e-privacy 2010

Firenze, 28-29 maggio 2010

Deanonimizzazione:

perche' nessuno ci aveva pensato prima?

Marco A. Calamari - **marcoc@winstonsmith.org**

Progetto Winston Smith

Pubblicita'

Il ***Progetto Winston Smith*** e' una organizzazione informale di persone preoccupate per la privacy ed I diritti civili in Rete.

Realizza fin dallo scorso millennio iniziative di tipo tecnologico, legale e formativo per favorire l'uso delle tecnologie di comunicazione privata e sicura, e gestisce remailer, router Tor ed altri server per la privacy.

E' una organizzazione “ricorsiva”, infatti fornisce una prova di fattibilita' del completo anonimato raggiungibile con la tecnologia realizzandosi tramite una personalita' virtuale, collettiva ed anonima, che, citando il sempre piu' attuale romanzo “*1984*” prende il nome dal protagonista, *Winston Smith*.

Maggiori informazioni: *https://pws.winstonsmith.org*

Dati “pericolosi” ....

Ormai da anni persone, ed organizzazioni attive nella difesa dei diritti digitali ed in particolare della privacy in Rete enunciano i problemi che la circolazione usuale dei dati pone.

Normalmente ci si concentra solle applicazioni, sulla reti, sugli attori piu' o meno importanti, buoni o cattivi del mondo della Rete.

I difensori della privacy ripetono che memorizzare dati personali o generati dalle persone, in qualunque modo venga fatto costituisce u pericolo, che i dati vengono incrociati e che la conoscenza che rappresentano in questo processo si moltiplica.

Si citano nomi noti come Google, Doubleclick, meno noti come Netflix o AOL, o quasi ignoti come Acxiom.

... Dati pericolosi ?

Normalmente si considerano i dati divisi in varia categorie, riconducibili comunque a due.

Dati ordinari (innocui) e dati sensibili (pericolosi)

L'impianto delle leggi europee sulla privacy, ed anche della 196/2003 italiana, e che i dati sensibili, quelli veramente pericolosi, possono essere protetti segregandoli in ambienti limitati dove vengono usati senza essere fatti circoloare o copiati, oppure vengono resi anonimi sopprimendo le informazioni identificative, cominciando ovviamente dal nome e codice fiscale delle persone.

Questa tecnica, l'**anonimizzazione**, in linea di principio sembra funzionare senza problemi .....

Madornale errore!

No, non funziona, e ve lo dimostriamo

Non esistono dati innocui

Non esistono dati anonimi

(se sono davvero anonimi non servono a niente)

La somma di dati e' maggiore della somma delle parti

Dati censuari

How many other people in the United States share your specific combination of ZIP code, birth date, and sex?

According to a landmark study, for **87% of the American population**, the answer is zero; these three pieces of information **uniquely identify each of them**.

Reidentificazione ...

Latanya Sweeney, professor of computer science, who crunched 1990 census data and discovered that 87.1% of people in the United States were uniquely identified by their combined five-digit ZIP code, birth date (including year), and sex.

According to her study, even less specific information can often reveal identity, as 53% of American citizens are uniquely identified by their city, birth date, and sex, 18% by their county, birth date.

Netflix database

If an adversary knows the precise ratings a person in the data-base has assigned to six obscure movies, and nothing else, he will be able to identify that person 84% of the time.

If he knows approximately when (give or take two weeks) a person in the database has rated six movies, whether or not they are obscure, he can identify 99% of the people in the database.

In fact, knowing when ratings were assigned turns out to be so powerful, that knowing only two movies a rating user has viewed (with the precise ratings and the rating dates give or take three days), an adversary can reidentify 68% of the users.

Nome e cognome

“Ma cosi' non sanno il mio nome”

(in effetti l'identificazione di cui abbiamo parlato nelle slide precedenti non identifica la persona con nome e cognome, ma semplicemente l'individua nell'insieme di tutti gli altri membri del suo gruppo – nel caso degli U.S.A. 250.000.000 persone)

Puo' sembrare che identificare l'utente in questo modo rappresenti l'uno percento del lavoro necessario per violare la sua privacy. Rappresenta invece semmai il 99%. Basta che in un qualsiasi altro database il nome che ci serve sia associato ad una parte dei dati del database deanonimizzato ed il gioco e' fatto. Ed il gioco puo' adesso continuare associando i dati di altri database deanonimizzati nello stesso modo.

Ed ecco aperta la strada per il *“Database del mondo”*....

Riassumendo ....

Tutti i dati sono personali

(se non lo sono disaggregati, possono diventarlo quando saranno aggregati, e siccome valgono soldi, qualcuno cetamente prima o poi li aggreghera')

E che dire allora dei dati memorizzati obbligatoriamente su tutti i cittadini italiani (decreto Pisanu) o quelli che tanti cittadini italiani donano a Google e Facebook?

Dove trovare informazioni

Questa presentazione, ed in particolare i brani inglesi delle slide precedenti, sono estratti da una recente paper [1] di **Paul Ohm**

Le promesse infrante della privacy: rispondere al sorprendente fallimento dell'anonimizzazione

La sua lettura e' fortemente consigliata anche ai non addetti ai lavori; la sua aggiornata bibliografia e' un ottimo punto di partenza per approfondimenti.

La sua lettura e' ancora piu' consigliata agli addetti ai lavori, particolarmente ai membri dell'Autorita' Garante della protezione dei dati personali

Bibliografia essenziale

- [1] ***Broken promises of privacy: responding to the surprising failure of anonymization*** - Paul Ohm, 2009 Univ. Colorado Law Legal Studies Paper #09-12
- [2] **A Practical Attack to De-Anonymize Social Network Users -** *Gilbert Wondracek, Thorsten Holz, Engin Kirda, Christopher Kruegel*, 2009 - Technical Report TR-iSecLab-0110-001 – Technical University Vienna, Austria
- [3] **Untraceable Electronic Mail, Return addresses, and Digital Pseudonyms** - *David L. Chaum, 1981* http://www.weidai.com/mix-net.txt
- [4] **Tor: The Second-Generation Onion Router** – *R. Dingledine et al., 2004* – http://www.torproject.org/svn/trunk/doc/design-paper/tor-design.pdf
- **Bibliografia omnicomprensiva sull'anonimato tecnologico.** http://freehaven.net/anonbib/full/date.html
- **Progetto Winston Smith** http://www.winstonsmith.info

Grazie a tutti per l'attenzione

ci sono domande ?

Potete contattarmi qui: ***marcoc@winstonsmith.org***

Il Progetto Winston Smith

mail: info@winstonsmith.org

web: http://pws.winstonsmith.org

tor: http://5zaspldty2calvcq.onion/

freenet: `USK@RU-C2q5kN7K62WO3seMMjSTUY8izF2vCFyVFOnLf~Q0,wxvGO2QMT6IN9c7dNUhHeHnXVVwhq8YLbQL~DlMA7YE,AQACAAE/pws/7/`