Universidad Autónoma de Madrid
Escuela politécnica superior

Grado en Ingeniería informática

Trabajo Fin de Grado

TECNOLOGÍAS PARA LA PROTECCIÓN CONTRA LA PRIVACIDAD Y EL ANONIMATO.

Autor: Francisco Andreu Sanz
Tutor: David Arroyo Guardeño
Ponente: David Arroyo Guardeño

Enero 2018

TECNOLOGÍAS PARA LA PROTECCIÓN CONTRA LA PRIVACIDAD Y EL ANONIMATO.

Autor: Francisco Andreu Sanz
Tutor: David Arroyo Guardeño
Ponente: David Arroyo Guardeño

Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Enero 2018

Resumen

Resumen

La tecnología ha tenido un gran impacto social en los últimos años, y cada vez va formando más parte de nuestras vidas.
Las grandes empresas de software lo saben, y muchas de ellas tienen objetivos de dudosa moralidad, como la recopilación de información personal para poder ofrecernos unos servicios más apropiados para cada usuario. Entre esta información podemos encontrar alguna tan íntima como la dirección, el número de teléfono o incluso los lugares que frecuentamos, y basta con un inicio de sesión en una red social para que dichos datos fluyan por la red. Esto mismo ha llevado a muchos el cuestionarse hasta qué punto nuestro derecho a la privacidad se está viendo comprometido y, pese a que mucha otra gente no de la importancia que merece a este tema, lo cierto es que no son pocas las herramientas que han aparecido para ayudarnos a que podamos navegar por internet de una forma *menos pública*. De ahí nace la motivación de este proyecto, de proporcionar varias formas de realizar tareas en Internet de la forma más anónima posible, a partir de una única herramienta. Ésta herramienta ha sido codificada en *Python*, y cuenta con varios paquetes de módulos, cada uno con una funcionalidad distinta.
El primer paquete cuenta con varias funcionalidades relacionadas con el envío y recepción de correos electrónicos. Un módulo de ese paquete nos permite hacer uso de un *remailer* para proteger lo máximo posible la identidad del emisor de dicho correo, otro módulo que permite enviar correos pero esta vez permitiendo al usuario *ofuscar* el contenido del mensaje y por último una bandeja de entrada temporal que permite ver en tiempo real los correos que envían a dicha cuenta.
El segundo paquete está relacionada con la automatización de procesos de registro, inicio de sesión e incluso ciertas acciones en páginas web. Para ello se ha hecho uso de *web-scraping* en redes sociales, diarios y foros de discusión. Permite, de esta forma, utilizar cuentas de correo volátiles con el fin de *anonimizar* la identidad al ingresar dichas páginas. Además, se deja a disposición del usuario el poder utilizar el *servicio TOR* para así mejorar aún más la privacidad. Las cuentas de usuario generadas en el registro de las páginas son almacenadas en una base de datos local *encriptada*.
El último paquete está relacionado con la posibilidad de *enmascarar tráfico web*, uso de *VPNs*, de *proxies*, y más funcionalidades que están descritas en este documento.
Cabe destacar que el presente proyecto usa una metodología *open source* y permite fácilmente la inclusión de módulos adicionales para aumentar aún más su funcionalidad.

Palabras Clave

remailer, ofuscar, web-scraping, anonimizar, servicio TOR, encriptada, enmascarar tráfico web, VPN, proxies, open source.

Abstract

Technology has had a major social impact in the last decades, and has become part of our daily life.
The most influential software companies are conscious about this, and they have ethically questionable objectives, like a complete collection of personal information in order to offer us better and customized services. Amongst this information there is sensible data, such as our personal address, our telephone number or even the places we usually visit. Logging in a page is the only thing we need to do in order to let our private information spread on the internet.

This condition has led many people think to what extent is our privacy compromised. Many others do not concern this topic but, even so, there has been a recent increase in the number of tools which let us navigate on the web in a less-public way. This project is born from this idea, the idea of providing many simple and daily functionalities in the most anonymous way possible. It has been coded in Python and it is composed by several source packages, eac h one with different purposes.
The first package of the tool has three main functionalities, all of them related to the seng and reception of e-mails. The first functionality lets us make use of a *remailer* which will protect the identity of the sender. Another module provides us a service of sending e-mails, with an option of *obfuscating* the content of the message and. Lastly, a real-time volatile inbox, which lets us see from console each e-mail that arrives to our temporary account.
The second package is related with the automation of signing up, logging in and other functionalities in web pages, such as newspapers, social networks, and discussion forums. In order to deal with these tasks it has been required to make use of *web-scraping*. This package basically lets us to sign in a page using a volatile e-mail, without the neccessity of using our personal one with the purpose of *anonimyzing* our identities. It also has an option of using the *TOR service* in these proccesses to hide our identity even more. Accounts generated in this package is stored in an *encrypted* database.
Last package is related with the possibility of *masking web traffic*, using *VPNs*, *proxies*, and more functionalities which will be described in this document.
Finally, there is a noteworthy effort of developing this project with an *open source* metodology, letting the final user include more modules in it easily.

Key words

remailer, obfuscate, web-scraping, anonimyzing, TOR service, encrypted, masking web traffic, VPN, proxies, open source.

Agradecimientos

Me gustaría agradecer el apoyo que me han brindado mis padres, mi hermana y mi novia cada día. Por no dejar de animarme cuando lo necesitaba y por saber aguantarme cuando no podía más, gracias de corazón.

A mis compañeros de la carrera, en especial a Fran, Iván, Juan y Pablo entre muchos otros, porque sin ellos este camino habría sido muy diferente, y desde luego no lo habría afrontado con las mismas ganas e ilusión.

Por último, y no por ello menos importante, a cada uno de los profesores del grado, que han aumentado día a día mi motivación por aprender.

Entre todos ellos quiero dedicar una especial mención a mi tutor David, por depositar la confianza en mí de poder realizar este trabajo, y por incentivarme aún más a aprender sobre lo que más me apasiona, la seguridad informática.

