Análisis y Simulación de un Ataque de Phishing

Víctor Barroso Beltri

Junio 2021

Director: Manel Medina

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

FIB

Resumen

El principal objetivo de esta tesis es el análisis de un ataque de Phishing, explicando en detalle el funcionamiento de estos, los distintos tipos que podemos encontrar, de qué vulnerabilidades se aprovechan, como ha evolucionado a lo largo del tiempo y cuáles son los datos más recientes relacionados con el tema.

Posteriormente, se utilizará el framework GoPhish para realizar una simulación de un ataque lo más real posible imitando dos webs distintas.

Para comprobar la consciencia de la gente sobre este tipo de ataque se realizará un cuestionario y se estudiarán los resultados.

Por otro lado, se investigarán los filtros de Spam, los cuales a día de hoy son la principal defensa contra ataques de Phishing, y se utilizará Jupyter Notebook para generar un modelo de Machine Learning que cumpla la función de un filtro sencillo.

Por último, se darán una serie de recomendaciones que pueden servir al lector para mejorar la seguridad de sus datos e intentar evitar caer en un ataque de este tipo.

Agradecimientos

Antes de entrar en materia querría agradecer a ciertas personas el apoyo y la ayuda que me han otorgado en el transcurso de este trabajo.

En primer lugar quería agradecer al director de este trabajo, Manel Medina, así como a su colega Matias Altamirano, tanto por los consejos en el desarrollo del trabajo como por la ayuda inicial a nivel inspiracional.

Por otro lado quería agradecer a mi compañera Nora Just i Bartolomé, por el apoyo emocional en los momentos más duros en la realización de este trabajo. También agradecer sus distintos consejos artísticos en lo referente al diseño de las páginas de Phishing.

Por último agradecer a mi familia por el apoyo emocional a distancia en el transcurso de la titulación y la confianza puesta en mi y en las decisiones de vida que he tomado.

Tabla de Contenidos

1 Introducción

1.1 Contexto

En primer lugar, expondré una breve introducción sobre los ataques de Phishing, explicando en qué consisten y qué se pretende realizar en este trabajo.

1.1.1 Introducción

La idea de este trabajo surge de mi interés por el área de la ciberseguridad, que a pesar de su importancia, sólo supone una asignatura obligatoria de la especialidad y una optativa. El semestre pasado cursé esta última impartida por Manel Medina, surgiendo así la posibilidad de realizar el trabajo final bajo su dirección.

Manel Medina es un profesional con muchos años de experiencia en el campo de la seguridad, fundador y director del esCERT-UPC [1], equipo español de respuesta a incidentes de seguridad en la red.
Actualmente, el Phishing es uno de los ataques de los que el esCERT-UPC está desarrollando proyectos. Paralelamente a lo realizado por el grupo, yo expondré este trabajo donde aportaré un contenido de gran valor.

1.1.2 Conceptos básicos

Para comprender el trabajo que expondré, conviene familiarizarse antes con una serie de conceptos que probablemente serán nombrados más adelante con frecuencia.

1.1.2.1 Spam

En el informe publicado en el IEEE sobre la definición de Spam 2.0 [2] definen además el Spam como el abuso de sistemas de mensajería electrónicos para enviar correo no solicitado, especialmente mediante e-mail. De esta forma podemos considerar los correos con objetivos ilícitos como Spam, y no sólo aquellos con un contenido comercial.

1.1.2.2 Framework

Un framework [3] es una plataforma para el desarrollo de software. Supone un entorno donde los desarrolladores de software pueden trabajar cómodamente utilizando generalmente bibliotecas, compiladores y otras herramientas que le faciliten el proceso de desarrollo.

1.1.2.3 Servidor

Dispositivo de un sistema que permite resolver las distintas peticiones de otros dispositivos del sistema, denominados clientes [4].

1.1.2.4 Deep Web/Dark Web

Pese a que cada vez el término Deep Web es más escuchado, tiende a confundirse con Dark Web [5]. El primero hace referencia a todas aquellas webs que los motores de búsqueda no pueden encontrar debido a herramientas de encriptación. Entre estas webs encontramos las pertenecientes a la Dark Web, que son aquellas con fines delictivos o contenido ilegal.
Las webs de la Dark Web pertenecen a la Deep Web, pero esta última además está formada por todas la bases de datos de usuarios, páginas que requieren un registro previo y demás páginas con contenido habitual.

