TRABAJO FIN DE GRADO

INGENIERÍA EN INGENIERÍA INFORMÁTICA

Analizador de mensajes de correo

Subtítulo del proyecto

Autor
Pedro Luis Fuertes Moreno

Directores
Alberto Guillén Perales
Gabriel Maciá Fernández

Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación

—

Granada, mes de 201

Analizador de mensajes de correo: Subtítulo del proyecto

Pedro Luis Fuertes Moreno

Palabras clave: Correo electrónico, ciberseguridad, phishing, malware, virus

Resumen

Este proyecto surge debido a la falta de herramientas, tanto para usuarios técnicos como domésticos, para analizar un correo sospechoso una vez que llega a la bandeja de entrada.

El objetivo de este proyecto es, por tanto, intentar proporcionar una ayuda extra en la identificación de correos maliciosos mediante la extracción y relación de características comunes. Por otro lado, se ofrecerá un servicio público y funcional para que los usuarios puedan aprovechar toda la información recopilada y analizar sus propios correos.

A lo largo del documento se hablará de los distintos tipos de correos maliciosos, de algunas técnicas usadas por los atacantes para engañar o manipular a sus víctimas, de los patrones que se van a extraer y cómo, de las relaciones que se van a hacer.

En la parte de diseño se analizarán y compararán tanto lenguajes de programación como bases de datos, teniendo la idea en mente de migrar todo el servicio a la nube para tener un SaaS, siendo especialmente relevante el servicio de Azure Functions para ejecutar el código. En la parte funcional, el servicio debe permitir tanto analizar como buscar resultados, así como mostrar información relevante derivada del análisis.

También se indicarán problemas encontrados, soluciones aplicadas y posibles mejoras.

Finalmente, y como objetivo último se intentarán obtener ingresos económicos por parte del servicio de cara a crear una posible startup o venta del servicio.

Analizador de mensajes de correo: Subtítulo del proyecto

Pedro Luis Fuertes Moreno

Keywords: e-mail, cybersecurity, phishing, malware, virus

Abstract

Write here the abstract in English.

Yo, **Pedro Luis Fuertes Moreno**, alumno de la titulación Grado en ingeniería informática de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI XXXXXXXXX, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Pedro Luis Fuertes Moreno

Granada a X de mes de 201 .

D. **Alberto Guillén Perales**, Profesor del Área de XXXX del Departamento Departamento de ... de la Universidad de Granada.

D. **Gabriel Maciá Fernández**, Profesor del Área de XXXX del Departamento Departamento de ... de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado ***Analizador de mensajes de correo, Subtítulo del proyecto***, ha sido realizado bajo su supervisión por **Pedro Luis Fuertes Moreno**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a X de mes de 201 .

Los directores:

Alberto Guillén Perales **Gabriel Maciá Fernández**

Agradecimientos

A mi familia, en especial a mi padre, a mi madre, a mi hermano y a mi hermana.

A mis amigos.

A mis profesores.

Índice general

Capítulo 1

Introducción

Aunque en la actualidad poseemos una gran cantidad de aplicaciones con las que comunicarnos con otras personas, el correo electrónico sigue siendo una de las herramientas de comunicación más usadas, especialmente en el mundo empresarial.

Esto es debido a que fue uno de los primeros servicios en permitirnos enviar y recibir tanto texto como otros tipos de archivos de manera rápida y sencilla.

Según un estudio de hace tres años de The Radicati Group el correo electrónico tenía 3.7 mil millones de usuarios, que enviaban 269 mil millones de correos cada día. [1]

Viendo las cifras anteriores se puede entender la importancia de tener un servicio de este tipo lo más seguro posible, ya que un ataque bien diseñado puede afectar a millones de personas.

Sin embargo, debido a cómo y cuándo crea, la seguridad nunca ha sido uno de sus puntos fuertes y eso ha llegado hasta nuestros días.

1.1. Breve historia del correo electrónico

El correo electrónico se remonta a 1962 en el MIT, cuando compraron a IBM un ordenador que permitía que distintos usuarios iniciaran sesión y guardaran archivos en él. Estos lo aprovecharon para intercambiar mensajes, lo que provocó que para 1965 se desarrollara un servicio que facilitase esa comunicación entre los distintos usuarios y lo llamaron MAIL.

