Mitos y realidades de la red Tor:
Análisis de tráfico en un nodo de salida

Castro Rendón Virgilio

Premio Universitario ESET

Noviembre de 2017

Tabla de contenido

Abstract

A lot of people use the Tor network to feel safer in internet as it prevents hackers and governments to spy their communications. Or not? This investigation will try to solve this question by showing how is it posible to get personal information from Tor by mounting and analyzing a tor exit relay. During the process, I will also show how difficult it is to help the Tor project and what is the network mostly used for.

Objetivo

Analizar tráfico que viaja a través de la red Tor mediante la implementación de un nodo de salida y captura de tráfico y, en el proceso, mostrar a los lectores y usuarios novatos de dicha red para qué es usada, desmentir algunos mitos sobre la red y buscar actividad maliciosa dentro de ella. En general se pretende responder a las preguntas: ¿es la red Tor sinónimo de privacidad?, ¿es la red Tor usada exclusivamente para entrar en la "*dark net*"? y ¿es difícil contribuir al proyecto Tor?".

Antecedentes

¿Qué es Tor?

[1] La red Tor es un grupo de servidores administrados por voluntarios que ayuda a las personas a mejorar su privacidad y seguridad en la red. Los usuarios de Tor utilizan esta red, como una serie de túneles virtuales en lugar de hacer una conexión directa entre un cliente y un servidor, por lo que permite a organizaciones e individuos compartir información en redes públicas sin comprometer su privacidad. Al usar Tor, los sitios no pueden rastrear a sus usuarios, pues realmente verían en sus registros las direcciones IP correspondientes a un nodo Tor. Asimismo, Tor cifra todos los datos enviados a través de la red.

¿Quién utiliza Tor?

[1] Tor es utilizada por periodistas, denunciantes y disidentes para que no se les pueda rastrear. Asimismo es utilizada por personas en países que tienen bloqueos regionales a internet y, de esta manera, evitar dicho bloqueo. Igualmente es usado por personas que quieren hacer uso de los "servicios cebolla" y, en general, por cualquiera que quiere mantener su privacidad mejor protegida.

¿Qué es un nodo Tor?

Un nodo Tor es un servidor que ayuda a la red Tor a proveer ancho de banda. Los nodos son los que forman los túneles virtuales que evitan las conexiones directas entre un cliente y un servidor. La conexión entre el cliente y la red Tor, así como todas las conexiones entre los diferentes nodos, van cifradas, por lo que no se puede analizar el tráfico en esos puntos. Un nodo puede ser usado para,

simplemente, reenviar el tráfico al siguiente nodo en la conexión o, también, como un nodo de salida.

¿Qué es un nodo de salida?

Es el último nodo o servidor en el túnel virtual, por lo tanto es el que establece la comunicación directamente con los servidores finales (Por ejemplo: *google.com*). Este servidor se encarga de descifrar los datos que viajaron cifrados en todo el trayecto, pues tiene que ser así para que el sitio destino original sea capaz de entender la información. Es la dirección IP de este último nodo la que realmente aparecerá en los registros de los sitios visitados.

¿Qué son los servicios cebolla?

También conocidos como "*hidden services"* u *"onion services"*, representan lo que comúnmente se conoce como "*Dark Net"* pues no pueden ser accedidos de forma convencional y se requiere de una conexión a través de Tor. Son sitios alojados en nodos Tor (Por ejemplo: *https://facebookcorewwwi.onion*), esto implica que no es necesario salir de la red Tor para visitarlos y, por lo tanto, el tráfico entre el cliente y el servidor siempre se encuentra cifrado. Debido a esta cualidad, los servicios cebolla no son vulnerables a ser analizados en un nodo de salida y, por lo tanto, salen del alcance de este trabajo.

Preparación del ambiente

[2] En esta sección se explica detalladamente todo lo que se hizo para poner en marcha un nodo de salida en la red Tor.

El primer paso fue adquirir un servidor con una dirección IP pública, por lo que se rentó un servidor virtual. Es común que los proveedores de servidores virtuales tengan prohibido explícitamente en sus políticas el instalar y configurar un nodo de Tor. Esto se debe a que es común que la red sea utilizada por atacantes para cubrir su localización, lo que implica que todas las quejas llegarán directamente al proveedor, por lo que prefieren prohibir este tipo de aplicaciones.

Por lo tanto se investigó sobre algún proveedor que no prohibiera ésto. Finalmente se decidió rentar con una empresa que promete ancho de banda ilimitado (característica especialmente útil para este trabajo) y no prohíbe explícitamente el levantar un nodo Tor.

Una vez teniendo acceso al servidor, puesto que se trata de un servidor público y queda expuesto a amenazas, se llevaron a cabo tareas básicas de configuración y seguridad.

- Se instaló y configuró en un sistema operativo Debian 9.

```
[redacted]@server:~$ cat /etc/*release
PRETTY_NAME="Debian GNU/Linux 9 (stretch)"
NAME="Debian GNU/Linux"
VERSION_ID="9"
VERSION="9 (stretch)"
ID=debian
```

Ilustración 1. Versión del sistema operativo

- Se hizo una actualización de los paquetes.

```
[redacted]@server:~$ sudo apt-get update
Hit:1 http://security.debian.org/debian-security stretch/updates InRelease
Ign:2 http://ftp.de.debian.org/debian stretch InRelease
Hit:3 http://ftp.de.debian.org/debian stretch-updates InRelease
Hit:4 http://ftp.de.debian.org/debian stretch Release
Reading package lists... Done
```

Ilustración 2. Actualización de los paquetes que se pueden instalar en el servidor

- Se instaló y usó *chkrootkit* para determinar si en el sistema estaba instalado algún rootkit por defecto. Al no encontrar ninguna amenaza de este estilo, se continúa normalmente con las configuraciones.

```
[redacted]@server:~$ sudo chkrootkit
ROOTDIR is `/'
Checking `amd'...                                           not found
Checking `basename'...                                      not infected
Checking `biff'...                                          not found
Checking `chfn'...                                          not infected
Checking `chsh'...                                          not infected
Checking `cron'...                                          not infected
Checking `crontab'...                                       not infected
```

Ilustración 3. Ejecución de chkrootkit

- Se instalaron y configuraron las actualizaciones automáticas de paquetes para mantener el sistema con las últimas actualizaciones de seguridad todo el tiempo.

```
[redacted]@server:~$ sudo apt-get install unattended-upgrades apt-listchanges -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
apt-listchanges is already the newest version (3.10).
unattended-upgrades is already the newest version (0.93.1+nmul).
```

Ilustración 4. Instalación de actualizaciones automáticas

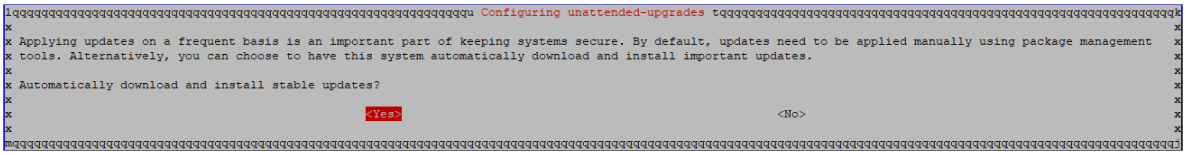


Ilustración 5. Configuración de actualizaciones automáticas

- Se configuró SSH para no permitir conexiones al usuario *root*.

```
[redacted]@server:~$ sudo grep Root /etc/ssh/sshd_config
PermitRootLogin no
# the setting of "PermitRootLogin without-password".
```

Ilustración 6. Usuario root no permitido

- Se instaló el programa “tor” mediante paquetes.

```
[redacted]@server:~$ sudo apt-get install tor
Reading package lists... Done
Building dependency tree
Reading state information... Done
tor is already the newest version (0.2.9.12-1).
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
```

Ilustración 7. Instalación de tor

- Se configuró Tor para funcionar como un nodo de salida. Esto se logra agregando la directiva "*ORPort*" en el archivo de configuración de Tor (/etc/tor/torrc). El ancho de banda elegido fue de 2MB, suficiente para ser aceptado como nodo de salida. Asimismo, se estableció una política reducida que se obtuvo a partir de las políticas propuestas en la página oficial [3]. Sólo se agregaron los puertos que comúnmente utilizan los chats IRC.

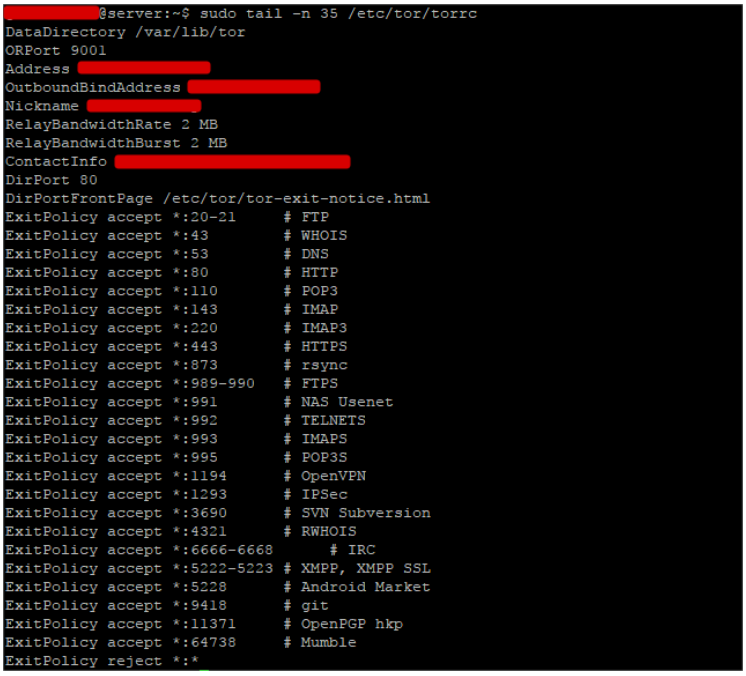


Ilustración 8. Configuración de tor

- Una vez definidos los puertos permitidos, se aplicaron las correspondientes reglas en el firewall del servidor para permitir esas comunicaciones.