1.1.3 Problema a resolver

El Phishing sigue siendo uno de los ataques más utilizados a día de hoy. Con este trabajo se pretende concienciar a los usuarios sobre cómo funciona y qué pueden hacer para evitarlo.

Actualmente, todo el mundo posee un smartphone, con este se realizan cada día más acciones que requieren de nuestros datos. Por ello es conveniente mentalizar a los usuarios sobre los peligros a los que se exponen.

Es por eso que este trabajo mostrará cómo se realiza uno de los ataques más comunes a día de hoy y aportará algunas recomendaciones para intentar evitarlos.

1.1.4 Actores implicados

Son diversos los grupos que pueden beneficiarse del contenido de este trabajo. Dependiendo del motivo de su interés podemos dividirlos en:

-Especialistas: En primer lugar tenemos el esCERT-UPC y otras instituciones del ámbito de la ciberseguridad que están estudiando aspectos del Phishing así como de los filtros de Spam.

-Educación: Por otro lado tenemos instituciones educativas del área de la informática, que buscan enseñar a sus estudiantes el funcionamiento en detalle de un ataque Phishing así como ciertas defensas a ellos.

-Usuarios: Por último, tenemos a cualquier usuario interesado en mejorar la seguridad de sus datos, ya sea para uso propio o bien porque es el responsable de datos de otras personas, como puede suceder en el ámbito empresarial.

1.2 Justificación

Como se puede ver en la siguiente imagen (Figura 1), cada vez son más los dispositivos que están conectados a Internet.

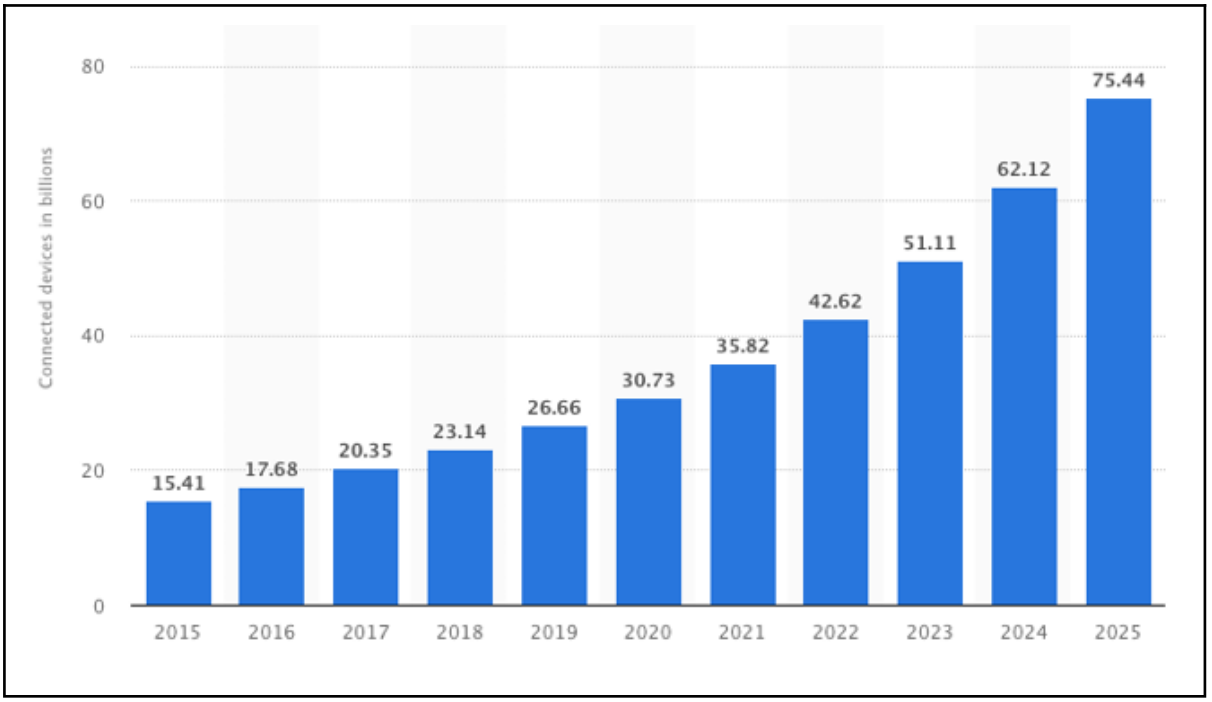

Figura 1. Evolución dispositivos conectados según La Vanguardia[6]

Estos dispositivos (tanto smartphones como ordenadores) son susceptibles de ciberataques, con el fin de hacerse con nuestros datos personales y nuestro dinero.
Uno de los ataques más comunes a día de hoy es el Phishing. Mediante esta técnica los cibercriminales suplantan una página web de nuestra confianza y nos redirigen a ella mediante un correo fraudulento, buscando que introduzcamos nuestros datos.