Hay que tener en cuenta que, en ese servicio, los mensajes no salían de dicho ordenador. Habría que esperar hasta 1971 para ver lo que sería el primer "correo electrónico" enviado a través de una red, en concreto de ARPANET y fue gracias a Ray Tomlinson, que adaptó un programa que permitía enviar mensajes a distintos terminales de distintos usuarios de un mismo ordenador, para poder enviar mensajes entre distintos terminales, aunque no estuviesen en el mismo ordenador. Precisamente el "@" del co-

rreo electrónico viene de la necesidad de Tomlinson de tener que separar al usuario del equipo, ya que anteriormente esto no era necesario.

No es hasta 1977 cuando se crea el primer rfc del correo electrónico, concretamente el rfc733 [2], aunque, este protocolo no es usado en la actualidad. El primer rfc del primer protocolo que aún se usa es el rfc821 [3] de SMTP de 1982 y el rfc918 [4] de POP de 1984.

La situación por aquél entonces de lo que ahora conocemos como Internet, era muy distinta. Internet estaba reservado a universidades, centros de investigación e instituciones gubernamentales. Esto hizo que cuando se desarrollasen estos protocolos, no se pensara en la seguridad de ellos.

Y este es uno de los grandes problemas que tiene el correo en la actualidad, ya que, aunque tanto SMTP como POP (E IMAP [5] aunque no se ha mencionado antes) han ido recibiendo actualizaciones, están basados en unos protocolos diseñados y pensado para un entorno radicalmente diferente en el que se siguen utilizando.

1.2. Origen del proyecto

Este proyecto surge debido a la gran cantidad de demandas que he recibido en los últimos años por parte de familiares y amigos para que, les ayudase a verificar si un correo sospechoso que les había llegado a su bandeja de entrada era o no malicioso.

Y esta tarea, que, para mí era trivial en la mayoría de los casos, no lo era para ellos. Aunque algunos estaban tan bien preparados que incluso a mí me costaba diferenciarlos.

Todo esto me llevó a pensar dos cosas. La primera, que una vez que un correo llega a la bandeja de entrada del usuario, sólo le queda su intuición para confiar o no en el mensaje, intuición que puede fallar incluso si se tienen conocimientos técnicos.

Por otro, que en caso de querer investigar dicho mensaje y obtener más información, no existe ninguna herramienta específica para ello, por lo que se tiene que hacer todo a mano y siendo imposible obtener algún tipo de relación con otro mensaje parecido, lo cual dificulta mucho el proceso.

1.3. El proyecto

El objetivo de este proyecto es tratar de paliar esta carencia de herramientas tanto para la identificación de correos maliciosos una vez que han pasado los filtros de spam, como para la investigación de un mensaje en los casos más complicados.

Por este motivo se va a crear por un lado, un servicio que permita analizar mensajes *on-line*, para que cuando se reciba un mensaje sospechoso, el usuario tenga una segunda validación con más información y por otro lado,

se va a permitir al usuario poder "navegar" entre los datos encontrados en el correo para poder obtener más información, lo que puede facilitar en gran medida un análisis más profundo del correo por parte de una persona más técnica en caso de ser necesario.

Para llevar esto acabo se debe crear una gran base de datos con todos los correos analizados, así como los distintos datos que se han extraído de los mismos y sus relaciones.

También se deben identificar patrones maliciosos conocidos, para poder alertar a los usuarios menos especializados sobre los patrones encontrados sin necesidad de que ellos sepan en qué consisten.

Finalmente, [como ejercicio académico] se presentará un estudio de modelo de negocio para poder hacer económicamente viable esta propuesta de servicio...

Capítulo 2

Motivación

En la actualidad los ataques por correo electrónico siguen siendo un hecho y prácticamente todo el mundo ha recibido algún correo de este tipo alguna vez, lo que demuestra que, las herramientas actuales no son capaces de solucionar de manera efectiva este problema.