Índice general

Índice de Figuras

Índice de Tablas

1 Introducción

1.1. Motivación del proyecto

El derecho a la privacidad en Internet es algo que todo usuario debería valorar y, por desgracia, el gran público no le da la importancia que debería.

No son pocas las noticias que están apareciendo últimamente sobre empresas como Google relacionadas con la invasión a la privacidad. Ésto, en gran parte, se ha visto incrementado debido a la llegada de los *smartphones* al mercado(algo relativamente reciente, hace alrededor de 10 años). El poder llevar en nuestro bolsillo todo un ordenador tiene el inconveniente de que grandes empresas como las anteriormente mencionadas pueden tener acceso a información en tiempo real de nosotros, como por ejemplo a la hora a la que nos levantamos, la localización de nuestra propia casa e incluso la ubicación real en todo momento(y sí, de poco sirve deshabilitar la ubicación por GPS en tu smartphone pues también la pueden averiguar mediante el inicio de sesión en una red WiFi).
Por otro lado está el tema de las redes sociales. Con el auge de Facebook, Instagram y Twitter, gran parte de la población(en el caso de Norteamérica, casi dos terceras partes) tiene perfil propio en la plataforma Facebook.

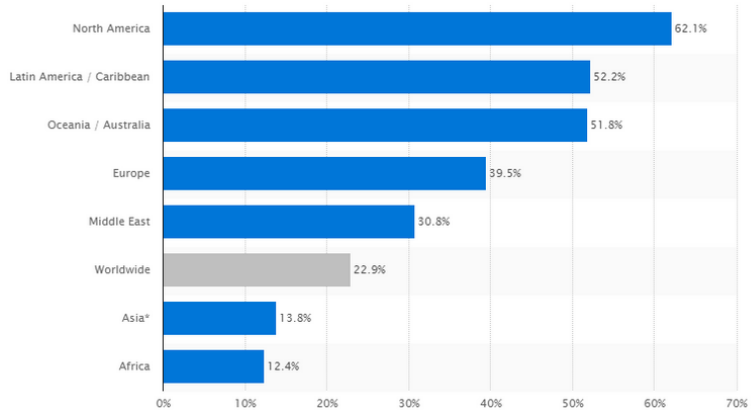

Figura 1.1: Porcentaje de población con perfil en Facebook

Ésto de por sí no es un dato negativo, el problema viene cuando para realizar un registro en una página (como por ejemplo, la web de un diario), la forma más sencilla es conectando con tu perfil personal de Facebook. Ésto causa que, al interaccionar con dicha página(ya sea publicando un comentario, o cualquier tipo de actividad), te arriegues a que aparezca tu nombre real, con todo lo que ello conlleva. Desde este punto, saber todo acerca de ese usuario es tan sencillo como buscar en Google su nombre completo y entrar a su perfil de Facebook, donde aparecen fotos, su dirección, entre otros.
En otros casos el iniciar sesión con la cuenta de Google ó Facebook sirve para que dichas empresas conozcan mejor tus gustos y hobbies, para así ofrecerte publicidad a medida.

De aquí nace la verdadera motivación de este proyecto. La lucha contra la privacidad en la red es algo que hemos ido perdiendo poco a poco, pero mediante el uso de tecnologías para anonimizarse como las que serán vistas en éste documento se pretende erradicar o, al menos, mitigar éste problema.

Parte de la motivación también reside mi interés en el ámbito de la seguridad informática que, desgraciadamente, no está demasiado presente en el temario desarrollado en la carrera. Sin embargo, es un tema de suma importancia y además sirve para poner en práctica metodologías y lenguajes estudiados en el grado.

En definitiva, el proyecto abarca un software modular compuesto de varias herramientas funcionales por sí mismas y donde además el requisito principal es la seguridad del sistema(*security-by-design*).

1.2. Objetivos y enfoque

Principalmente se pretenden lograr dos objetivos fundamentales en este proyecto.

El primero es el de hacernos conocedores más a fondo de las diferentes vías a la hora de anonimizarnos en Internet, las variadas herramientas que pretenden conseguir éste objetivo (así como las que pretenden identificar a un usuario), las disparidades entre anonimato y privacidad... En conclusión, realizar una investigación exhaustiva sobre la privacidad en la red.

Por otro lado, y quizá el objetivo más importante, es el de poner en práctica los conocimientos adquiridos en la investigación anteriormente dicha. En este caso se ha diseñado, desarrollado y probado una herramienta con numerosas y diversas funciones, cuyo principal propósito es el de garantizarnos una experiencia lo más anónima posible en todo momento.

1.3. Metodología y plan de trabajo

Éste documento se organiza de la siguiente manera:

- Estado del Arte: El segundo capítulo explica todos y cada uno de los conceptos de los que trata este proyecto, es decir, el término privacidad, anonimato y la importancia de los mismos hoy en día. Además, se mostrarán ejemplos de herramientas y metodologías para anonimizarse.

- Análisis: El capítulo tres consta de la serie de requisitos, definidos según los objetivos deseados en las aplicaciones finales y delimitados por el alcance del proyecto.La funcionalidad de la herramienta desarrollada se resume tanto en el catálogo de requisitos como de casos de uso.
- Diseño: Este capítulo trata con detalle la fase de diseño, teniendo en cuenta la estructura de la aplicación y el flujo de navegación de la misma
- Desarrollo: En el quinto capítulo se encuentra explicado el método de desarrollo, las librerías utilizadas, los lenguajes en los que está programada la herramienta, las características Software del equipo de desarrollo y el porqué de dicha elección.
- Integración, pruebas y resultados: Aquí se tratan las pruebas unitarias realizadas, así como los resultados de las mismas y cómo se han integrado todos los módulos en la aplicación final.
- Conclusiones/Trabajo futuro: Por último, en este capítulo resumimos las conclusiones de la aplicación y el futuro trabajo que se requeriría para que la herramienta vaya creciendo.