Pese a la existencia de una amplia documentación sobre Phishing, hay una carencia a la hora de hallar documentación teórica acompañada de una demostración práctica realista. Este realismo se puede lograr con herramientas como GoPhish [7], el framework que utilizaré en el trabajo. Este framework permite simular un ataque de Phishing de una manera muy similar a la realidad. Gracias a esto, se podrá poner en práctica una gran parte de los conocimientos teóricos mencionados, para que los usuarios puedan observar cómo se llevan a cabo estos ataques.

Además, este proyecto abordará también el funcionamiento de los filtros de Spam, una herramienta muy importante a la hora de luchar contra el Phishing, pero que pocos trabajos se centran en ella.

Para finalizar el trabajo, se buscará resumir cuales son las buenas prácticas que se pueden realizar para protegerte de las distintas vulnerabilidades que usa el Phishing.

1.3 Alcance

Como se ha mencionado con anterioridad, este trabajo buscará aunar un contenido teórico con su implementación práctica. Para ello, el trabajo constará de varias partes según el objetivo que se pretenda alcanzar.

1.3.1 Objetivos

1.3.1.1 Conocer el funcionamiento del Phishing

El primer objetivo del trabajo será buscar la información necesaria para conocer en profundidad el funcionamiento del Phishing así como la teoría en la que se basa. Esta información se detallará en el trabajo para poner en contexto al lector sobre lo que se simulará más adelante.

1.3.1.2 Generar una simulación con GoPhish

Se realizará, con el framework GoPhish, una simulación de un ataque de Phishing. Para lograrlo, habrá que cumplir una serie de subobjetivos:

- Diseñar una web que falsifique un servicio de confianza
- Diseñar un correo falso que dirija a la víctima a la web fraudulenta
- Aunar lo diseñado en GoPhish
- Comprobar que es funcional y creíble.

1.3.1.3 Analizar el Phishing en la población

Mediante el desarrollo de un cuestionario y su posterior estudio, se analizará cómo de preparada está la población actual ante este tipo de ataques.

1.3.1.4 Generar Filtros de Spam

En esta parte del trabajo se recopilará información sobre su funcionamiento y se tratará de implementar uno propio. Para ello es necesario completar unos pasos previos:

- Entender el funcionamiento de los filtros de Spam.
- Comprender cómo están implementados actualmente.
- Desarrollar nuestro propio filtro funcional.

1.3.1.5 Dar recomendaciones eficaces a los usuarios

El último objetivo del trabajo será hacer que los usuarios comprendan qué prácticas pueden realizar para estar más seguros o qué debería requerir en su servicio de correo de confianza.

1.3.2 Requisitos

Podremos considerar que nuestro trabajo cumple sus objetivos si superamos una serie de requisitos :

- La simulación realizada con GoPhish resulta funcional y creíble.
- Los filtros de Spam implementados son capaces de filtrar el contenido no deseado
- Con la aplicación de las recomendaciones dadas se puede asegurar una alta seguridad contra el Phishing.

1.3.3 Posibles inconvenientes

Durante la realización del proyecto pueden surgir distintos problemas que si no son previstos con anterioridad pueden causar que el trabajo no se complete a tiempo o que el mismo pierda calidad.

A continuación se exponen algunos de los inconvenientes que a día de hoy se consideran posibles:

- **Programar Webs en HTML:** Para la realización del ataque será necesario escribir código en HTML, que si bien se ha trabajado a lo largo de la carrera, tampoco es en lo que más horas se ha invertido.

- **Problemas con GoPhish:** Incomprensión de la plataforma o falta de documentación de calidad pueden ser algunos de los problemas que surgen a la hora de utilizar este framework.

- **Programar filtros de Spam:** La programación de los filtros de Spam puede resultar más complicada de lo que en un principio se espera.

- **Situación sanitaria:** Los cambios en el panorama global presente pueden traducirse en problemas muy diversos, modificando rutinas y horarios de trabajo.

1.4 Metodología y rigor

1.4.1 Metodología Kanban

Para poder organizar las tareas del proyecto y el estado en que se encuentran, he decidido utilizar la metodología Kanban [8], que ya he utilizado con anterioridad en otros proyectos.