Esto se suma a que cada vez los ataques son más y más sofisticados, y, por tanto, complicados de detectar, ya no solo por estas herramientas, sino, por las propias personas, sean o no profesionales del sector. Y es que cuando un usuario recibe un mensaje de este tipo no tiene ninguna herramienta extra que le ayude a comprobar si es o no malicioso.

Esto es así hasta tal punto, que la solución para identificar estos correos de grandes empresas antivirus dedicadas a la ciberseguridad es que los usuarios sigan su intuición [6], otras ofrecen pequeños cursos para identificarlos [7].

Todo esto refleja como los usuarios están en una clara situación de vulnerabilidad, especialmente los usuarios menos técnicos que no conocen cómo funciona realmente la tecnología y que están usando ¿Y es que acaso deberían?

Bajo mi punto de vista, el enfoque de las grandes compañías sobre cómo atacar los problemas de seguridad que tienen el correo electrónico, está dirigido a personas con unos conocimientos que la mayoría de las personas no tienen y por ta nto sólo es útil para una minoría de usuarios del servicio.

Pero además, es que dichas compañías tampoco están exentas de ataques de esta naturaleza, esto se pone de manifiesto en algunos ataques llevados a cabo con éxito a empresas tecnológicas de máximo nivel, como pueden ser Google o Facebook, a las que un atacante les consiguió robar 121 millones de euros [8]

Todo esto lleva a pensar que actualmente sigue habiendo un gran agujero de seguridad en el correo electrónico y que las herramientas existentes no son suficientes ni siquiera para los expertos del sector.

2.1. Aplicaciones relacionadas

A continuación, se van a analizar un conjunto de herramientas que, si bien no ofrecen soluciones completas a este problema, pueden ser buenos servicios en los que apoyarse para tratar de dar una solución más amplia y completa que las actuales.

2.1.1. Herramientas específicas del correo electrónico

Las herramientas descritas en esta sección únicamente son válidas en correos electrónicos, por lo que, si en el futuro también se desean analizar otro tipo de mensajes, o no se tienen todas las cabeceras asociadas al mensaje, no serán válidas.

Filtros de correo no deseado

Tal vez sea la mejor solución que se tiene actualmente para mitigar este tipo de problemas. Son filtros que analizan cada uno de los mensajes que se reciben y en base a distintos criterios informan al usuario si el correo es o no legítimo.

Aunque, este tipo de herramientas tienen varios problemas asociados:

- Su efectividad: Conseguir una herramienta con una efectividad del 100 % es prácticamente imposible y esto es algo que se debe asumir, sin embargo, hay ciertos filtros cuya efectividad bastante reducida.
- La imposibilidad de poner un filtro de este tipo: hay muchos servicios de correo electrónico que no permiten al usuario poner un filtro personalizado de correo no deseado. Ejemplos de estos servicios son Gmail u Outlook, 1500 [9] y 400 [10] millones de usuarios respectivamente (¡Son muchos!)
- Una vez que el mensaje pasa el filtro, esta herramienta deja de ser efectiva, lo cual puede ser muy sencillo para un atacante, simplemente va probando con distintos correos hasta que consigue uno que lo pase.

Un servicio de este tipo puede ser muy útil para un primer análisis si el mensaje es compatible. Normalmente tienen una larga trayectoria, son rápidas y tienen una efectividad comprobada.

2.1.2. Analizador de cabeceras

Un analizador de cabeceras de correo electrónico da información relevante y normalmente invisible al usuario sobre dicho correo. Estas cabeceras pueden aportar información muy útil como por ejemplo la ruta seguida por

el mensaje, la ip del servidor de correo desde donde se envió o si ha pasado o no los filtros antispam del proveedor de correo.

Actualmente hay varias páginas que ofrecen este servicio, como por ejemplo Google [11], Mircrosoft [12] o Mxtoolbox [13].

Ofrecer este tipo de análisis es muy interesante, ya que permite de manera sencilla hacer un análisis más profundo por parte de un profesional del sector.

2.2. Herramientas generales

Las herramientas que se describen a continuación, si bien no están destinadas al correo electrónico como tal, se pueden usar de manera efectiva para analizar los distintos mensajes que se reciban, sean o no correos electrónicos.

