Navigare “privatamente”

Author Giovambattista Vieri (ENT S.R.L.)
g.vieri@ent-it.com

Navigare privatamente

- Oggi navigare su internet e' facile.
- Ottenere il rispetto della propria privacy potrebbe essere meno facile.
- Esistono pero' molte soluzioni per mantenere private le nostre mete, e sperabilmente anche I nostri dati.

Struttura del talk

- Situazione attuale
- Breve riferimento a delle tecnicalita'
- Approfondimento della situazione
- Possibili strategie,
- Verifiche
- Link a tecnicalita' e documentazione
- Q&A

Situazione attuale

- Siamo nel mezzo di una rivoluzione nell'uso dei browser.
- Sempre piu' siti usano un protocollo cifrato (https) che potrebbe proteggerci dai curiosoni occasionali.
- La legge (anche a livello sovranazionale) comincia a occuparsi del problema (leggi sui cookie in europa e, sentenza sul safe harbor)

Situazione Attuale

- D'altro canto moltissimi servizi “indispensabilmente social” sono offerti a titolo gratuito. Pure le societa' che li offrono sono redditizie e, distribuiscono dividendi.
- Quindi potremmo immaginare che noi, o meglio le nostre azioni su internet siano oggetto di studio e analisi e, che questi studi e analisi piu' o meno anomizzati, siano oggetti di “commercio”.

Le nostre abitudini

- Prendiamo ad esempio la mattina del lunedi' ..
- Dopo esser usciti e magari aver fatto colazione, ci si reca al lavoro.. sul mezzo pubblico.
- Nel tragitto cominciamo ad accedere alla nostra e-mail, ai siti di notizie e magari a dei social network.
- Ognuna di queste abitudini, lascia tracce. Piu' o meno diffuse. Piu' o meno precise.

Abitudini social

- Quando accediamo a un social network, tipicamente ci autentichiamo.
- Ergo, volentemente conferiamo la conoscenza della nostra persona, gusti, abitudini, storia personali a entita' che potrebbero anche essere site fuori dall'europa.
- Questa volontarieta' fa' si' che le abitudini social non saranno soggette di questo talk se non a latere di altre azioni.

Sempre di mattina

- Accediamo a web-mail, (ergo riceviamo cookie) e, il fornitore del servizio conosce gli indirizzi e-mail dei ns corrispondenti. Il fornitore del servizio potrebbe anche ispezionare su base meramente statistica le keyword contenute nei messaggi.
- Accediamo a piu' siti di notizie che potrebbero avere uno stesso fornitore di pubblicita' o di “proxying” dei contenuti

Privacy, advertising e proxying dei contenuti

- Molti siti fanno uso di pubblicita'. Alcuni di noi usano anche degli “add-on” per bloccarla.
- Molti siti usano dei proxy di contenuti per ridurre il loro costo e, offrire un servizio migliore.
- Tecnicamente significa che il nostro browser non mandera' solo richieste al sito che noi vogliamo vedere ma anche ad altri.
- Che, POTREBBERO incrociarli.

Incroci di traffico

- Incrociare I dati di traffico e' possible.
- Ora con le nuove normative sui “cookie” dovrebbe essere piu' difficile.
- Sicuramente potrebbe essere illegittimo spingere la profilazione oltre quanto da noi esplicitamente consentito.

Qualche tecnicalita'

- Per navigare su internet ci avvaliamo di programmi chiamati browser.
- Questi programmi possono essere fortemente integrati o meno nel sistema operativo usato dal nostro computer/tablet/mobile phone, o essere forniti da terzi
- Tutti I browser “parlano” dei protocolli condivisi con le loro controparti (web server)

protocolli

- I protocolli usati dai browser sono per la maggior parte HTTP e HTTPS.
- Per semplicita' ci limiteremo a HTTP. Questo protocollo e' descritto abbondamente e chiaramente.
- E' da notare che e' un protocollo “testuale” e, in chiaro (non cifrato). Quindi potremo ispezionarlo a nostro piacimento.

Cenni di http

- Testuale e “Stateless”
- Con keyword simili al linguaggio naturale “GET” “PUT” “Set-Cookie” etc.
- Il protocollo http si appoggia su un protocollo TCP/IP (anche lui molto ben descritto p.es in TCP ILLUSTRATED)
- Questo protocollo e' relativamente facile da “intercettare”.

Transazioni HTTP

- Saltiamo tutta la parte di “rete” e “dns”

```
GET / HTTP/1.0

HTTP/1.1 401 Authorization Required
Date: Mon, 12 Oct 2015 21:19:30 GMT
Server: Apache/2.2.15 (CentOS)
WWW-Authenticate: Basic realm="area riservata"
Vary: Accept-Encoding
Content-Length: 401
Connection: close
Content-Type: text/html; charset=iso-8859-1
```

Tutto qui?

- Beh no.

```
GET / HTTP/1.1\r\n
Host: www.ziogianni.com\r\n
User-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:38.0) Gecko/20100101 F
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0
Accept-Language: en-US,en;q=0.5\r\n
Accept-Encoding: gzip, deflate\r\n
Cookie: __utma=219111826.686701955.1422306524.1423579202.1431554445.3
Connection: keep-alive\r\n
If-Modified-Since: Tue, 02 Jun 2015 09:38:37 GMT\r\n
If-None-Match: "1e60d27-5ea-51785b4ebe140"\r\n
Cache-Control: max-age=0\r\n
\r\n
```

Vedete la differenza?