La metodología ágil Kanban permite, mediante un método visual, conocer el estado actual del proyecto y asignar nuevas tareas de manera muy efectiva. Algunas de las ventajas de esta metodología son:

- Transparencia (todo el mundo sabe cuál es su tarea y en qué momento está)
- Evita las tareas ineficientes
- Control de las tareas
- Flexibilidad

1.4.2 Trello

Para aplicar esta metodología de una forma sencilla y que también permita ver al director la evolución del proyecto, se ha optado por Trello. [9]

Trello es un software destinado a la administración y gestión de proyectos. Basa su funcionamiento en la metodología Kanban o "sistema de tarjetas".

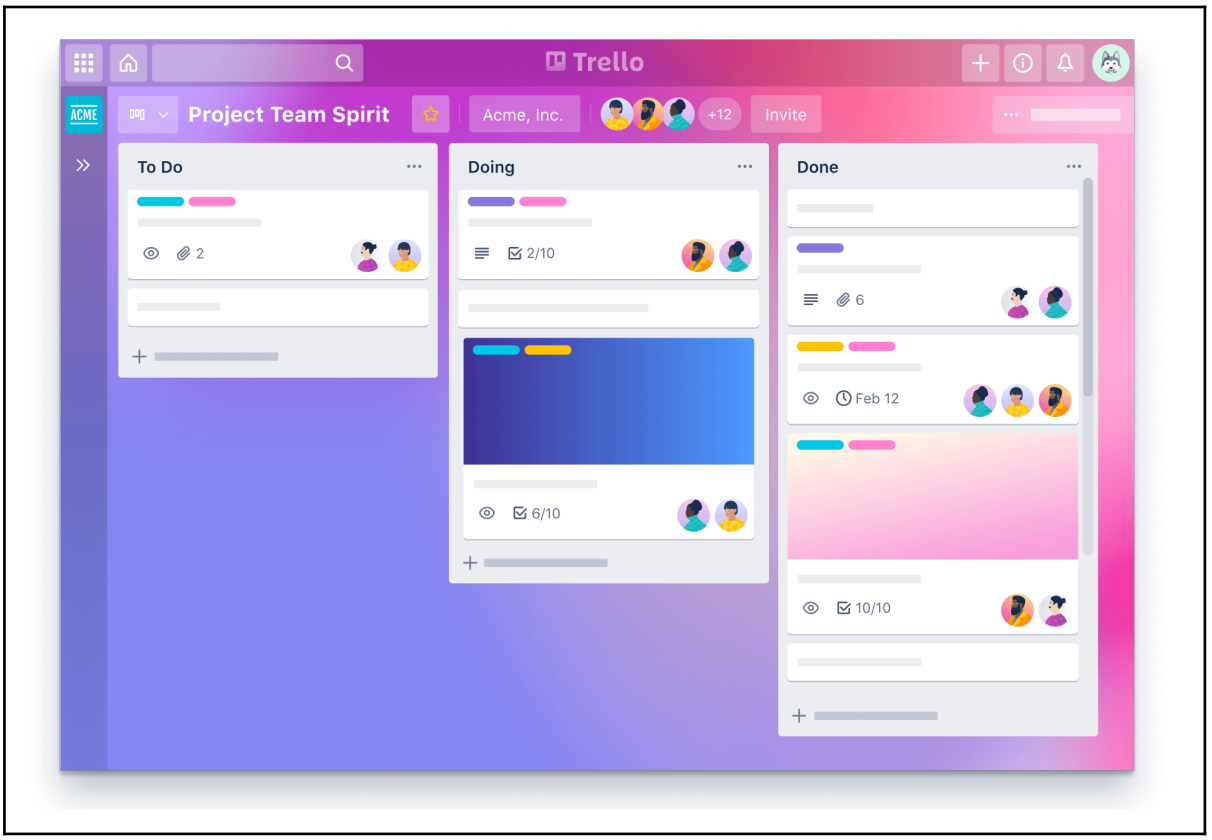

Figura 2. Representación de la interfaz de Trello [9]

Esta página web permite la realización de “tarjetas virtuales” y la modificación del estado de las tareas. En la Figura 2 podemos observar un ejemplo de su uso, separando las tareas en 3 estados: To Do, Doing y Done.

1.4.3 Github

Como repositorio para almacenar el código generado se utilizará Github. [10] De este modo, se podrá mantener un control de las versiones, permitiendo no sólo poder acceder al código desde distintos dispositivos, sino también una continua supervisión por parte de mi director Manel Medina.