Virus Total

Virus Total [14] es una web propiedad de Google que permite analizar archivos, y direcciones web en busca de programas malignos y aunque no está directamente relacionada con el correo electrónico puede servir de ayuda para analizar posibles archivos adjuntos, así como posibles direcciones sospechosas.

Tener una herramienta que enlace directamente con el servicio puede ser de gran ayuda tanto para profesionales del sector como para usuarios menos especializados, que lo único que tendrán que hacer es clicar en un botón para analizar una dirección de su correo electrónico.

Virus total ofrece una api rest que permite analizar tanto archivos como urls, dominios e ips. [15]

Metadefender

Metadefender [16] es un analizador de archivos, url's, dominios e ips similar a Virus Total de la empresa de ciberseguridad Opswat.

Puede ser una buena alternativa a Virus Total y al igual que este tiene una api pública [17] en la que realizar consultas.

Have I been pwned

Have I been pwned [18] es una web que permite saber si una dirección de correo electrónico ha aparecido en alguna brecha de seguridad, y en caso de que haya aparecido te dice en qué brecha ha sido.

Saber si el mensaje proviene de una dirección comprometida puede ser de relevancia, siendo más probable que un mensaje malicioso provenga de una dirección comprometida que de una dirección que no lo sea.

Have I been pwned tiene una api [19] donde realizar consultas sobre direcciones de correo electrónico, además su base de datos se va actualizando

con las últimas brechas de seguridad que van surgiendo y contando ya con más de 9.500 millones de cuentas de correo.

Capítulo 3

Objetivos y requisitos

3.1. Obtención y relación de patrones

Sacar mediante expresiones regulares datos de interés de correos electrónicos y relacionarlos entre sí. Esto permitirá a los expertos del sector contar con una herramienta para poder realizar análisis de mayor profundidad al poder relacionar datos comunes entre los distintos correos de la base de datos.

Un ejemplo sencillo de esto puede ser el análisis de un correo electrónico nuevo, pero con una url ya conocida y presente en otros correos electrónicos.

Un ejemplo más complejo podría ser, el análisis de un mensaje con una url nueva, pero con un dominio de cuya IP sí se tienen registros, lo que puede dar lugar a una nueva línea de investigación sobre si dicho dominio pertenece (O no) a un cibercriminal, aunque no se tengan registros previos ni del dominio ni de la url en cuestión.

Se debe extraer al menos los siguientes tipos de datos:

- Direcciones IP
- Dominios
- Enlaces
- Direcciones de correo electrónico

Aunque el servicio se debe pensar para que en un futuro se puedan extraer más tipos de datos como carteras de criptomonedas o números de teléfono.

3.2. Evaluar de manera relativa la maliciosidad

Se deben detectar ciertas técnicas comúnmente usadas por los ciberdelincuentes para atacar a sus víctimas y así asignar un valor de maliciosidad tanto a los datos extraídos como a los mensajes en sí.

Algunas de estar técnicas podrían ser:

- Mostrar un enlace distinto al que se redirige.
- Usar una gran cantidad de subdominios de subdominios
- Mostrar un correo electrónico distinto del real

3.3. Ofrecer un servicio de análisis público

La segunda parte del proyecto será ofrecer a todos los usuarios la posibilidad de analizar sus mensajes mediante una web donde podrán o bien copiar y pegar el mensaje, o bien subir un archivo eml para un análisis más completo.

La página devolverá varias listas con todos los tipos datos extraídos del análisis, así como enlaces a cada uno de ellos donde se muestre un informe más detallado.

Esto permitirá que a un usuario que le llegue un correo y no sepa si es o no malicioso, pueda analizarlo en la página y obtener más información sobre este.

3.4. Cumplir de manera efectiva con la ley de protección de datos

Es de vital importancia diseñar el servicio pensando en las leyes de protección de datos, especialmente en la europea por ser la más estricta hasta el momento.

Los usuarios deben poder eliminar, tanto los mensajes, como los datos obtenidos de ellos si así lo desean de manera automática y sin necesidad de que nadie intervenga en el proceso.

