RELAZIONE PROGETTO GESTIONE DI RETI

Introduzione

Nell'industria della pubblicità online il tracciamento delle abitudini di navigazione degli utenti al fine di offrire pubblicità mirata è una pratica molto diffusa.
Visitando una pagina web, c'è una buona probabilità che questa contenga codice javascript e web beacon di svariate aziende, il cui scopo è tracciare e registrare l'attività dell'utente. La diffusione di questi elementi in grandi porzioni del web permette a queste aziende di seguire l'attività dell'utente attraverso diversi siti e costruire un profilo dettagliato di interessi e abitudini.
L'analisi del comportamento dell'utente, della sua interazione con le pagine, di ricerche, acquisti recenti e provenienza geografica permette di offrire una pubblicità "su misura", aumentando la probabilità che l'utente clicci sull'annuncio.
Questi meccanismi pongono notevoli problemi di privacy dal momento che il tutto avviene che l'utente sia a conoscenza o possa avere controllo su quanti e quali dati vengono utilizzati, da chi e a che fini.

Il progetto

Lo scopo del progetto è analizzare il traffico di una rete per visualizzare i collegamenti ai domini che effettuano attività di tracking, mostrando statistiche che aiutano a comprendere la portata del fenomeno.
A questo scopo ho scelto la libreria open source di deep packet inspection nDPI, modificando il file di esempio ndpiReader.

ndpiReader

L'argomento -a <nomefile> (o –tracking <nomefile>) accetta in input una lista di domini, uno per riga, che viene utilizzata per il riconoscimento del traffico. Il file script.sh genera una lista adatta allo scopo.
parse_tracking_file è la funzione che all'avvio del programma legge il file dato come argomento e inizializza l'automa per il pattern matching fornito dalla libreria nDPI.
Se tra i flussi raccolti il nome di dominio (*host_server_name* o *ssh_ssl.client_info*) corrisponde con uno di quelli prescelti, accanto al flusso è stampato il tag [TRACKING] ed è chiamata **collect_tracker_stats**, che popola la tabella hash contenente le statistiche sui domini con la struttura **tracker_stats**

```
struct tracker_stats {
    char *domain;                /* key */
    u_int64_t packets;
    u_int64_t bytes;
    u_int32_t flows;
    UT_hash_handle hh;         /* makes this structure hashable */
};
```

Per ogni dominio individuato sono stampati il numero di pacchetti, bytes e flussi, e infine la percentuale di flussi individuati sul totale dei flussi, divisa fra i protocolli dns, http, ssl.

script.sh

È uno script bash che genera una lista di domini che effettuano attività di tracking e advertising, utilizzando fonti pubbliche.
easylist ed easyprivacy, le principali liste dei browser addon che bloccano le pubblicità nelle pagine web, come Adblock Plus e uBlock, contengono un lunghissimo elenco di url da bloccare. Per ottenere una lista di soli domini di tracking sono estratti con una espressione regolare solo le righe che bloccano un intero dominio, senza considerare quelle che bloccano singoli indirizzi o elementi html, per eliminare la possibilità di falsi positivi.
Altre lite utillizzate sono quella di Disconnect, un addon che blocca i siti di tracking, Peter Lowe's List e NoTrack Blocklist.
Queste liste sono poi unite, ordinate e private di elementi duplicati.

Compilazione

È necessario ottenere nDPI e applicare la patch

```
git clone https://github.com/ntop/nDPI.git
mv tracker.patch nDPI
cd nDPI
git apply tracker.patch
```

Se l'applicazione della patch fallisce si può utilizzare la revisione su cui ho lavorato

```
git checkout f336f1a340b4d6bf417fab56e76b5a7937b8b368
```

oppure provare il 3-way mege

```
git apply -3 tracker.patch
```

Una volta applicate le modifiche, è possibile compilare nDPI

```
./autogen.sh
./configure
make
```

ndpiReader si trova in examples/

Esempio di esecuzione

```
./script.sh list
```

```
./ndpiReader --tracking list -i eth0
```

Segue il risultato dell'esecuzione durante una visita alla home page del sito d'informazione repubblica.it
È possibile notare la grande quantità di domini di tracking contattati

```
Tracking and Advertising Stats:
       accounts.us1.gigya.com                 packets: 26           bytes: 8900         flows: 2
       ad.doubleclick.net                     packets: 373          bytes: 260894       flows: 2
       ade.googlesyndication.com              packets: 26           bytes: 7029         flows: 2
       ads.rubiconproject.com                 packets: 25           bytes: 10006        flows: 2
       adx.g.doubleclick.net                  packets: 70           bytes: 19088        flows: 2
       b.scorecardresearch.com                packets: 24           bytes: 6820         flows: 2
       beacon-eu-ams3.rubiconproject.com      packets: 4            bytes: 956          flows: 1
       cdn-gl.imrworldwide.com                packets: 120          bytes: 69236        flows: 3
       cdns.gigya.com                         packets: 97           bytes: 67549        flows: 2
       cdns.us1.gigya.com                     packets: 53           bytes: 29664        flows: 2
```

```
        ds-aksb-a.akamaihd.net                  packets: 25           bytes: 7762          flows: 2
        geo.moatads.com                         packets: 27           bytes: 7730          flows: 2
        googleads.g.doubleclick.net             packets: 50           bytes: 13122         flows: 3
        googleads4.g.doubleclick.net            packets: 44           bytes: 9967          flows: 2
        gruppoespresso01.webtrekk.net           packets: 19           bytes: 4713          flows: 2
        gscounters.us1.gigya.com                packets: 25           bytes: 9452          flows: 2
        it-gmtdmp.mookie1.com                   packets: 15           bytes: 2650          flows: 2
        optimized-by.rubiconproject.com         packets: 16           bytes: 4328          flows: 2
        pagead2.googlesyndication.com           packets: 157          bytes: 87673         flows: 4
        ping.chartbeat.net                      packets: 19           bytes: 5048          flows: 2
        px.moatads.com                          packets: 22           bytes: 9068          flows: 2
        s0.2mdn.net                             packets: 398          bytes: 282981        flows: 4
        secure-assets.rubiconproject.com        packets: 47           bytes: 27665         flows: 2
        secure-it.imrworldwide.com              packets: 106          bytes: 25117         flows: 10
        securepubads.g.doubleclick.net          packets: 206          bytes: 125705        flows: 2
        static.chartbeat.com                    packets: 35           bytes: 16417         flows: 2
        tpc.googlesyndication.com               packets: 425          bytes: 283264        flows: 6
        www.googletagservices.com               packets: 17           bytes: 4221          flows: 2
        z.moatads.com                           packets: 106          bytes: 72849         flows: 2
        TOTAL                                   packets: 2577         bytes: 1479874       flows: 75

Protocols: (tracking flows/total flows)
        DNS:     10 / 38    26.32%
        HTTP:     9 / 36    25.00%
        SSL:     14 / 14    100.00%
```