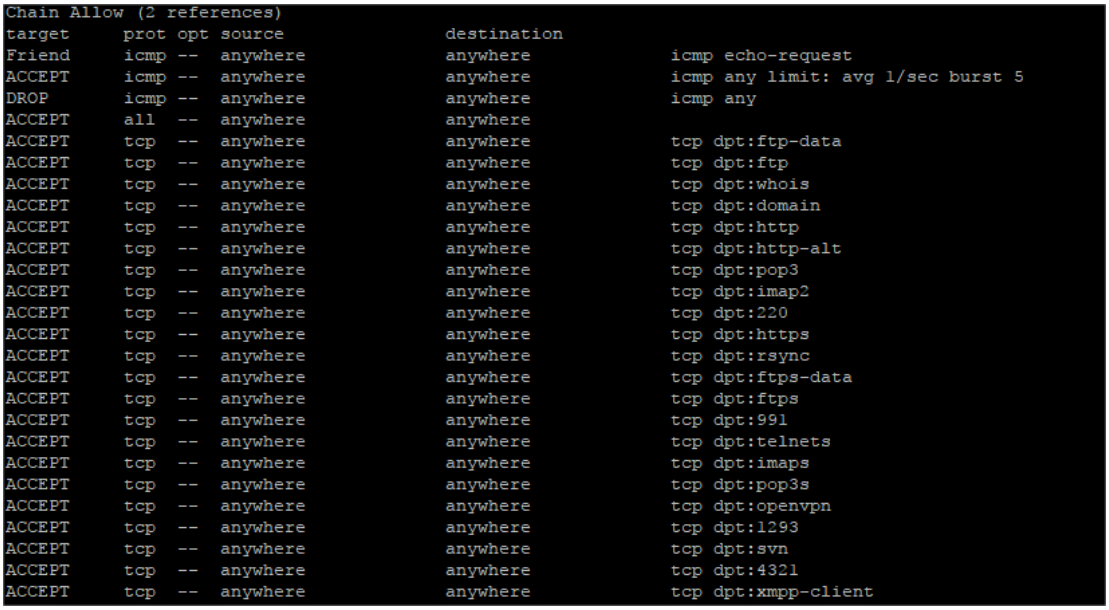


Ilustración 9. Reglas activas en el firewall

- Puesto que es común que los atacantes usen Tor para protegerse, se agregó una página explicativa en el servidor web sobre el funcionamiento de la red y el papel del servidor en ella. Esto debido a que se ha demostrado que, al estar enteradas las víctimas de que se trata de un nodo de salida Tor, las denuncias se reducen considerablemente. Dentro de la página se incluyó mi correo electrónico para recibir los correos de "*abuse*" y poderles dar una respuesta oportuna.

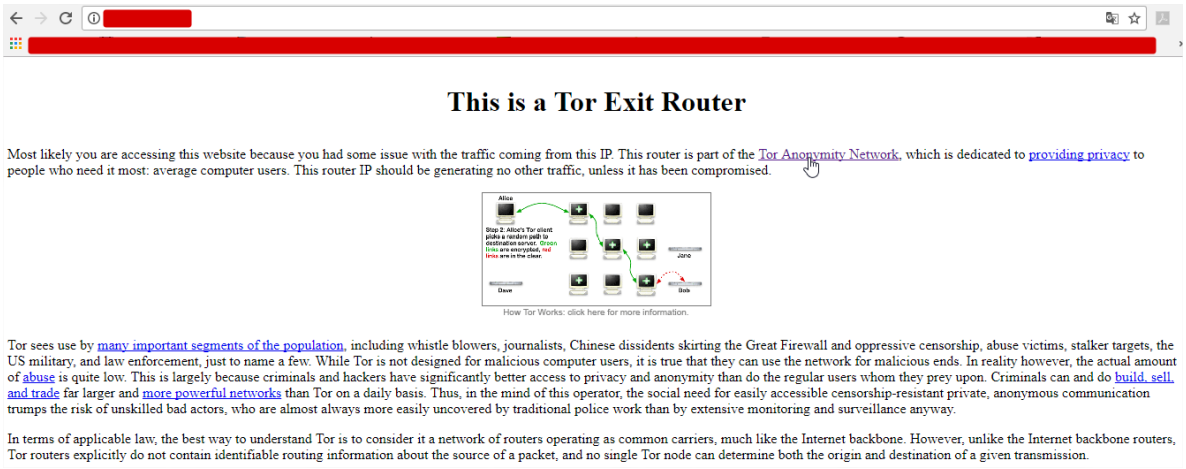


Ilustración 10. Servicio web con descripción del nodo

- Una vez teniendo listo todo lo anterior, sólo fue cuestión de tiempo para que la red nos aceptara como un nodo de salida. Automáticamente se hacen pruebas para ver si el nodo no está redirigiendo tráfico a otros sitios o cambiando las consultas DNS y si tiene ancho de banda suficiente, entre otras pruebas.

- Se puede revisar fácilmente si está aceptado como un nodo de salida en la página de búsquedas [4] de Tor, pues muestra todas las características del nodo, incluida la bandera "*exit*" en caso de ser aceptado como un nodo de salida.

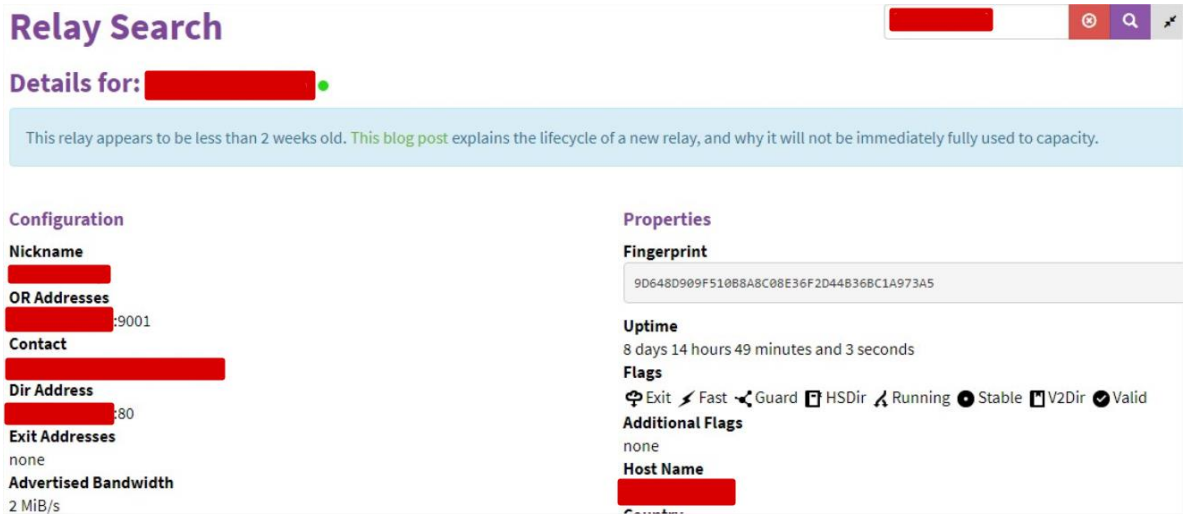


Ilustración 11. Descripción del nodo en la página oficial

- A continuación se muestra el comando utilizado para hacer la captura de tráfico. Se decidió hacer un archivo nuevo cada hora, con el fin de evitar que tuvieran tamaños excesivos. Asimismo, sólo se captura lo que sale de nuestra dirección IP y se excluye nuestra conexión por SSH.

```
[redacted]@server:~$ cat capture_traffic.sh
#!/bin/bash
/usr/sbin/tcpdump -i ens18 -w /home/[redacted]/traffic/dump_%Y-%m-%d_%H:%M:%S.pcap -G 3600 -W 1 'src [redacted] and not (src port 22)'
```

Ilustración 12. Comando para capturar tráfico

- Se creó una nueva tarea en cron para ejecutar el script de la ilustración 12 cada hora

```
[redacted]@server:~$ sudo crontab -l
# m h  dom mon dow   command
#0 * * * *  /bin/bash /home/[redacted]/capture_traffic.sh
```

Ilustración 13. Tarea programada para capturar tráfico

Análisis de hosts conectados

El objetivo de este análisis es obtener información de primera mano de los hosts con los que se está comunicando el servidor. La información de los hosts fácilmente se obtiene con el comando *netstat* y se puede hacer un filtro de las direcciones IP correspondientes a otro nodo Tor y un servidor común y corriente.

Como se determinó en el archivo de configuración de tor, el puerto 9001 se usa para las comunicaciones con otros servidores tor. Por lo tanto, cualquier conexión que inicie con un puerto diferente corresponde a una conexión establecida con un servidor común. Como se puede ver en la ilustración 14, existen más de 4300 conexiones simultáneas.

```
4306 tcp        0      0 [redacted]:9001     91.8.18.235:55376       ESTABLISHED
4307 tcp        0      0 [redacted]:44607    95.213.209.60:80        TIME_WAIT
4308 tcp        0      0 [redacted]:9001     195.154.242.122:47602   ESTABLISHED
4309 tcp        0      0 [redacted]:9001     178.255.42.86:40110     ESTABLISHED
4310 tcp        0      0 [redacted]:9001     91.8.18.235:55376       ESTABLISHED
4311 tcp        0      0 [redacted]:9001     178.255.42.86:40110     ESTABLISHED
4312 tcp        0      0 [redacted]:9001     91.8.18.235:55376       ESTABLISHED
```

Ilustración 14. Salida de netstat que muestra conexiones

Al ejecutar el comando *'awk '{print $4 "\t"$5}' netstat.txt'* se puede filtrar los datos del archivo para obtener únicamente las direcciones IP y puertos de los hosts.

```
[redacted]:44791    208.91.197.54:80
:9001       83.162.202.182:39577
:9001       163.172.69.166:33896
:9001       146.0.139.88:47319
:9001       94.130.28.151:34903
:43613      147.14.11.114:443
:9001       216.158.226.216:47654
:9001       192.71.245.137:60840
```

Ilustración 15. Hosts conectados

Teniendo la información anterior, se puede crear un script como el siguiente para escribir en archivos separados los hosts que pertenecen a la red Tor (los que salen del puerto 9001 de nuestro servidor) y los hosts que son servidores comunes, es decir, los visitados por los usuarios de la red.

```
1 #!/usr/bin/python
2 with open('hosts','r') as ifile, open('tor_servers','w') as o_tor, open('common_server','w') as o_server:
3     lines = set(ifile.readlines())
4     for l in lines:
5         hosts = l.split()
6         if hosts[0].endswith(':9001'):
7             o_tor.write(hosts[1].split(':')[0]+'\n')
8         else:
9             o_server.write(hosts[1].split(':')[0]+'\n')
0
```