2

Estado del arte

2.1. Privacidad y anonimato: conceptos

La privacidad es un concepto bastante complejo, una palabra con tantas acepciones que en algunos casos puede resultar engañosa o, incluso, sin sentido alguno. Los temas donde se trata éste concepto van desde las leyes y derechos hasta la tecnología, pasando por campos tan ambiguos como la filosofía.

Por otro lado, el contexto en el que suele ser utilizada va desde los ajustes de un navegador hasta uno de los debates más importantes sobre el desarrollo de la sociedad.

En resumen, los usos del concepto de privacidad abarcan un número incalculable de temas y es por ello por lo que dicho término es difícil de definir. Sin embargo, en este proyecto nos atañe el uso relacionado con la red, y aquí se puede definir como el control de la información que posee un determinado usuario que se conecta a Internet, interactuando con diversos servicios en línea con los que intercambia datos durante la navegación.

Cabe mencionar que muchos de los usuarios que navegan día a día no son realmente conscientes de los datos personales que circulan por la red.

La privacidad no debe confundirse con el **anonimato en la red**. Éste último término refiere a aquellas acciones destinadas a garantizar que el acceso a la red se efectúa de forma que no se conoce quien realiza la conexión.

Por último conviene hablar de otra expresión que aparecerá también muy frecuentemente en este proyecto, y es el de **ofuscación**.

En su sentido más abstracto, la ofuscación es la producción de ruido modelado en una señal existente con el objetivo de hacer una recopilación de datos más ambigua, confusa, difícil de *explotar* y, por ende, menos valiosa.

2.1.1. Procedimientos para incrementar el anonimato

Este subcapítulo tiene como objetivo explicar los distintos tipos de procedimientos para lograr un mayor grado de anonimato en la red.

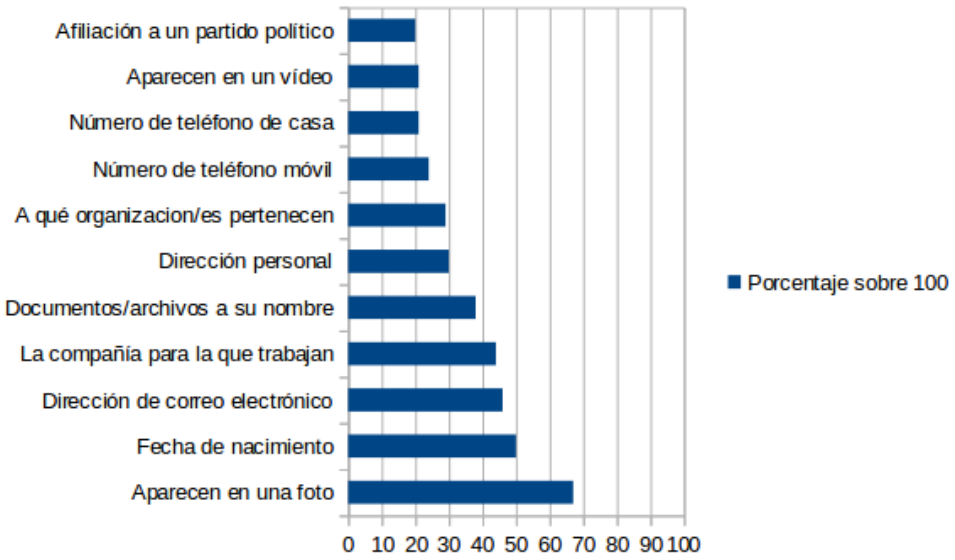

Figura 2.1: Resultados del estudio en Pew Research Center's and American Life Project hecho en Julio de 2014

Antes de nada conviene diferenciar dos conceptos comúnmente confundidos como son el de pseudoanonimato (*pseudonymity* en inglés) y anonimato.

El primero de ellos refiere al hecho de usar un pseudónimo con el fin de camuflar una identidad real. Su significado literal según su etimología es "llamado engañosamente". Por otra parte, el segundo, cuyo origen etimológico significa "sin nombre", refiere cuando no hay información identificable a nada ni nadie.

Por ende, la principal diferencia entre ambos términos radica en que mientras en el anonimato la identidad es totalmente desconocida, en el pseudoanonimato se aprovecha el hecho de utilizar un pseudónimo para esconder una identidad real. En éste proyecto se ha trabajado con métodos tanto para lograr un anonimato como para lograr un pseudoanonimato.

Una vez aclarados sendos términos, procedemos a listar las diferentes vías para lograr el anonimato:

- Anonimato de emisor: Este tipo consiste en un origen que efectúa un mensaje a un determinado receptor, y el emisor no puede ser reconocido por ningún observador.
- Anonimato de receptor: En este caso, al contrario del anterior, es el receptor el que no puede ser identificado por el observador.
- Anonimato de comunicación: Este capítulo trata con detalle la fase de diseño, teniendo en cuenta la estructura de la aplicación y el flujo de navegación de la misma

2.2. Hechos relevantes con respecto a la seguridad en la red

Conviene empezar el capítulo recalcando que Internet no fue concebido como un protocolo de comunicación seguro. Es por esto que a lo largo de su historia han ocurrido varios sucesos que han puesto en riesgo (por diversos motivos) la privacidad del usuario en la red.

Podemos marcar como primera incursión histórica con respecto a la seguridad un libro publicado por Jon Von Neumann en el año 1949 llamado *The Theory of Self Reproducing Automata*. Dicha publicación sirvió como base para el desarrollo de los primeros **virus informáticos**.

El primer virus que causó un gran impacto social fue el lamado **virus Creeper**, el cual era un programa experimental autoreplicante creado por **Bob Thomas** diseñado con fines experimentales y que no causaba un daño real entre las máquinas en las que se iba moviendo. Realmente fue desplazándose de ordenador en ordenador alrededor de toda la red ARPANET (la precursora de lo que es a día de hoy internet).