Aunque esta característica no se implemente en este proyecto, sí que se debe pensar el sistema para que se pueda implementar en el futuro de manera sencilla y modificando la mínima cantidad de código posible.

3.5. Monetización

Como parte complementaria se intentará pensar cómo obtener ingresos económicos del desarrollo. Estos ingresos nunca deben impedir que usuarios particulares puedan usar el servicio, aunque sí pueden impedir acceder a todas las funcionalidades de este.

3.6. Integración con terceros

Aunque esto está un poco al margen del TFG, sería muy interesante la integración directa desde la página con otros servicios como el de Virus Total o la de Have I been pwned.

Esto permitiría por un lado facilitar a los usuarios menos técnicos un análisis rápido desde la propia página y por otro puede dar información relevante a los expertos.

3.7. Crear un servicio funcional

Al final del proyecto se debe crear un servicio que dé un soporte real y online, no se debe limitar a tener un servicio local únicamente de prueba. Por lo que durante el apartado del diseño se deben elegir tecnologías factibles para su posterior despliegue en internet y por tanto no limitadas a un entorno de local pruebas.

Capítulo 4

Análisis y diseño

4.1. Elección del lenguaje de programación para la extracción de datos

4.1.1. Java

4.1.2. C++

4.1.3. Pyhton

4.1.4. Node.js (JavaScript)

4.1.5. PHP

4.2. Elección de la base de datos en la que almacenar los datos

4.0.1. NO-SQL

MongoBD

Neo4j

4.0.2. SQL

MariaBD

PostgreSQL

4.1. Visualización de los datos

4.2. Adecuación a la ley de protección de datos

Capítulo 5

Implementación

5.1. Expresiones Regulares

5.2. MySQL

5.3. PHP

Capítulo 6

Pruebas

Capítulo 7

Conclusiones y vías futuras

Bibliografía

[1] The Radica ti Group. Email statistics report, 2017-2021. http://www.radicati.com/wp/wp-content/uploads/2017/01/Email-Statistics-Report-2017-2021-Executive-Summary.pdf, 2017.

[2] Internet Engineering Task Force. Standard for the format of arpa network text messages. https://tools.ietf.org/html/rfc733, 1977.

[3] Internet Engineering Task Force. Simple mail transfer protocol. https://tools.ietf.org/html/rfc821, 1982.

[4] Internet Engineering Task Force. Post office protocol. https://tools.ietf.org/html/rfc918, 1984.

[5] Internet Engineering Task Force. Interactive mail access protocol - version 2. https://tools.ietf.org/html/rfc1064, 1988.

[6] Malwarebytes. What is phishing? https://www.malwarebytes.com/phishing/.

[7] Google. ¿puedes detectar cuándo te están engañando? https://phishingquiz.withgoogle.com/?hl=es.

[8] ABC. El hombre que logró robar 121 millones de dólares a facebook y google con facturas falsas. https://www.abc.es/tecnologia/redes/abci-hombre-logro-robar-121-millones-dolares-facebook-y-google-facturas-falsas-201903261035_noticia.html, 2019.

[9] Profesional Review. Gmail tiene ya 1.500 millones de usuarios activos. https://www.profesionalreview.com/2018/10/28/gmail-tiene-ya-1-500-millones-de-usuarios-activos/, 2018.

[10] Lifewire. How many people use email worldwide? https://www.lifewire.com/how-many-email-users-are-there-1171213, 2019.

[11] Google. Message header. https://toolbox.googleapps.com/apps/messageheader/.

[12] Microsoft. Message header analyzer. https://mha.azurewebsites.net/.

[13] Mxtoolbox. Email header analyzer. https://mxtoolbox.com/EmailHeaders.aspx.

[14] Virus Total. https://www.virustotal.com/gui/.

[15] Virus Total. Api. https://developers.virustotal.com/reference.

[16] Metadefender. https://metadefender.opswat.com/?lang=en.

[17] Metadefender. Api. https://onlinehelp.opswat.com/mdcloud/.

[18] Have I been pwned. Eldfsfds. https://haveibeenpwned.com/.

[19] Have I been pwned. Api. https://haveibeenpwned.com/API/v3.