Ilustración 16. Script para filtrar hosts conectados

Una vez teniendo por separado las direcciones IP de servidores que pertenecen a la red Tor y de los que no pertenecen, podemos ejecutar comandos como *whois* de forma secuencial y obtener información de cada uno. Haciendo esto puede ser que se encuentre mucha o poca información; en este caso, podemos determinar que se están buscando páginas que se encuentras alojadas en China.

```
% [whois.apnic.net]
% Whois data copyright terms    http://www.apnic.net/db/dbcopyright.html

% Information related to '58.208.0.0 - 58.223.255.255'

% Abuse contact for '58.208.0.0 - 58.223.255.255' is 'anti-spam@ns.chinanet.cn.net'

inetnum:        58.208.0.0 - 58.223.255.255
netname:        CHINANET-JS
descr:          CHINANET jiangsu province network
descr:          China Telecom
descr:          A12,Xin-Jie-Kou-Wai Street
descr:          Beijing 100088
country:        CN
admin-c:        CH93-AP
tech-c:         CJ186-AP
mnt-by:         APNIC-HM
mnt-lower:      MAINT-CHINANET-JS
mnt-routes:     MAINT-CHINANET-JS
remarks:        ------------------------------------------------------
remarks:        To report network abuse, please contact mnt-irt
remarks:        For troubleshooting, please contact tech-c and admin-c
remarks:        Report invalid contact via www.apnic.net/invalidcontact
remarks:        ------------------------------------------------------
status:         ALLOCATED PORTABLE
last-modified:  2016-05-04T00:01:43Z
source:         APNIC
mnt-irt:        IRT-CHINANET-CN

irt:            IRT-CHINANET-CN
address:        No.31 ,jingrong street,beijing
address:        100032
e-mail:         anti-spam@ns.chinanet.cn.net
```

Ilustración 17. Información de un host visitado

Se puede obtener la misma información de las direcciones correspondientes a los demás nodos usando el comando '*whois'* y, haciendo una búsqueda un poco más profunda, se puede determinar qué tipo de organizaciones o personas son las que apoyan en mayor o menor medida el proyecto Tor. Por ejemplo, es común encontrar páginas cuyo principal tema es la seguridad informática apoyando con un nodo, como se ve en las ilustraciones 18 y 19. Se recuerda que es posible que cualquier nodo Tor podría estar alojando un *hidden service* (sitio .onion), pero esto no se puede determinar con este tipo de análisis.

Ilustración 18. Sitio web de un nodo Tor

Ilustración 19. Sitio web de un nodo Tor 2

Análisis del tráfico

Se capturaron 222 GB de tráfico de red en 48 horas, repartidos en 48 archivos en formato *pcap*. Esto significa un promedio de 4.6 GB de tráfico por hora. Lo que demuestra que la red es aún muy utilizada y aunque esta cantidad de tráfico es poco (en comparación con los más de 100Gbits/s que viajan por la red) [5], se considera como tráfico suficiente para ser analizado.

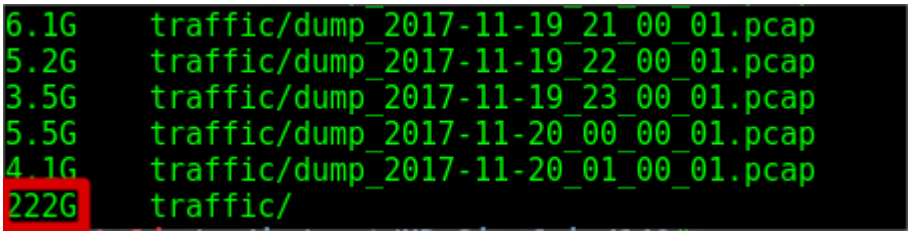


Ilustración 20. Capturas de tráfico y su tamaño

Consultas DNS

Sitios más buscados

El objetivo de este análisis es mostrar los sitios más comúnmente visitados por los usuarios de Tor. Para obtener esta información, se hace un filtro con *tshark* , el cual permite obtener los nombres de dominio que fueron consultados. Para automatizar la búsqueda en todos los archivos *pcap*, se hizo un script sencillo.

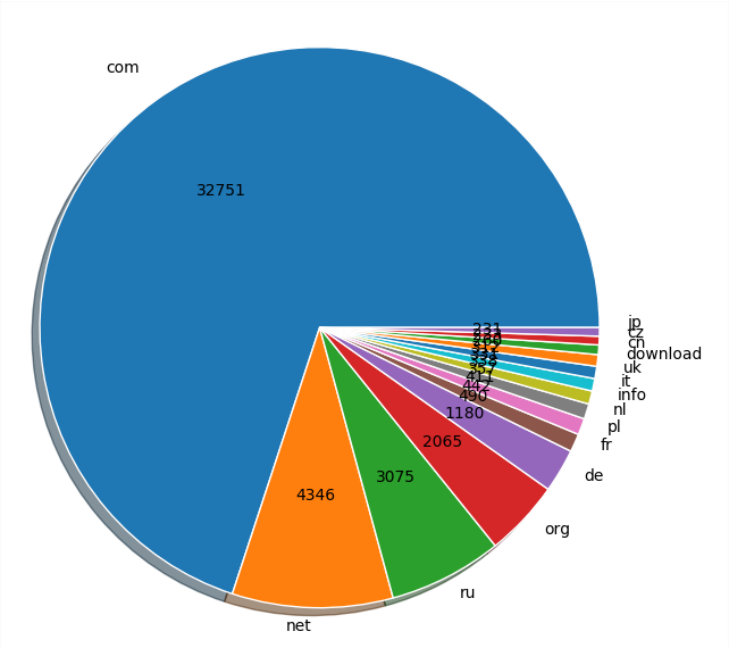
```
1 for dump in traffic/*; do
2     tshark -r "$dump" -T fields -e dns.qry.name -Y "dns.flags.response eq 0" >> dns_queries
3 done
4
```

Ilustración 21. Comando para obtener peticiones DNS

En total se hicieron 53,433 consultas DNS en aproximadamente diez horas. Se desarrolló un script en *python* que permite obtener los dominios de nivel superior y segundo nivel más buscados (top 15) y mostrarlos gráficamente.

```
 1 #!/usr/bin/python
 2 with open('dns_queries','r') as ifile:
 3     domains = [d.lower() for d in ifile.readlines()]
 4
 5 tld = {}
 6 sld = {}
 7
 8 for d in domains:
 9     #Agrega el dominio de orden superior al diccionario o aumenta su cuenta
10     subdomains = d.split('.')
11     if subdomains[-1] not in tld:
12         tld[subdomains[-1]] = 1
13     else:
14         tld[subdomains[-1]] += 1
15
16     #Agrega dominios de segundo nivel a su diccionario o aumenta su cuenta
17     if len(subdomains) >= 2:
18         dom = '%s.%s' % (subdomains[-2],subdomains[-1])
19         if dom not in sld:
20             sld[dom] = 1
21         else:
22             sld[dom] += 1
23     else:
24         if subdomains[-1] not in sld:
25             sld[subdomains[-1]] = 1
26         else:
27             sld[subdomains[-1]] += 1
28
```

Ilustración 22. Script para obtener la cuenta de dominios



Se observa que "*com*" es el TLD más buscado, sin embargo sorprende que "ru" ocupa el tercer lugar, lo que demuestra que los sitios rusos son un objetivo regular de los usuarios de la red.

Ilustración 23. Top de dominio de nivel superior

En cuanto a los nombres de dominio hasta el segundo nivel, hay que destacar que se trata de dominios comunes como *Facebook.com* y *google.com*. Sin embargo hay dominios extraños como *q3537.download.*

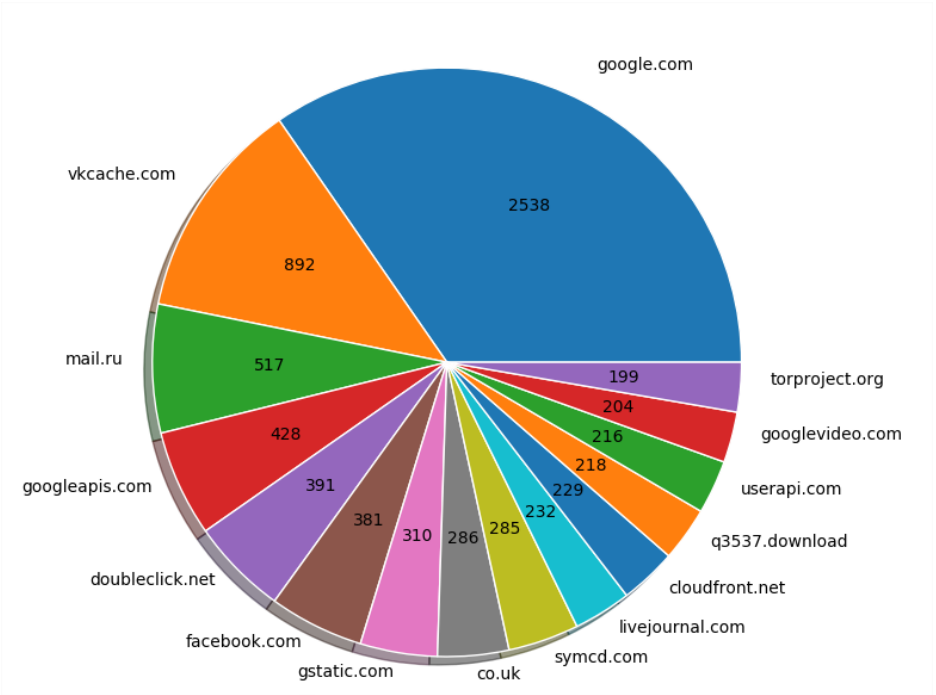


Ilustración 24. Top de dominios hasta segundo nivel

Búsqueda de actividad maliciosa

Como se puede ver en las gráficas anteriores, la mayor cantidad de consultas se hace hacia sitios comunes, sin embargo revisando los dominios, se puede encontrar con sitios extraños o poco comunes. En estos casos se procedió a obtener más información de dichos dominios a través del sitio *Virus Total,* el cual se encarga de analizar archivos y sitios en busca de malware.

El primero en analizarse llamó mi atención, pues es un dominio con una longitud muy grande, algo que es muy poco común. Sin embargo, no fue detectado como un dominio malicioso en Virus Total.