Con el objetivo de acabar con dicho virus apareció **Reaper**, lo que hoy día llamaríamos antivirus pero que en realidad no era más que otro programa autoreplicante cuyo propósito era la eliminación de Creeper en cada uno de los sistemas donde se encontraba instalado.

En 1973 Robert Metcalfe, un trabajador de ARPANET y el cuál fundó 3Com (uno de los fabricantes de redes informáticas más importantes), advertía que una incursión a la red interna desde el exterior era algo extremadamente sencillo y, de hecho, son atribuidas durante la década de los 70 varias intrusiones a la red por parte de estudiantes de secundaria. Durante esta etapa no se produjeron descubrimientos destacables con respecto a la seguridad informática. De hecho, en el año 1978 un grupo de científicos propusieron un proyecto de cifrado de paquetes TCP/IP pero encontraron muchas trabas, algunas de ellas incluso por la Agencia de Seguridad Nacional. Por ello, dicho proyecto (que bien podría haber marcado otro camino en la historia de la seguridad en la informática) fue abandonado.

En 1981 apareció el segundo virus reconocido a nivel mundial, el llamado **Elk Cloner**. Atacaba computadoras Apple II, aunque su único propósito era el de reproducirse en otros dispositivos y no efectuaba ningún daño propiamente dicho. Uno de los datos más impactantes es que fue diseñado por un joven de 15 años. Su propagación era mediante el disquete. Este virus sentó la base para los siguientes que fueron apareciendo, los cuales contendrían todo tipo de código destructivo (robo de datos, manipulación de los mismos, destrucción de software y hardware...) y se propagarían por más medios, como el correo electrónico e Internet. Debido a la aparición de los numerosos softwares maliciosos en esta época fueron apareciendo empresas que proporcionaban herramientas para proteger los equipos, los **antivirus**.

En el año 1983, se hizo mandatorio que los usuarios de la red ARPANET utilizasen el protocolo TCP/IP. Éste hecho estableció un estándar en la comunicación entre redes y favoreció la aparición de la World Wide Web. Éste fue además el año en el que se utilizó por primera vez el término *virus informático* en una tesis académica, dirigida por Fred Cohen.

En 1986 se aprueba una ley llamada La *Ley De Fraude Y Abuso Cibernético* la cual aparece como contramedida al virus más dañino hasta la fecha como fue Brain, el primer virus compatible con máquinas IBM. La ley defendía a los usuarios del robo de datos, del acceso a la red no autorizado y demás delitos relacionados con la tecnología.

1987 fue un año de grandes avances en el campo de la seguridad informática en todo el mundo. En primer lugar, se produjo la primera eliminación total de un virus dañino a gran escala por parte de Bernd Fix. Por otra parte Andreas Lűning y Kai Figge lanzan al mercado el primer antivirus diseñado para la plataforma Atari. Además, aparece la primera empresa estadounidense de antivirus, creada por John McAfee. Por último, añadir que aparecen los primeros antivirus basados en heurísticas, como son Anti4us y Flushpot.

Los tres años siguientes siguieron apareciendo compañías que velaban por la seguridadm como Symantec, la cual lanzó el conociod antivirus Norton en el año 1991. Sin embargo, ésto no hizo que el número de robos de información cesara. Al contrario, pues con la aparición del primer navegador web surgieron nuevas formas de ataque y *phising*. Asimismo surgieron los primeros ataques de denegación de servicio.

En el año 1996, con la aparición del complemento de navegador Flash(que permite la reproducción de vídeo y música) surgen por ende nuevas formas de ataques, debido a las numerosas vulnerabilidades del plugin. El correo electrónico, que cada vez se encuentra más vigente, tam-

bién permite recibir correspondencia con un objetivo de robar información personal. En éste año también aparece el primer virus creado para el sistema Linux, llamado Staog.

El año 2000 el número de gusanos informáticos que se encientran en equipos domésticos es desproporcionado. Asimismo, en ésta década los ciberdelincuentes aprenden a anonimizarse más sofisticadamente y hacen más difícil su identificación. Para el año 2005 el número de malware únicos asciende a más de 300.000, y ha ascendido más de un 1000 % en diez años. Para el año 2008 dicha cifra asciende 5 millones de malware únicos.

En los últimos años la variedad de ataques de robo de información ha cambiado mucho. Con la aparición de los smartphones, surgen nuevas amenazas para nuestra privacidad como *payloads* que corren como un proceso en segundo plano y permiten tracear nuestra ubicación, espiarnos a través de la cámara del dispositivo y un largo etcétera. Ésto ha ocasionado que surjan también aplicaciones antivirus en nuestros teléfonos móviles, que actúan de forma *pasiva* vigilando que ningún proceso monitorice de forma maliciosa nuestro dispositivo.

Sin embargo, con el incremento del número de amenazas a nuestra privacidad también ha sido necesario que aparezcan herramientas que actúen de forma *activa* y nos permitan utilizar los servicios de Internet de una forma más segura.

2.3. Herramientas para la protección de identidad

Hoy en día existen numerosas aplicaciones que permiten aumentar nuestro grado de anonimato a la hora de realizar tareas en Internet, usualmente a costa de una más óptima velocidad de conexión,por ejemplo, el servicio **Tor**, o a cambio de una suscripción temporal, como los **servicios VPN de pago**.

2.3.1. Servidor proxy

Un servidor proxy es básicamente un mediador entre un usuario que realiza una petición y otro servidor. Su funcionamiento es relativamente simple: Cuando un cliente de la red desea acceder a un recurso, es el servidor proxy el que realiza la comunicación y el que lleva el resultado de la petición al usuario final.

Los usos de dicha aplicación en ejecución van desde aumentar el rendimiento de algunas operaciones sirviendo como memoria caché, hasta proteger la identidad del usuario que lo utiliza. Dicha finalidad depende del tipo de proxy que se esté utilizando. Hay varios tipos, los cuales se resumen a continuación.