- Una linea...
- Molte linee...
- Escludiamo I cookie (per ora)
- User-Agent:
- Accept:

Http header

- Accept-* (ci son molti Accept): in sintesi, e' un header che comunica a un server le preferenze e/o le capacita' dei browser.
- User-Agent: e' un header che fornisce la carta di identita' del browser al server, e incidentalmente a tutti coloro in ascolto.
- Un esempio di user-agent: Mozilla/5.0 (iPad; U; CPU OS 3_2_1 like Mac OS X; en-us) AppleWebKit/531.21.10 (KHTML, like Gecko) Mobile/7B405

User-agent

- Ci dice il sistema operativo,
- Il tipo di browser e la versione,
- La lingua,
- Le versioni di “librerie”
- … e altro …

Cominciamo a capire?

- Poi ricordiamo che il server puo' sapere anche l'ip di provenienza della ns connessione (ergo provider e, piu' o meno precisamente la localita' geografica da cui ci connettiamo)
- E noi, NOI, gli comunichiamo os, e altro …

Inoltre I browser

- Sono in grado di processare altre estensioni:
  - Flash ?
  - Javascript ?
  - Java ?
  - Immagini con formati piu' o meno esoterici
- Possono avere plugin …

e..

- I plugin abilitati, processano delle informazioni... che a loro volta potrebbero farne caricare altre.
- NON DOBBIAMO DIMENTICARE CHE OGNI “CARICAMENTO” DI DATI IMPLICA UNA RICHIESTA AD UN SERVER.
- Quindi una pagina che contiene contenuti in flash ™ o java ™ in javascript o altro, che a loro volta caricano altri contenuti fornisce informazioni sulla configurazione del browser e, sulle nostre abitudini e attitudini.

Un browser comunica al server:

- Ip address di partenza
- User agent (con os, tipo di browser etc etc)
- Una parte dei plugin caricati (che a loro volte caricano altri dati)

Quindi un utente

- Fornisce molte informazioni al primo collegamento.
- Queste informazioni in casi particolari, possono identificare da sole univocamente un utente.

Inoltre:

- Le informazioni precedenti unite a una analisi statistiche dei comportamenti dell'utente puo' consentire una profilazione ancora piu' fine.
- Potremmo immaginarci un ipotetico server che tra se e se mormora:
- “ecco l'utente linux con konqueror da debian che vuol conoscere le ultime informazioni sul calcio... beh prepariamo le informazioni finanziarie che ci chiedera' tra 5 minuti”... ?

TECNICAMENTE SI.

- Ovviamente se I server mormorassero. :-)
- Di certo e' ragionevole che certi server profilino in modo spinto I propri utenti. Immaginiamo l'utilita' di una antifrodi che analizzando il comportamento di un utente Romano, magari anziano e non particolarmente “geek”/”nerd” che improvvisamente si collega (dopo 30 minuti) con un browser “non maintream” magari da un paese in un altro continente.

Il problema e' noto?

- Agli esperti sicuramente si'.
- https://developer.mozilla.org/en-US/docs/Browser_detection_using_the_user_agent
- http://www.wilderssecurity.com/threads/improving-the-privacy-with-generic-browser-user-agent-strings.358284/
- https://www.eff.org/deeplinks/2010/01/tracking-by-user-agent
- https://wiki.mozilla.org/Fingerprinting
- http://programmers.stackexchange.com/questions/122372/is-browser-fingerprinting-a-viable-technique-for-identifying-anonymous-users

continua

- http://www.w3.org/wiki/images/7/7d/Is_preventing_browser_fingerprinting_a_lost_cause.pdf
  http://www.computerworld.com/article/2517643/internet/eff--forget-cookies--your-browser-has-fingerprints.html
- https://www.reddit.com/comments/1ic6ew
- https://blog.whitehatsec.com/i-know-a-lot-about-your-web-browser-and-computer/

Dunque:

- Usare sistemi operativi non mainstream
- Disabilitare javascript
- Non usare flash
- Usare browser non mainstream
- Insomma non essere UGUALI agli altri componenti del “gregge” utenti

Porta a una maggiore profilazione

- Cio' e' deprimente ?
- Pensate agli utenti TOR o altri “router” anonimizzatori...
- La profilazione teste' descritta li fa' emergere senza problemi...

Che fare ?

- Anonimi nel gregge.
- Ovvero usare “User-Agent” diffusi.
- Usare configurazioni main-stream ove possibile.
- Avere una routine di uso associato a un “profilo browser”.
- Usare plugin di “greggizzazione”.

Seriamente:

- Il problema esiste.
- Il differenziarsi diventa “pericoloso”.
- Queste “nuove” tecniche unite alle vecchie (cookie?) possono avere una precisione interessantissima.
- L'attuale abitudine di usare javascript intensivamente e estensivamente, fornisce altre informazioni ai profilatori (versioni dell'idioma, per esempio)

Consigli pratici:

- Usare browser che consentano di modificare lo user_agent
- Usare browser che paiano mainstream
- Navigare anonimamente.
- Usare piu' “profili utente”:
  - Un profilo generico per leggere I giornali
  - Un profilo diverso per I social
  - Altri profili utente puntuali all'uso desiderato.

Tecnicalita':

- Di seguito procedure e link a tool per verificare quanto detto
- https://github.com/gvieri/conferences_material/tree/master/ws_2015
  Slide, Dockerfiles e scripts per analizzare quanto esposto
- http://www.howtogeek.com/113439/how-to-change-your-browsers-user-agent-without-installing-any-extensions/