Ilustración 25. Algunos dominios consultados 2

Ilustración 26. Análisis en Virus Total de dominio sospechoso 2

El segundo dominio analizado llamó mi atención pues, además de ser más largo que el resto de los dominios encontrados, termina en "*download*", dando una referencia a que se trata de una página de descargas. En este caso sí fue determinado como un sitio malicioso por 5 escáneres.

```
spartel-marketing.com:  1
ads.realitytraffic.com: 1
momsclubofvenicefl.com: 1
adsmod.com:     1
46852.43483.wymigx.zjwpq1.kskddh.co0qvm.uul0jd.dllm3.www.q3537.download:        1
www.proprealtors.com:   1
eastchesterhousevalue.com:      1
www.teaforte.com:       1
biotecher.com:  1
jallenaluminum.com:     1
```

Ilustración 27. Algunos dominios consultados

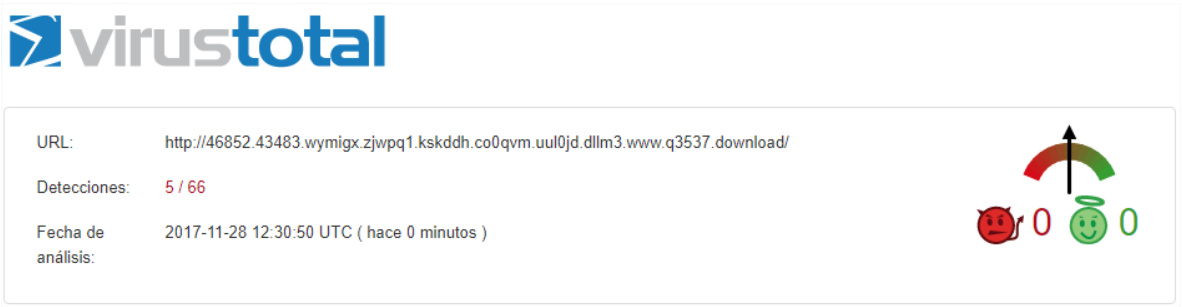


Ilustración 28. Análisis en Virus Total de dominio sospechoso

Puesto que, con el script de python, se determinó que *q3537.download* fue uno de los dominios más veces buscado, llamó mi atención aún más. Se accedió a través de un navegador y, efectivamente, es anunciado por el navegador como un sitio malicioso. Se pasó por alto la advertencia para ver el contenido del sitio y resultó ser una página en idioma chino.

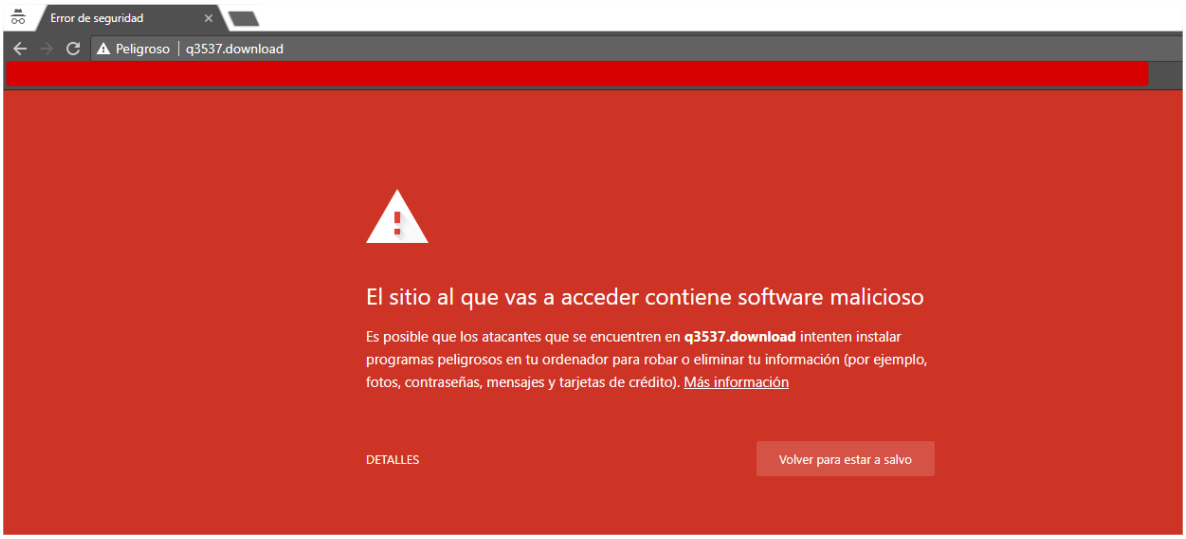


Ilustración 29. Advertencia en navegador sobre sitio malicioso

Ilustración 30. Página principal de sitio malicioso

Se volvió a utilizar el comando *whois*, pues de esta forma se puede encontrar información que podría ser útil para advertir que el sitio está siendo utilizado para difundir malware. Sin embargo eso sale del objetivo del proyecto y se decidió terminar aquí.

```
ESET-2017 > whois q3537.download
Domain Name: q3537.download
Registry Domain ID: D446566-DOWNLOAD
Registrar WHOIS Server: whois.west263.com
Registrar URL: West263.com
Updated Date: 2017-10-20T11:17:00Z
Creation Date: 2016-10-14T07:31:58Z
Registry Expiry Date: 2018-10-13T23:59:59Z
Registrar: Chengdu West Dimension Digital Technology Co., Ltd.
Registrar IANA ID: 1556
Registrar Abuse Contact Email: westdomain@gmail.com
Registrar Abuse Contact Phone: +86.2886263960
Domain Status: ok https://icann.org/epp#ok
Registry Registrant ID: C432199-DOWNLOAD
Registrant Name: wang hai rong
Registrant Organization: wang hai rong
Registrant Street: Feng Xiang Xian Fan Jia Zhai Xiang Zhang Jia Gou Cun Si Zu 047Ha
Registrant Street:
Registrant Street:
Registrant City: Bao Ji
Registrant State/Province: SN
Registrant Postal Code: 721400
Registrant Country: CN
Registrant Phone: +86.13243841569
Registrant Phone Ext:
Registrant Fax: +86.13243841569
Registrant Fax Ext:
Registrant Email: wwwymco@126.com
Registry Admin ID: C432200-DOWNLOAD
Admin Name: hairong wang
Admin Organization: hairong wang
Admin Street: Feng Xiang Xian Fan Jia Zhai Xiang Zhang Jia Gou Cun Si Zu 047Ha
```

Ilustración 31. Información registrada del sitio malicioso

Dominios mexicanos

Se investigó qué tipo de sitios son los que los mexicanos buscan. Esto no es una aseveración definitiva, pues un mexicano podría buscar cualquier tipo de sitios y un extranjero podría buscar sitios mexicanos por cualquier razón. Aun así, no deja de ser un poco interesante y divertido.

Se hizo un filtro de todos los dominios usando las cadenas ".mx" y "mex".

```
ESET-2017 > fgrep .mx dns_queries_sorted >> dominios_mex
ESET-2017 > fgrep mex dns_queries_sorted >> dominios_mex
ESET-2017 > sort dominios_mex | uniq > dominios_mex
ESET-2017 >
```

Ilustración 32. Comandos para filtrar dominios mexicanos

El resultado se muestra a continuación, siendo *"www.correosdemexico.gob.mx"* y *"www.amazon.com.mx"* los dominios más consultados.

Dominio	Consultas	Dominio	Consultas
app.cfe.gob.mx	1	*sevenservice.com.mx*	1
arxbuysell.com.mx	1	*terra.com.mx*	1
bclegalconsulting.com.mx	1	*themexicantaco.org*	1
blogs.eluniversal.com.mx	1	*transexualesmexico.net*	1
cdn.mxpnl.com	6	*unidep.mx*	1
chorizomexicano.biz	1	*vertelenovelasyseries.blogspot.mx*	1
com.mx	1	*whois.mx*	1
folex.com.mx	1	*www.actbc.mx*	1
grupocyc.mx	1	*www.amazon.com.mx*	24
hospedame.mx	1	*www.audioonline.com.mx*	1
ibotana.mx	1	*www.correosdemexico.gob.mx*	31
jimaja.com.mx	1	*www.costco.com.mx*	1
jumpseller.mx	1	*www.dimercom.mx*	1
koolteck.com.mx	1	*www.elfinanciero.com.mx*	1
livinginthemexicancaribbean.com	1	*www.fiestasmexicanas.net*	1
mexashare.com	1	*www.filosofia.mx*	1
mexicolindojewelry.com	1	*www.homedepot.com.mx*	1
nextme.mx	1	*www.i-m.mx*	1
oarsa.mx	1	*www.iscor.com.mx*	1
paginas.seccionamarilla.com.mx	2	*www.lamudi.com.mx*	1
p.ato.mx	1	*www.linio.com.mx*	1
quicksteptelcomgimex.com	1	*www.medicinatradicionalmexicana.unam.mx*	1
s.ato.mx	2	*www.mxdout.com*	1
scientologymexico.org	1	*www.sams.com.mx*	1

segurodeviajero.mx	1	*www.transexualesmexico.net*	1

Análisis de flujos: tcpflow

Un flujo se refiere a todas las comunicaciones establecidas entre la misma dirección IP origen y la misma dirección IP destino. Esto vuelve más fácil el mostrar cuáles son las direcciones IP que más datos transmitieron, así como los protocolos más usados a través del nodo. Para lograr este objetivo, se utilizó el programa *tcpflow.*

Para el primer análisis, se usó el archivo *dump_2017-11-19_00_01.pcap*, el cual tiene un tamaño de 1.9 GB y contiene el tráfico en esa hora en particular.

Se puede observar que la mayor parte del tráfico, por mucho, corresponde al puerto 9001. Como bien recordaremos, este es el puerto configurado para establecer las comunicaciones con otros nodos de la red. Esto nos dice que prácticamente la mayor parte del tráfico que redirigimos no va hacia hosts finales a pesar de ser un nodo de salida.

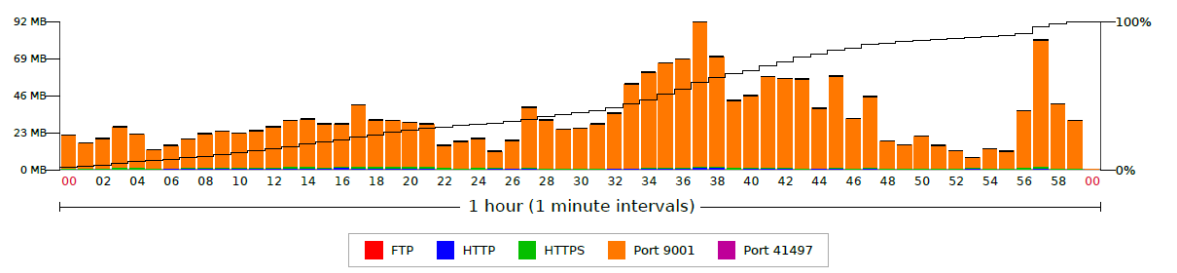


Ilustración 33. Protocolos más usados por minuto

A continuación se puede ver el top de puertos origen y puertos destino. Esto corrobora la información, pues el puerto origen más usado fue el 9001 y en los puertos destino sólo figura el 443 de los puertos bien conocidos.