Proxy Web

Un **servicio proxy ó proxy web** es un proxy para una aplicación concreta, y permite el uso de los protocolos FTP y HTTP/S.

Éste tipo de proxy es muy comúnmente utilizado para proteger la privacidad y se puede utilizar junto con Tor (el cual veremos más adelante) para mejorar el grado de anonimato en la red.

El esquema de funcionamiento es el siguiente:

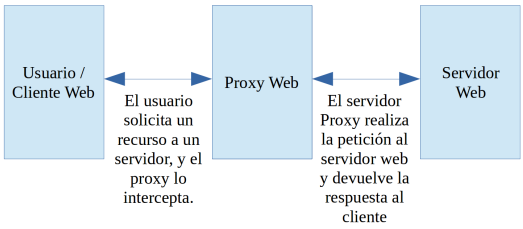

Figura 2.2: Funcionamiento de un Web proxy o Proxy service

Proxy Cache

Su propósito es el de guardar el contenido solicitado por el usuario para así mejorar la velocidad de respuesta en futuras solicitudes de recursos. Conviene destacar a su vez que un proxy web puede actuar también almacenando las páginas web solicitadas, actuando de cierta manera como un proxy caché. Funciona de la siguiente manera:

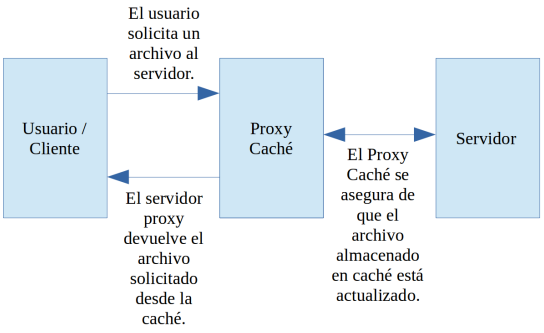

Figura 2.3: Funcionamiento de un Proxy Caché

Transparent Proxy

También es conocido como proxy forzado, y tiene la peculiaridad de no modificar la petición realizada por el cliente ó respuesta más allá de la autenticación del propio proxy. Se le llama transparente puesto que el usuario final no necesita realizar ningún tipo de configuración adicional en el navegador. Su uso es principalmente el de filtrar ciertas conexiones (se combina con un *cortafuegos*) y para proporcionar seguridad.

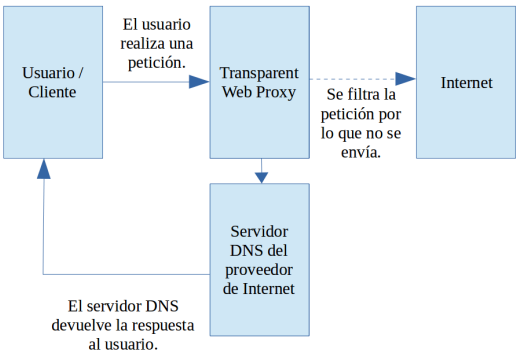

Figura 2.4: Funcionamiento de un Proxy transparente

Reverse Proxy

Este proxy tiene la peculiaridad de estar alojado en uno o más servidores web. Es decir, mientras que un proxy normal es el intermediario entre sus clientes para realizar peticiones a cualquier servidor, un proxy inverso es el intermediario entre sus servidores asociados para ser contactados por cualquier cliente.

NAT proxy ó enmascaramiento

El uso de este proxy es también llamado **enmascaramiento de IP**. En este caso, las direcciones de destino de los paquetes IP son reemplazadas por otras. En este caso la actuación de mediador es entre los equipos de la red interna y la red exterior.

Los aquí citados son los principales tipos de proxies. No obstante, existen algunos más, como el **proxy abierto** y el **Cross-Domain** proxy.

Una vez dejados claros los conceptos básicos sobre proxies, mencionaremos algunas herramientas útiles que hacen uso de éstos para navegar de una forma más anónima y segura por la red.

Proxychains

Proxychains es un programa disponible únicamente para GNU/Linux y Unix que nos permite crear cadenas de proxies, escondiendo así nuestra dirección IP pública en **todo tipo de conexiones** (HTTP, FTP, SSH, etcétera). Esto se traduce en que podemos navegar por Internet o realizar cualquier operación en la red de redes sin descubrir nuestra identidad real.

Mientras que en las figuras anteriores mostramos una conexión a la red con la utilización de un sólo proxy, en el caso de *proxychains* (como su propio nombre indica) utilizaremos cadenas de servidores proxy para anonimizar el tráfico que generemos(no necesariamente debe ser tráfico web).

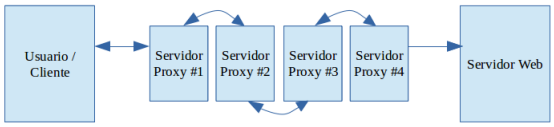

Figura 2.5: Ejemplo de una cadena de proxies

Para hacer funcionar proxychains en un equipo es necesario modificar su fichero de configuración (*proxychains.conf*). En él existen varias opciones en cuanto a la formación de ccadenas de proxies:

- *Dynamic chains*: Supongamos que tenemos 4 servidores proxies añadidos en nuestro archivo de configuración, en este orden: A, B, C y D. En el caso de que, por ejemplo, el servidor B esté caído y el resto funcionen perfectamente, la conexión se realizará de la siguiente manera:

  Es decir, aunque uno (o varios) de los servidores proxy que componen la cadena no funcione, siempre se intentará realizar la conexión omitiéndolo.

