on steroids

Quanto preciso?



M. Enev, S. Gupta, T. Kohno, and S. N. Patel. *Televisions, video privacy, and powerline electromagnetic interference.* (CCS '11)

Come proteggersi?

Batterie non incluse

- Tentare di uniformare il consumo di casa
  - ma non posso tenere costantemente acceso tutto....
- Vietare la registrazione dati di consumo troppo precisi
  - Funziona, ma riduce i vantaggi del metering
- Eliminare i picchi di consumo con sistemi a batteria tampone
  - Funziona molto bene per piccoli carichi

Conclusioni

Conclusioni

- I problemi legati alla privacy sono stati amplificati dalla capacità di trattare quantità di dati enormi
  - Ricordatevi dell' effetto "megafono"
- Considerate tutti i dati personali che avete fornito in rete come un unicum e non come tanti frammenti
- La "digitalizzazione" di informazioni apre nuove possibili perdite di privacy
  - La soluzione in questi casi è puramente tecnica[a]

[a] ma qualcuno deve progettarla