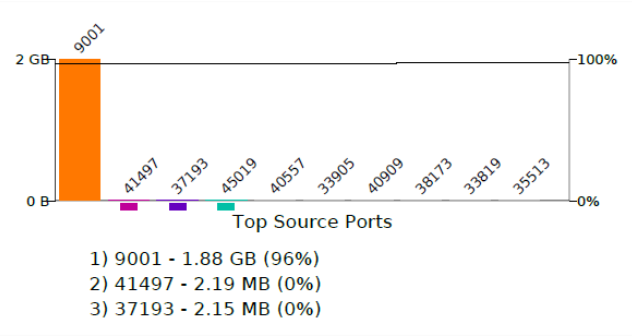


Ilustración 35. Puertos origen más comunes

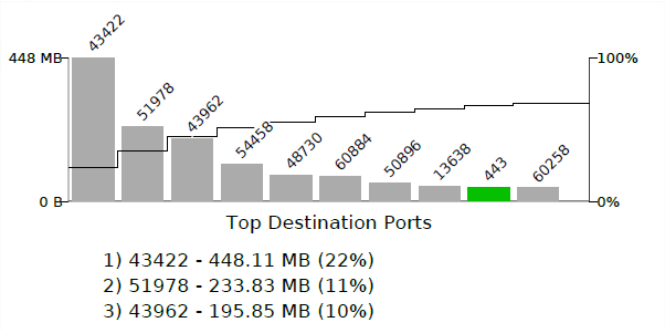


Ilustración 34. Puertos destino más comunes

Se procedió a hacer un filtro sobre una captura diferente, en este caso se eligió el archivo *dump_2017-11-20_06_00_01.pcap* el cual originalmente tenía un tamaño de 4.9 GB y después de filtrar todo el tráfico correspondiente al puerto 9001, resultó un archivo de 220 MB.

Ahora es más evidente la cantidad de tráfico que viaja cifrado con respecto al que viaja no cifrado dentro de la red. Si bien sí hay una gran cantidad que viaja cifrado, es preocupante ver que sigue habiendo un porcentaje alto de tráfico que no lo hace.

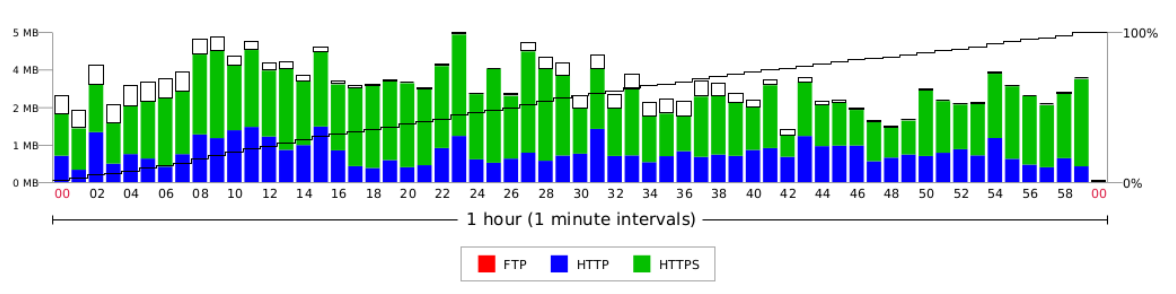


Ilustración 36. Protocolos más usados por minuto 2

Puesto que se eliminó el tráfico correspondiente al enrutamiento de la red, se puede ver cuáles fueron las direcciones que más tráfico generaron durante esa hora en particular, así como una mejor vista de los protocolos mayormente usados.

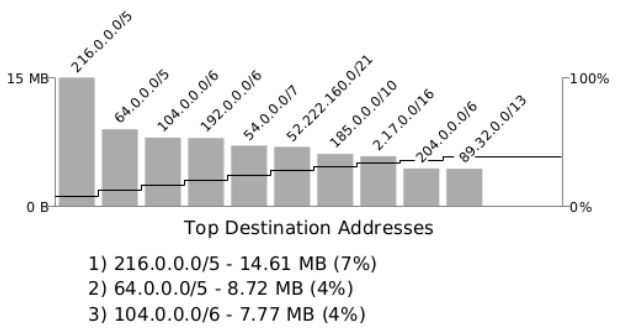


Ilustración 38. Direcciones destino más comunes

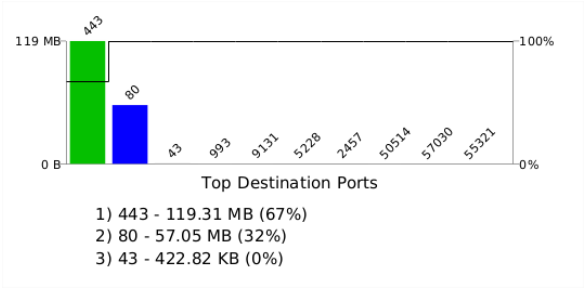


Ilustración 37. Puertos origen más comunes 2

Análisis de flujos: tcpdstat

Para complementar el análisis anteriormente hecho, se decidió usar un programa un tanto antiguo pero muy útil: tcpdstat. Este programa nos ofrece de forma rápida y concisa un desglose de los protocolos más usados.

Esto nos sirve para determinar cuánto tráfico puede recibir o enviar un solo host y ver cuáles son las costumbres de los usuarios dentro de la red. Para este análisis se usaron los archivos *dump_2017-11-19_11_00_01.pcap* (3.8GB), *dump_2017-11-19_18_00_01.pcap* (4.6 GB), *dump_2017-11-20_14_00_01.pcap* (5.5 GB), y *dump_2017-11-20_08_00_01.pcap* (7.3GB). Hay que recordar que estas capturas contienen mucho tráfico correspondiente al puerto 9001 (enrutamiento de Tor).

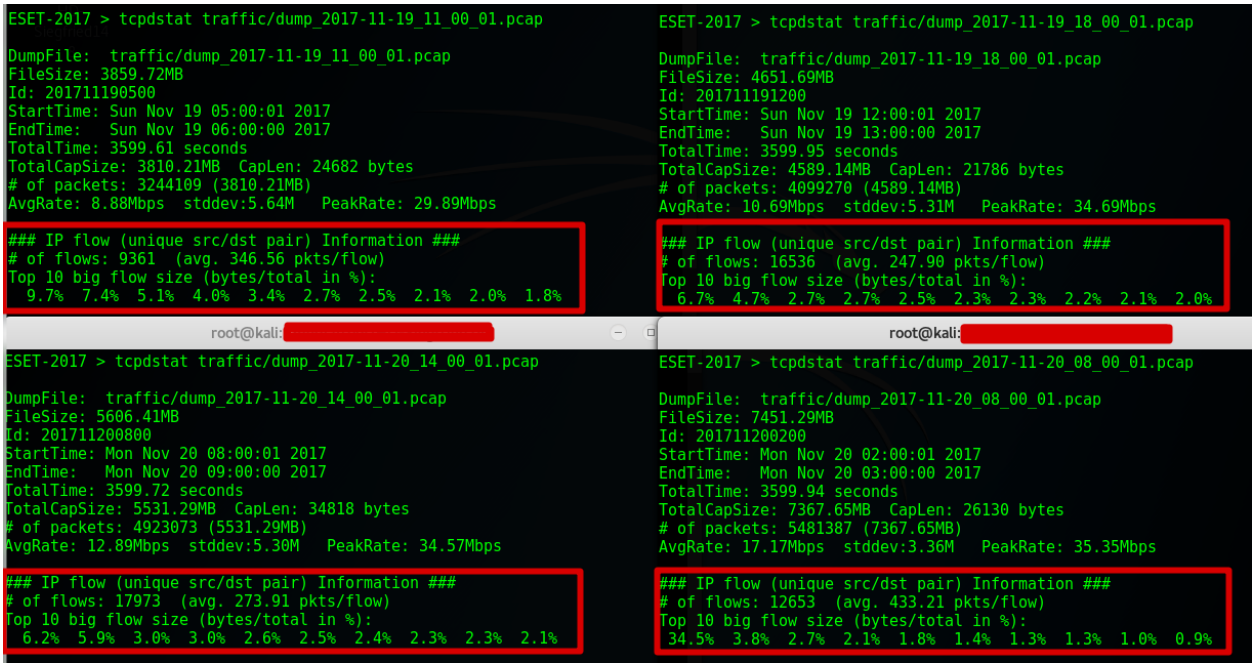


Ilustración 39. Resultado de tcpdstat en cuatro capturas

En este análisis se puede comprobar que, si bien hay una buena parte del tráfico cifrado, aproximadamente el 30% del tráfico de navegación de los usuarios corresponde a HTTP y no a HTTPS. Esto significa que sería muy sencillo para cualquiera que instale y configure un nodo de salida, obtener una gran cantidad de información de los usuarios, como veremos a continuación.

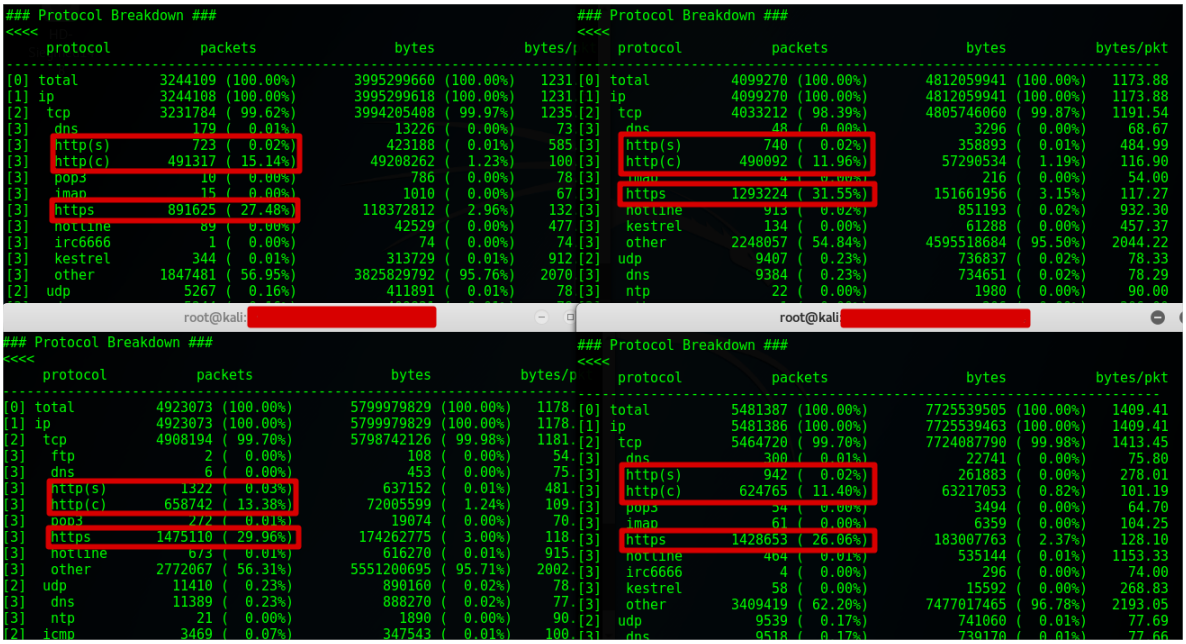


Ilustración 40. Resultado de tcpdstat en cuatro capturas 2

Análisis de tráfico HTTP: Wireshark

El objetivo del trabajo a partir de esta sección, es demostrar qué tan fácil podría ser para un administrador de un nodo de salida obtener información sensible de las personas que utilizan Tor.