- *Strict chains*: Si tomamos el ejemplo anterior, en el caso de una cadena estricta ocurre lo siguiente:

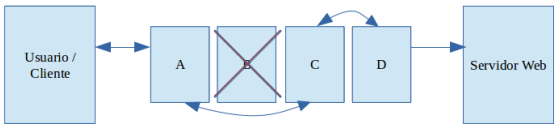

Figura 2.6: Cadena dinámica

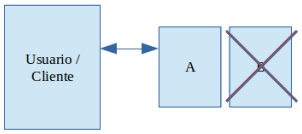

Figura 2.7: Cadena estricta

En el caso de que uno de los servidores proxy falle, la conexión no será satisfactoria. Por ello, las cadenas estrictas tienen la peculiaridad de que siguen el orden de los servidores rigurosamente.

- *Random chains*: Éste tipo de cadenas es totalmente distinta de las anteriores. Básicamente escoge uno de los servidores proxy que aparecen en el archivo de configuración de forma aleatoria.

Ahora bien, ¿cómo podemos añadir servidores proxy a nuestra cadena? El formato para añadirlos es el siguiente:

socks5 192.168.67.78 1080 user password

El primer elemento es el tipo de proxy. Las posibilidades son HTTP, socks4 y socks5. Como nuestro objetivo es el de anonimizarnos y proteger nuestra identidad **siempre que sea posible se intentará utilizar socks5**. El segundo y tercer elemento es la dirección IP y el puerto del proxy, respectivamente. Por último, el cuarto y quinto campo son opcionales y depende de si el proxy que utilizaremos cuenta con usuario y contraseña. Normalmente cuentan con contraseña los proxies que adquirimos por medio de plataformas de pago.

Hay una gran cantidad de páginas que ofrecen servidores proxies, tanto gratuitos como de pago. La diferencia radica en la carga de dichos servidores. Muchos de los servidores proxy gratuitos publicados en la red se encuentran saturados y limitan mucho la velocidad de conexión.

No obstante, es posible encontrar servidores proxy gratuitos que funcionan relativamente bien. Una buena página en la que encontrarlos es https://socks-proxy.net/. Ésta página permite visualizar servidores proxy gratuitos con disponibilidad actualizada cada 20 minutos. Además, permite filtrarlos según tipo y sobre todo, país de origen.

Con el objetivo de mantenernos anónimos en la red, conviene utilizar proxies ubicados en países que tengan buenas políticas de privacidad. Ejemplo de ello son países como **Rusia, China o Países Bajos**.

2.3.2. VPN

Una VPN (o red virtual privada) no es más que el uso de una red privada segura sobre una red pública más grande. La notoriedad que ha cosechado éstos últimos años reside en que nos permite gozar de un alto grado de anonimato al darnos acceso al envío y recibo de datos de la red pública, teniendo todas las políticas de privacidad de una red privada.

La forma de conseguir esto es, normalmente, estableciendo una conexión extremo a extremo mediante el uso de cifrado y/o conexiones dedicadas.

Pese a que el término se ha visto utilizado enormemente estos últimos años, lo cierto es que las redes privadas virtuales existen desde hace bastante tiempo. De hecho, la primera forma de VPN surgió con SwIPe (*Software IP Encryption Protocol*), un trabajo experimental surgido en el año 1993 por John Ioannidis y su equipo en la Universidad de Columbia y AT&T Labs. Éste proyecto pretendía garantizar confidencialidad, integridad y autenticación del tráfico de red.

Tras este experimento, en el año siguiente Xu Wei continuó investigando acerca de la seguridad del protocolo IP hasta formar la familia de protocolos IPSec, la cual autentica y cifra cada paquete compartido a través de una red pública. Después de un tiempo y tras la mejora en las velocidades de transmisión de paquetes, y de la función *plug-and-play*, fue posible la salida al mercado de **las primeras VPNs**.

A la vez que apareció IPSec, se realizó un trabajo en la Biblioteca de Investigación NAVAL con ayuda de DARPA (Defense Advanced Research Projects Agency) con el que surgió el **Protocolo de Seguridad de Encapsulación**. Con ello surgió un gran avance para la seguridad en internet y la tecnología VPN. La Carga de Seguridad Encapsulada, **ESP, ofrece la autenticidad, integridad y protección de la confidencialidad de los paquetes de datos**. Permite configuraciones de autenticación, encriptación, o ambos. Este protocolo es similar al de los Encabezados de Autenticación y proporciona una segunda capa de seguridad para las conexiones a Internet.

1995 fue el año en el cuál se creó el grupo de trabajo de IPsec dentro de la IETF, o Internet Engineering Task Force, el cual es una comunidad de ingenieros de Internet, proveedores, desarrolladores y otras personas interesadas en la evolución de internet y su buen funcionamiento.

El objetivo de este grupo era el de crear un conjunto estandarizado de protocolos disponibles libremente y examinados abordando los componentes, extensiones y la implementación de IPsec.

IPsec está formado por tres subprotocolos:

- *Autentication Header (AH)*: Este protocolo es el encargado de proporcionar integridad de datos en el caso de no haber conexión y autenticación de paquetes IP, además de protección contra ciertos tipos de ataques. La **autenticación** es importante porque asegura que los paquetes de datos que envías y recibes son los que deseas, no el malware u otros ataques potencialmente dañinos. Hay varias versiones con diferentes grados de protección a diferentes niveles. En todos los casos, la Carga de Paquetes IP, tus datos personales están protegidos.
- *Encapsulating Security Payload (ESP)*: Se ocupa de proporcionar la confidencialidad de esos paquetes, al igual que integridad de origen de los datos, la seguridad a los ataques y también seguridad para el tráfico de flujo. Cuando se utiliza en **Modo Túnel**, proporciona seguridad para todo el Paquete IP.
- *Security Associations (SA)*: Son algoritmos y datos que permiten que AH y ESP funcionen correctamente. Básicamente, los datos se cifran en paquetes en la fuente y luego se transfieren a través de internet de forma **anónima** para ser recibidos, autenticados y descifrados en el destino. Las asociaciones se crean sobre la base de la Internet Security Association And Key Management Program (ISKAMP) utilizando una serie de números.

Además de esto, existen dos modos de funcionamiento:

- *Transport Mode*: En Modo Transporte únicamente la carga Útil de IP es típicamente cifrada asegurando los datos, pero dejando visible la información que se origina.