Esta parte del análisis se comenzó usando una herramienta muy conocida: Wireshark, debido a su gran popularidad para el análisis de tramas y por todas las características que provee.

A continuación veremos que existen herramientas que automatizan la extracción de datos que, para un atacante o analista de seguridad, son mucho más útiles. Sin embargo, no está de más probar con una herramienta tan popular para demostrar claramente el análisis llevado a cabo.

Se aplicó un filtro para mostrar únicamente el tráfico HTTP y, para mi sorpresa, una de las primeras cosas en aparecer fue un par de credenciales de un formato de registro de un sitio.

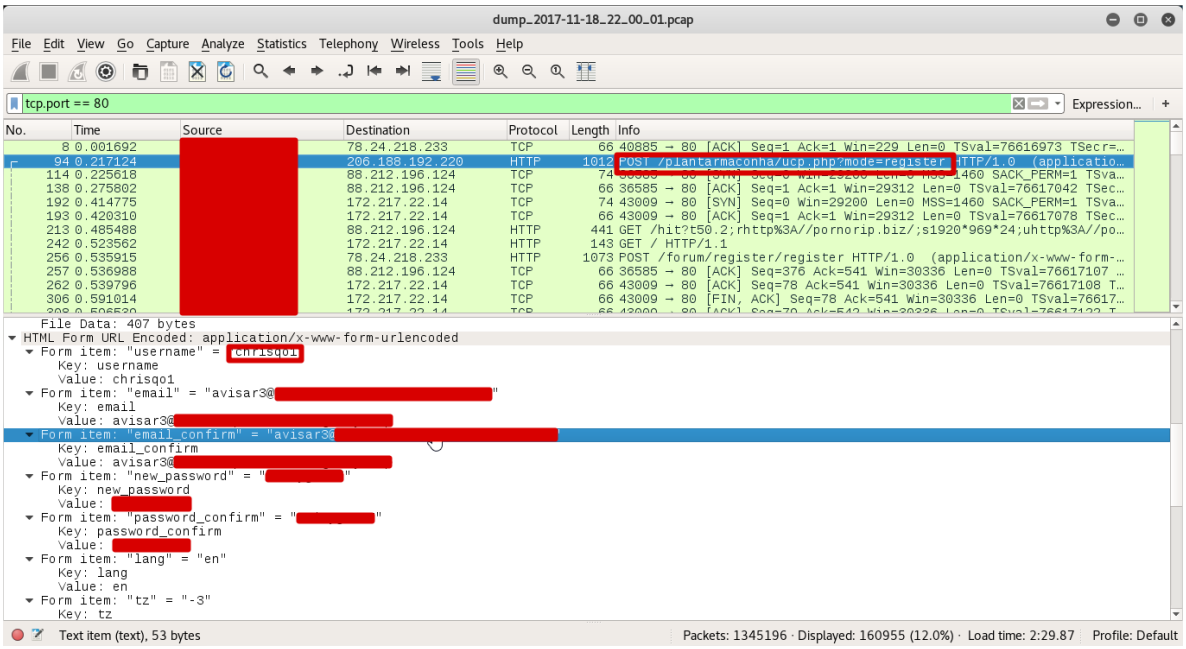


Ilustración 41. Credenciales encontradas en wireshark

En no muchos paquetes posteriores, se encontró con algo que bien podría ser considerada una situación similar. Inicios de sesión (o mejor dicho: intentos de inicio de sesión) a través del archivo xmlrpc.php que se encuentra activo por defecto en los sitios construidos con el manejador de contenidos (CMS) Wordpress.

Claramente se trata de un ataque de fuerza bruta al sitio mostrado, pues se encontraron muchos intentos para ingresar en esta captura de tráfico.

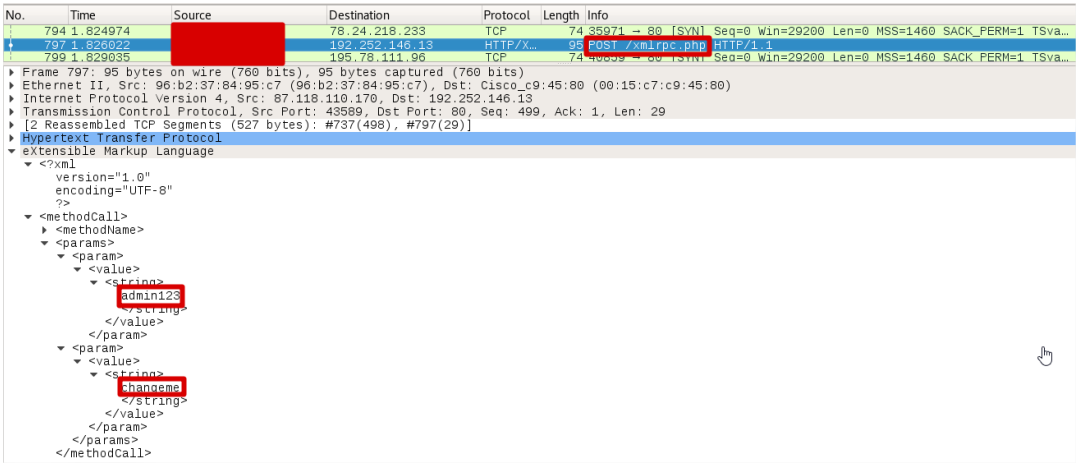


Ilustración 42. Ataque de fuerza bruta en wireshark

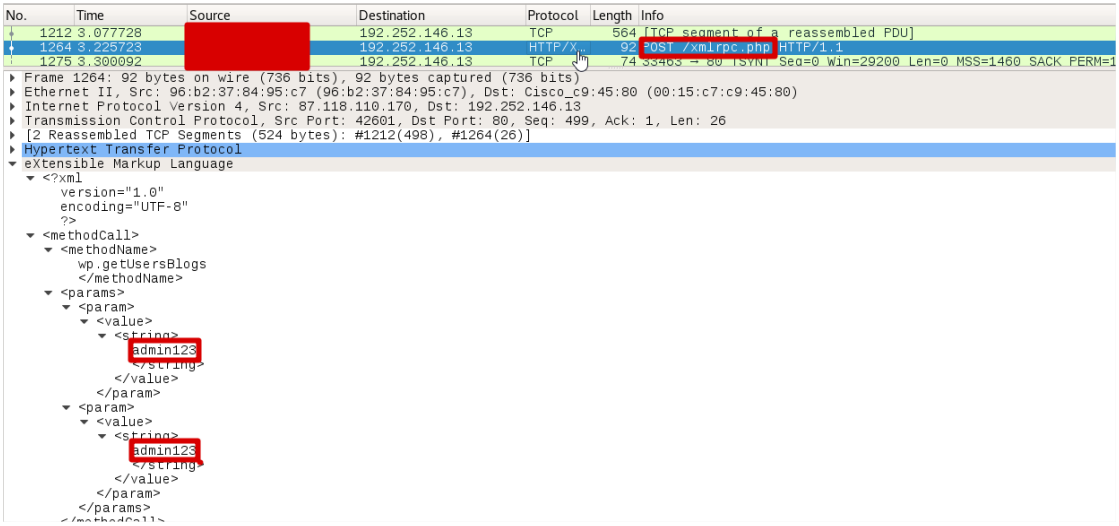


Ilustración 43. Ataque de fuerza bruta en wireshark 2

Además de probar repetidamente con muchas combinaciones diferentes, el atacante también suele cambiar el agente de usuario (HTTP User-Agent) en las diferentes peticiones en un esfuerzo de evitar ser bloqueado por la página atacada, lo cual sólo confirma que se trata de este tipo de ataque.

Ilustración 44. User-Agent en fuerza bruta

Ilustración 45. User-Agent en fuerza bruta 2

Ilustración 46. User-Agent en fuerza bruta 3

Extracción de contraseñas: Pcredz

Puesto que en general son archivos muy grandes, continuar con la búsqueda de información sensible usando wireshark no es viable. Por lo tanto se decidió usar una herramienta que automatice la extracción de este tipo de datos [6] y así poder determinar qué tanta información importante y sensible viaja de forma insegura a través de esta red.

Se ejecutó Pcredz y se seleccionó la captura *dump_2017-11-18_21_00_01.pcap*. Una vez que se ejecuta el programa se registran todas las posibles credenciales en un archivo de texto.

```
./Pcredz -f traffic/dump_2017-11-18_21_00_01.pcap
```

Ilustración 47. Ejecución de Pcredz

Se repitió el proceso con tres capturas más y la situación en general es alarmante. No intenté determinar si eran credenciales válidas, pues eso sale completamente del alcance de este trabajo, sin embargo, se encontraron más de 300 posibles credenciales en sólo cuatro horas.