- *Tunneling Mode*: En Modo de Túnel, todo el Paquete IP está cifrado y encapsulado, se le otorga un nuevo encabezado de autenticación y luego se envía. Modo túnel es la tecnología que impulsa la VPN de hoy en día.

Veamos una comparativa entre un paquete IP original, uno que hace uso del modo transporte y otro del modo túnel de IPsec:

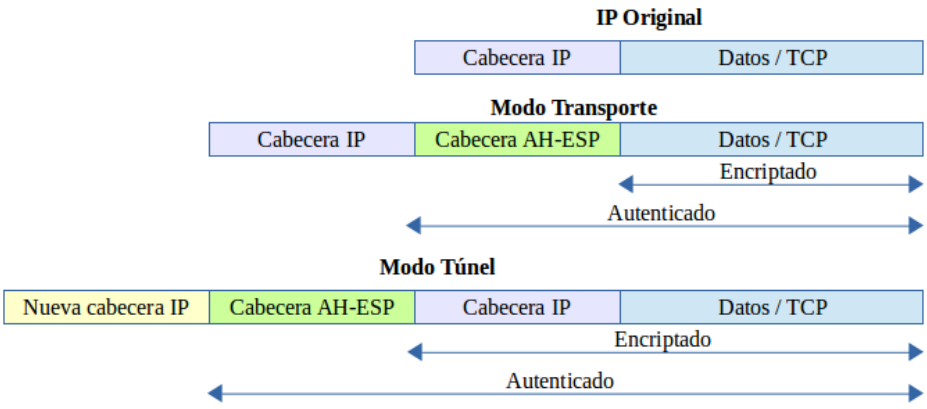

Figura 2.8: Ejemplo de paquetes IP con diferentes niveles de seguridad de encapsulación

El protocolo de túnel hace de las VPNs una gran opción si el objetivo es la protección de nuestra información personal. Permite a un usuario conectarse a Internet con una dirección IP que no es parte de la red local. La **tunelización** funciona encriptando y encapsulando los datos, es decir, lo que proporciona un tercer y muy buscado beneficio: **el anonimato y la privacidad**.

La forma en la que opera es un poco más compleja que con el modo de transporte, los paquetes que contienen la información que realiza el cifrado y el servicio de entrega se llevan a cabo dentro de la carga útil del mensaje original, pero operan a un nivel más alto que la propia carga útil, creando un escudo formado desde dentro y seguro de las influencias externas. Los mejores servicios cifrarán todo el paquete, el marcador de identificación y todo; a continuación, volverán a encapsularlo con una nueva dirección IP y marca de identificación para obtener una completa privacidad.

El modo túnel recoge su nombre de la forma en la que una red VPN opera:

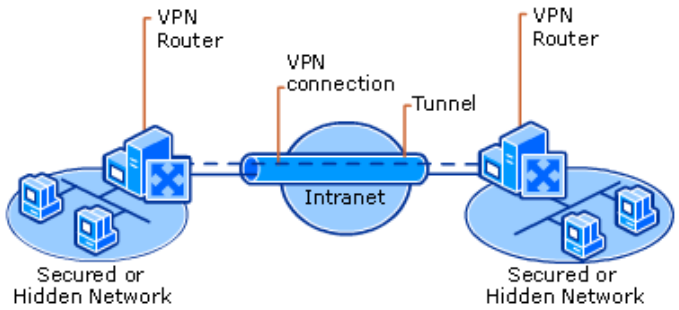

Figura 2.9: Fuente: *https://technet.microsoft.com/pt-pt/library/cc779919(v=ws.10).aspx*

Hoy en día, se usan diferentes tecnologías VPN, cada una con sus pros y sus contras.

Entre todas ellas las más destacables son:

- PPTP: *Point to Point Tunneling Protocol*, la cual está bajo licencia de Microsoft(fue el

primer protocolo de VPN compatible con Windows), crea una red privada virtual en redes dial-up.

Su implementación requiere poca sobrecarga de cómputos, lo cual lo hace uno de los protocolos de VPN **más rápidos** disponibles actualmente. El problema con esta tecnología reside en que **no es del todo segura**. Aunque ahora normalmente utiliza una encriptación de 128 bits, existen varias vulnerabilidades de seguridad, con la posibilidad de una autenticación MS-CHAP v2 no encapsulada como la más grave. Con todo esto, una red que usase PPTP podría ser decodificada en apenas días. La misma Microsoft, pese a haber corregido el fallo de seguridad, no recomienda el uso de este protocolo, y recomienda el uso de SSTP o L2TP.

- L2TP y L2TP/IPsec: El protocolo de túnel de capa dos, normalmente se implementa con los protocolos IPsec (explicados anteriormente) para encriptar datos antes de la transmisión, a fin de proveer a los usuarios privacidad y seguridad. Todos los dispositivos y sistemas operativos modernos compatibles con VPN tienen L2TP/IPsec incorporado. La configuración es tan rápida y fácil como la de PPTP, sin embargo en ocasiones puede ser problemático en el caso de que usar un cortafuegos NAT restrictivo.

- L2TP y L2TP/IPsec: El protocolo de túnel de capa dos, normalmente se implementa con los protocolos IPsec (explicados anteriormente) para encriptar datos antes de la transmisión, a fin de proveer a los usuarios privacidad y seguridad. Todos los dispositivos y sistemas operativos modernos compatibles con VPN tienen L2TP/IPsec incorporado. La configuración es tan rápida y fácil como la de PPTP, sin embargo en ocasiones puede ser problemático en el caso de que usar un cortafuegos NAT restrictivo. Por el momento, **no hay vulnerabilidades importantes** relacionadas con la encriptación por IPsec, pero John Gilmore, miembro fundador y especialista en seguridad de la Electric Frontier foundation, afirma que es probable que el protocolo sea debilitado intencionalmente por la NSA.

- SSTP: Secure Socket Tunneling Protocol fue presentado por Microsoft en el Service Pack 1 de Windows Vista. Este estándar está además ahora disponible para SEIL, Linux y RouterOS, aunque sigue siendo principalmente una plataforma únicamente para Windows. Utiliza SSL v3, y no tendría por qué tener problemas de seguridad aparentes. Sin embargo, hay que recordar que es propiedad de una gigante como Microsoft, y no puede ser analizado en busca de ingresos clandestinos.

- IKEv2: Internet Key Exchange, en su segunda versión, es un protocolo de túnel basado en IPsec, fue desarrollado por Cisco y Microsoft. Los dispositivos móviles son los más beneficiados con IKEv2 ya que el protocolo de movilidad y multioproveedor que se ofrece en forma predeterminada lo hace extremadamente flexible para cambiar de redes. Pese a que IKEv2 está disponible en menos plataformas comparado con IPsec, tiene buena reputación en términos de estabilidad, seguridad y rendimiento.

- OpenVPN: Es un estándar *open-source* relativamente nueva, utiliza los protocolos SSLv3/ TLSv1 y biblioteca OpenSSL para brindar a los usuarios una solución de VPN confiable y potente. El protocolo tiene amplia capacidad de configuración, lo que hace que sea muy difícil de bloquear para servicios como Google. La principal ventaja de esta tecnología es que OpenSSL, la biblioteca que utiliza, soporta **múltiples algoritmos criptográficos** tales como 3DES, AES, Camellia, Blowfish, CAST-128 y más, aunque Blowfish o AES son utilizados casi exclusivamente por proveedores de VPN. La rapidez con la que se desempeña el protocolo OpenVPN depende del nivel de encriptación utilizado, pero normalmente es más rápido que IPsec. Por contra, la configuración, es complicada en comparación con L2TP/IPsec y PPTP.

OpenVPN

Una vez explicadas las diferencias de este estándar con algunas de sus alternativas, vamos a ver cómo funciona éste en una plataforma GNU/Linux.

Lo primero que conviene hacer es cambiar **el servidor DNS por defecto.** Esto, pese a que no es algo explícito ni directamente relacionado con el funcionamiento de la VPN, sí es recomendado. En determinadas ocasiones, pese a utilizar un servicio VPN, la traducción de nombre de dominios a su correspondiente IP numérica (dicha petición debería hacerse mediante el túnel VPN) puede hacerse erróneamente por medio del proveedor de Internet. Ésto se conoce como **DNS leak**. Para evitarlo hay muchas soluciones. Entre ellas, algunos clientes de VPN (como Mullvad) permiten activar un campo en la configuración que evita estos problemas. Otra forma, en sistemas Unix, es acceder al archivo de configuración localizado en /etc/dhcp/dhclient.conf, descomentar la línea:

#prepend domain-name-servers 127.0.0.1;

Evidentemente, hay que cambiar la dirección del servidor de DNS que utilizaremos. Podemos encontrar múltiples servidores, todos ellos seguros, en *https://www.opendns.com.*

Tras esto (y reiniciar la red, evidentemente) podemos utilizar openVPN sin riesgo a que ocurran DNS-leaks.

Utilizar openVPN es tan sencillo como llamar al programa por línea de comandos pasándole como argumento un fichero *.ovpn*, los cuales podemos conseguir, por ejemplo, en *openvpn.com.*

Una duda que puede surgirnos es, si nuestro objetivo es mejorar nuestra privacidad y anonimato ¿cuál es una mejor alternativa, una VPN ó el uso de un proxy?

Lo cierto es que, como veremos, no son las únicas opciones a la hora de obtener un mayor grado de anonimización.Sin embargo, la respuesta no es rotunda puesto que depende de qué servicio VPN y qué proxy se utilice.

El uso de un proxy, hay tres protocolos principales, que como hemos visto son HTTP/HTTPS y SOCKS, Si usamos un proxy HTTP ó SOCKS, el uso de éste no proveerá de ningún tipo de encriptación de los datos, mientras que los proxies HTTPS ofrecen un nivel de encriptación igual que una web que funcione con el protocolo SSL. Además, por lo general, el uso de los proxies (sobre todo si se usan cadenas de los mismos) implican una muy baja velocidad de conexión.

El caso de los VPN, hasta el momento es imposible interceptar el tráfico que circula por su túnel. Sin embargo, ésto produce un **único punto de fallo**, y es el servicio VPN. Es necesario asegurarse de que el servicio VPN que utilizamos no guarda logs ni otros datos, pues en el momento en el que dichos logs saliesen a la luz, estaríamos totalmente expuestos.

2.3.3. Tor

El uso de Tor supone una tercera entre muchas alternativas para mejorar nuestra privacidad y anonimato.

Tor es un software libre que permite defenderte contra el análisis de tráfico, y avoca por la privacidad y libertad del usuario. La forma en la que lo consigue es haciendo rebotar"nuestra información en distintos **nodos Tor** intermedios con el objetivo de que el origen de la información se oculte lo máximo posible.

2.4. Herramientas de identificación de usuarios

2.4.1. Introducción

2.5. Debate social

2.5.1. LOPD GPDR proteccion de datos ley reciente

3

Sistema, diseño y desarrollo

3.1. Segmentación

3.2. Normalización

3.3. Codificación

3.4. Matching

4

Experimentos Realizados y Resultados

4.1. Bases de datos y protocolo

4.2. Sistemas de referencia

4.3. Escenarios de pruebas

4.4. Experimentos del sistema completo

5

Conclusiones y trabajo futuro

Glosario de acrónimos

- **IS**: Iris Subject
- **DCT**: Discrete Cosine Transform
- **WED**: Weighted Euclidean Distance

Bibliografía

[1] Autor Apellidos. Titulo del artículo. *Revista de publicación*, pages 65–73, 2008.

A

Manual de utilización

B

Manual del programador