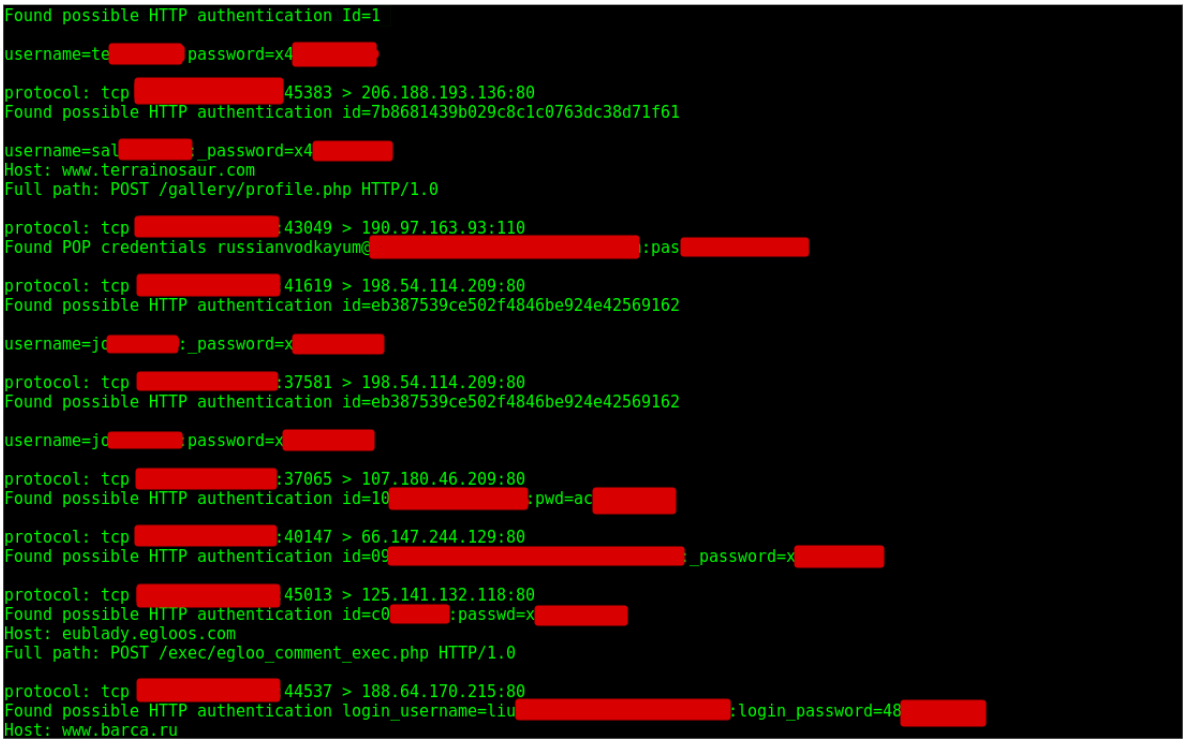


Ilustración 48. Credenciales obtenidas con Pcredz

En esta herramienta también se encontraron ataques de fuerza bruta a un sitio construido con Wordpress. Si bien esto claramente modifica la cifra antes mencionada (más 300 credenciales en cuatro horas), cabe resaltar que no cambia por mucho pues estos ataques están lejos de ser exhaustivos. Esto debido a que se probaron únicamente contraseñas por defecto o muy débiles.

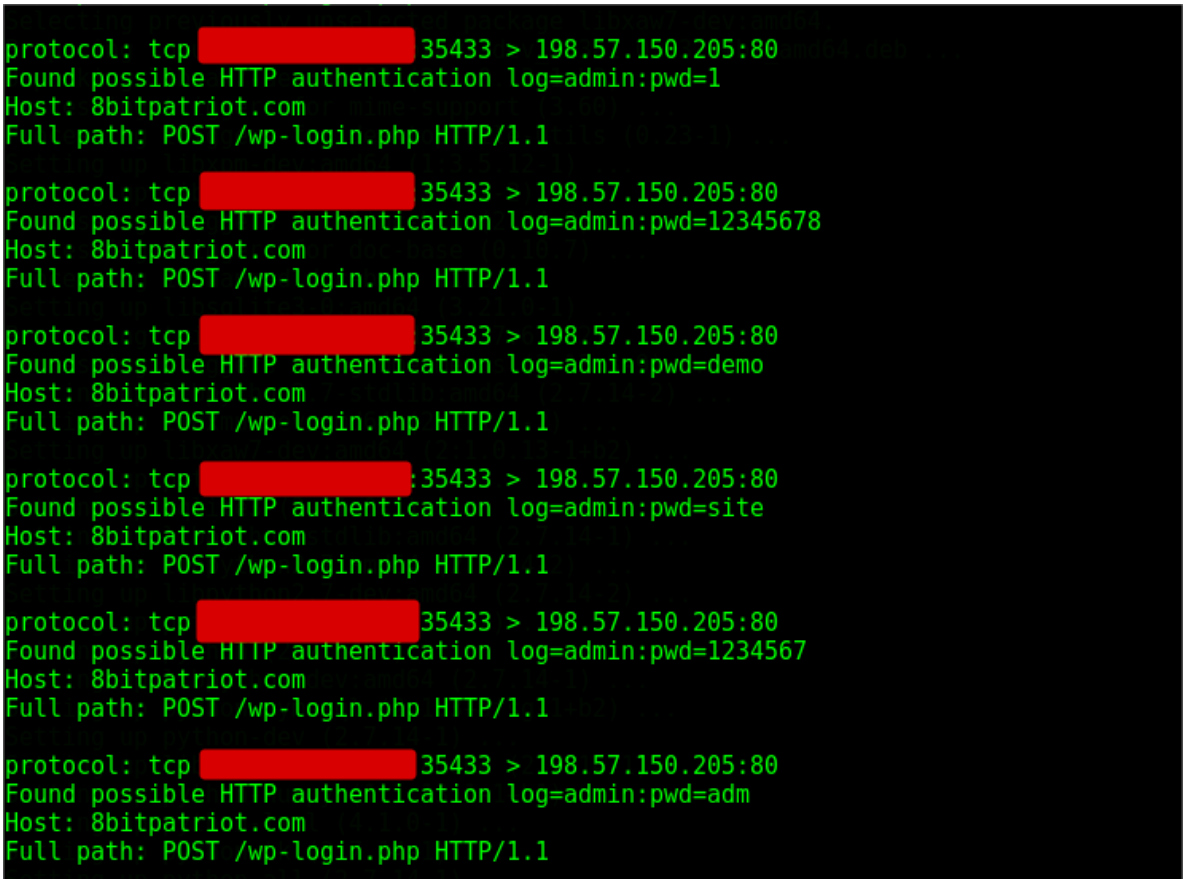


Ilustración 49. Ataque de fuerza bruta encontrado con Pcredz

Extracción de información: Network Miner

Anteriormente ya se extrajo información sensible que viajó a través del nodo, sin embargo, el programa Network Miner [7] facilita la extracción de archivos y provee muchos datos en un formato fácilmente entendible. Por ejemplo, muestra todos los hosts que se encuentran en determinada captura, así como información detallada de ellos.

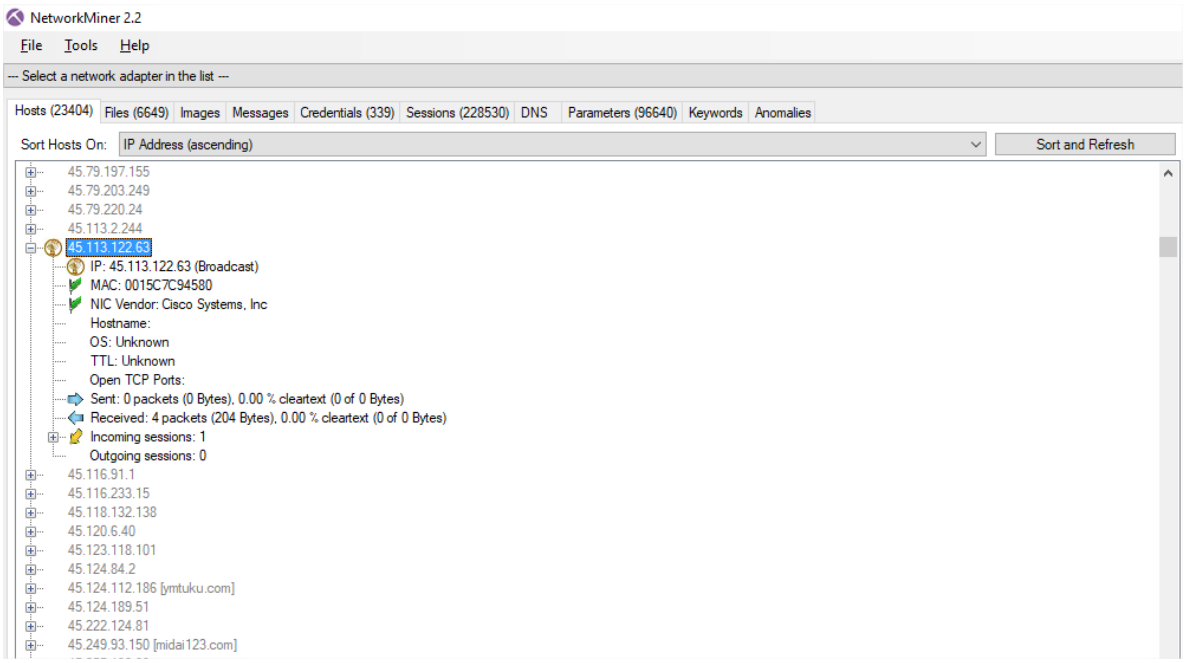


Ilustración 50. Hosts encontrados en una captura

Una de las características que implementa este programa es la extracción de credenciales. No lo hace tan bien como PCredz, pero obtiene mucha información que fácilmente podría utilizar un atacante, como los identificadores de sesión y de esta manera hacer un secuestro de sesión.

Server	Protocol	Username	Password	Valid login	Login timestamp
78.140.137.93 [rawpost.com]	HTTP Cookie	__mcja=172068000.1.2073218301.1510992660.0.151099...	N/A	Unknown	2017-11-18 02:1
78.140.137.93 [rawpost.com]	HTTP Cookie	__mcja=172068000.1.2073218301.1510992660.0.151099...	N/A	Unknown	2017-11-18 02:1
172.82.228.19 [adobetp.chegg.com]	HTTP Cookie	check=true; AMCVS_3FE7CBC1556605A77F000101%40A...	N/A	Unknown	2017-11-18 02:2
172.82.228.19 [adobetp.chegg.com]	HTTP Cookie	check=true; AMCVS_3FE7CBC1556605A77F000101%40A...	N/A	Unknown	2017-11-18 02:2
172.82.228.19 [adobetp.chegg.com]	HTTP Cookie	PHPSESSID=90561p66f221f6v2b1md1b8ti6; C=0; O=0; U...	N/A	Unknown	2017-11-18 02:2
172.82.228.19 [adobetp.chegg.com]	HTTP Cookie	PHPSESSID=041hueoeudehbsd6okcuf4sv61; C=0; O=0; ...	N/A	Unknown	2017-11-18 02:2
172.82.228.19 [adobetp.chegg.com]	HTTP Cookie	PHPSESSID=p5vb2q3jt2pl2oc2k8cg7gjrm0; C=0; O=0; U=...	N/A	Unknown	2017-11-18 02:2
216.137.61.57 [promo.webpayments.closeby.internet.apps....	HTTP Cookie	Set-Cookie: Path=/; Secure; HttpOnly; CloudFront-Key-Pair...	N/A	Unknown	2017-11-18 02:2
94.142.139.248 [skyback.ru]	HTTP Cookie	h3Pdxw7bFm=1; PHPSESSID=69eaade64da43e4254b1d...	N/A	Unknown	2017-11-18 02:2
104.28.8.154 [en.mytubes.xyz]	HTTP Cookie	100640546P=2-1510994328-1510994372-1510994328-; 1...	N/A	Unknown	2017-11-18 02:4
185.68.16.181 [napor.com.ua]	HTTP Cookie	da517e9f69575c62c3bd485786d712f6=3suu2p5pb98obkt...	N/A	Unknown	2017-11-18 02:4
84.200.32.149 [k183.clanzilla.de]	HTTP Cookie	PHPSESSID=mcmdt7015tr11r67gj99h1jbb3	N/A	Unknown	2017-11-18 02:4
84.200.32.149 [k183.clanzilla.de]	HTTP Cookie	PHPSESSID=mcmdt7015tr11r67gj99h1jbb3; wordpress_te...	N/A	Unknown	2017-11-18 02:4
84.200.32.149 [k183.clanzilla.de]	HTTP POST	ma[redacted]	9[redacted]	Unknown	2017-11-18 02:4
84.200.32.149 [k183.clanzilla.de]	HTTP Cookie	PHPSESSID=mcmdt7015tr11r67gj99h1jbb3; wordpress_te...	N/A	Unknown	2017-11-18 02:4
104.28.9.154 [en.mytubes.xyz]	HTTP Cookie	100640546P=5-1510994328-1510995177-1510994328-; 1...	N/A	Unknown	2017-11-18 02:5
5.35.172.152 [www.24video.in]	HTTP Cookie	raw_ref=https%3A%2F%2Fduckduckgo.com%2F; JSESSIO...	N/A	Unknown	2017-11-18 02:5

Ilustración 51. Información sensible de una captura

Se hizo una extracción de todas las cabeceras HTTP y en particular de los agentes de usuario (HTTP User-Agent), con el fin de encontrar actividad maliciosa. Esto en razón de que pueden encontrarse campañas de atacantes que se diferencian mediante esta cabecera. Asimismo, es posible encontrar actividad de malware que tenga un agente de usuario muy particular.

En general los agentes de usuario encontrados son bastante comunes (sin que esto signifique que sea tráfico no malicioso). Por otro lado, fue común encontrar el agente "*AfD-Verbotsverfahren JETZT!*", que no corresponde a una campaña de ataques o de malware, sino a una campaña para enviar un mensaje político. [8]

User-Agent	AfD-Verbotsverfahren JETZT!	4873378	TCP 37215	109.237.138.16 [muscle-car-parts.de] [payclever.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873380	TCP 41305	81.169.145.84 [tanzkapelle-calypso.de] [schloss-brunn.de] ...	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873381	TCP 33937	81.169.145.74 [beratungsmappe.de] [tli-tax.de] [vandevoor...	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873399	TCP 40711	81.169.145.162 [claudia-fiergolla.de] [van-der-sman.de] [w...	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873415	TCP 45163	81.169.145.84 [tanzkapelle-calypso.de] [schloss-brunn.de] ...	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873419	TCP 34601	109.237.140.22 [recruiting-erfolg.de] [al-bayati-gmbh.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873420	TCP 45515	81.169.145.70 [www.emh-service-gmbh.de] [mathe-kolleg....	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873421	TCP 36729	81.169.145.77 [e-ffect.de] [dsenergy.de] [walddorf.de] [ord...	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873443	TCP 37887	217.160.231.121 [rechtsgelenkt.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873444	TCP 38813	80.237.132.102 [xn--manthei-rechtsanwltin-m2b.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873461	TCP 39017	92.51.164.231 [keepyoureyesopen.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873462	TCP 34977	89.31.143.1 [xn--hrgertetrger-kcbe4x.de] [rabenhofer.de] [d...	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873471	TCP 42679	94.247.41.101 [www.abakus-musik.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873472	TCP 36805	37.17.224.10 [forrodedomingo.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873502	TCP 38539	109.237.130.96 [eddys-autoglas.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873521	TCP 43521	185.21.102.43 [club-imperia-hannover.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873565	TCP 37661	217.160.0.84 [baeckerei-schwerdtner.de]	TCP 80
User-Agent	AfD-Verbotsverfahren JETZT!	4873602	TCP 44567	46.252.18.36 [feuerfootbag.de] [feuer-footbag.de]	TCP 80

Ilustración 52. Agente de usuario de una campaña política

Del mismo modo se extrajeron las peticiones GET hechas en diferentes capturas. Fue relativamente rápido encontrar un escaneo de vulnerabilidades hecho con *Acunetix*, que intenta ver si un servidor es vulnerable a *SQL injection* y a *local file inclusion*, entre otras vulnerabilidades.

Filter keyword: GET — Case sensitive — ExactPhrase — Any colu

Para...	Parameter value	Destination host
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/T1rKRwmBGgt_gtWI?as 0001&ch=20150401&pm=custom&utm_campaign=WEIXIN&utm_content=Phantom4MB&utm_medium=cpc&utm_source=../../../../../../../../../etc/passwd%00.jpg	52.204.43.15 [click.dji.com] [s
GET	/	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/T1rKRwmBGgt_gtWI?as 0001&ch=20150401&pm=custom&utm_campaign=WEIXIN&utm_content=Phantom4MB&utm_medium=cpc&utm_source=/../..//../..//../..//../..//../..//etc/passwd%00.jpg	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/T1rKRwmBGgt_gtWI?as 0001&ch=20150401&pm=custom&utm_campaign=WEIXIN&utm_content=Phantom4MB&utm_medium=cpc&utm_source=../..//../..//../..//../..//../..//../..//../..//../..//etc/passwd	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/T1rKRwmBGgt_gtWI?as 0001&ch=20150401&pm=custom&utm_campaign=WEIXIN&utm_content=Phantom4MB&utm_medium=cpc&utm_source=../.../../../../../.../../../../../.../../../../../.../../../../../.../../../etc/p...	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/AIVv38GhzjgbNBgOANw	52.204.43.15 [click.dji.com] [s
GET	/assets/site/icons/icn-nav-cart-n-v01.svg	52.222.171.54 [c.cheggcdn.c
GET	/T1rKRwmBGgt_gtWI	52.204.43.15 [click.dji.com] [s

Ilustración 53. Escaneo de vulnerabilidades

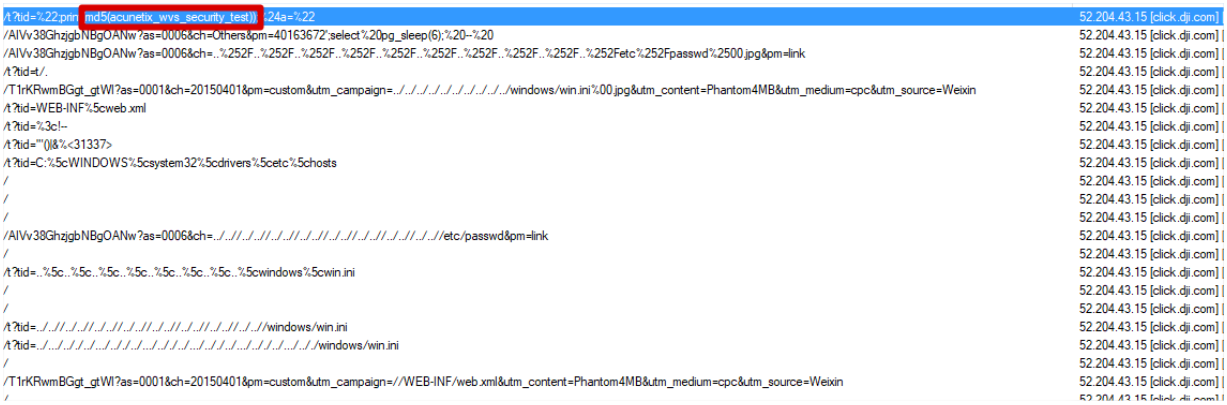


Ilustración 54. Escaneo de vulnerabilidades 2

Conclusiones

La red Tor es una red ampliamente usada que ayuda a muchas personas en diferentes situaciones que necesitan mantener su privacidad un poco mejor protegida. Se demostró que no es difícil instalar y configurar un nodo Tor y que, si alguien quisiera apoyar con un poco de ancho de banda para ayudar a estas personas, fácilmente puede lograrlo. Además, podría ser útil implementarlo para los investigadores de seguridad, ya que es un buen método para hallar actividad maliciosa en la red, así como campañas de ataques y, con algo de suerte, nuevas familias de malware.

En cuanto al uso que se le da a la red, si bien no se puede determinar el tráfico dirigido a los hidden services, está claro que la red está lejos de ser exclusiva para eso. Se encuentra todo tipo de sitios que pueden ser visitados en cualquier navegador, en cualquier lugar. Eso sí, las visitas a estos sitios son de todo tipo: escaneos de vulnerabilidades, ataques de fuerza bruta, campañas políticas, descargas de malware y navegación común.

El punto más importante de toda la investigación trata sobre la privacidad en esta red. Usar Tor está muy lejos de ser sinónimo de privacidad si no se hace con el cuidado adecuado. Quedó demostrado que en un nodo de salida administrado por un atacante o un gobierno, fácilmente se pueden extraer todo tipo de archivos, contraseñas, cookies y cualquier tráfico que no se encuentre cifrado. Esto debe convencer al lector que, si planea usar Tor para mantenerse privado, debería tener cuidado y asegurarse de viajar sólo a sitios seguros usando las versiones seguras de los protocolos, como por ejemplo HTTPS, SSH o IMAPS.

Fuentes

[1] https://www.torproject.org/about/overview.html.en

[2] https://blog.daknob.net/running-a-tor-exit-node-for-fun-and-e-mails/

[3] https://trac.torproject.org/projects/tor/wiki/doc/ReducedExitPolicy

[4] https://atlas.torproject.org/

[5] https://metrics.torproject.org/bandwidth.html

[6] https://github.com/lgandx/PCredz

[7] http://www.netresec.com/?page=NetworkMiner

[8] https://marius.bloggt-in-braunschweig.de/2017/10/09/politische-kampagnen-aus-dem-tor-netz/