1.4.4 Telegram

En cuanto a la comunicación directa con mi director, se usará principalmente Telegram [11] para comunicarle cualquier duda que pueda surgir, o informarle de actualizaciones que se realicen.

1.4.5 Skype

Para poder conversar de forma más fluida o debatir ciertos puntos que por Telegram puedan resultar confusos, y con el fin de evitar reuniones presenciales innecesarias dado el riesgo de salud que en la situación actual supone, se realizarán videollamadas mediante Skype. [12]

1.4.6 Google Documentos

La realización de documentación se hará utilizando la web gratuita Google Documentos. [13]

1.4.7 Google Hojas de cálculo

Para calcular presupuestos necesarios se usará la web gratuita Google Hojas de Cálculo. [14]

1.4.8 Google Presentaciones

De cara a la defensa final, se empleará Google Presentaciones[15] para realizar la presentación.

2 Marco teórico del Phishing

2.1 Definición

Antes de intentar definir qué es el Phishing, lo cual no es sencillo dada su continua evolución, debemos definir un par de conceptos más generales.

2.1.2 Spam

El primer término que debemos conocer es Spam.[25] La definición más sencilla que podemos dar es "correo no deseado en masa". Así, para que un correo sea considerado Spam, debe cumplir dos características:

- No solicitado
- Masivo

En el caso de que una de estas dos condiciones no se cumpla, el correo no se considerará Spam:

- Un correo no solicitado no tiene por qué ser Spam, como puede ser un correo de primer contacto.
- Un correo enviado en masa no tiene por qué ser Spam, como pueden ser los correos del boletín informativo de una web de tu interés.

Esta es la definición de Spam usada por el estándar de la industria, que la iguala al término UBE (Unsolicited Bulk Email). Sin embargo, la definición aportada en las legislaciones de cada país pueden variar ligeramente, provocando diferencias en la forma de clasificar ciertos correos.
Aun así, los proveedores de servicios a nivel mundial siguen la definición de la industria, prohibiendo y pudiendo llegar a actuar contra quienes envíen UBEs.

2.1.3 Ingeniería social

La segunda definición que necesitamos antes de definir el Phishing es la de ingeniería social.
Cuando hablamos de ingeniería social en este trabajo lo haremos desde el punto de vista tecnológico, es decir, definiremos el término ITSE (Information Technology Social Engineering).

Una primera aproximación a su definición la podemos encontrar en Wikipedia [26] :

"social engineering is the psychological manipulation of people into performing actions or divulging confidential information."
Traducción: ***[La ingeniería social es la manipulación psicológica de las personas para que realicen acciones o divulguen información confidencial.]***

Por otro lado, los autores del campo de la ciberseguridad Ira Winkler y Matt Bishop en sus respectivos libros *Zen and the Art of Information Security* [27] y *Computer Security Art and Science* [28], consideran que la ingeniería social debe de ser ajena a la tecnología.

Sin embargo, esta definición dejaría al Phishing fuera de las ingenierías sociales, por lo que utilizaremos otra definición más amplia como la de Charles y Shari Pfleeger en su libro *Security in Computing:* [29]

"Social engineering involves using social skills and personal interaction to get someone to reveal security-related information and perhaps even to do something that permits an attack. [...] The purpose of social engineering is to persuade the victim to be helpful,"
Traducción: [La ingeniería social involucra el uso de habilidades sociales e interacción personal para lograr que alguien revele información relacionada a la seguridad y quizás incluso haga algo que permita un ataque [...] El objetivo de la ingeniería social es persuadir a la víctima para que sea útil.]

2.1.4 Phishing

Ahora que conocemos los dos conceptos anteriores, resulta más sencillo entender el Phishing. Phishing [30][31][32] es el acto de enviar correos falsificados a un usuario, imitando un servicio legítimo, en un intento de estafar al destinatario para difundir información privada como la de una tarjeta de crédito o las credenciales de un servicio.
El Phishing, llamado así por una comparación con la pesca (fishing en inglés), es también llamado, en ocasiones, Brand Spoofing (Suplantación de Identidad de una marca). Se distribuye generalmente mediante correos de Spam y es, además, considerado fraude y falsificación en gran cantidad de países.

Este método usa la ingeniería social para provocar que las víctimas "piquen en el anzuelo". Por un lado, se aprovecha de la confianza de los usuarios hacia grandes empresas, imitando de la forma más precisa posible, una web que al usuario le resulte familiar y de la cual no tendría por qué dudar, como se puede ver en la falsificación de la Figura 3.

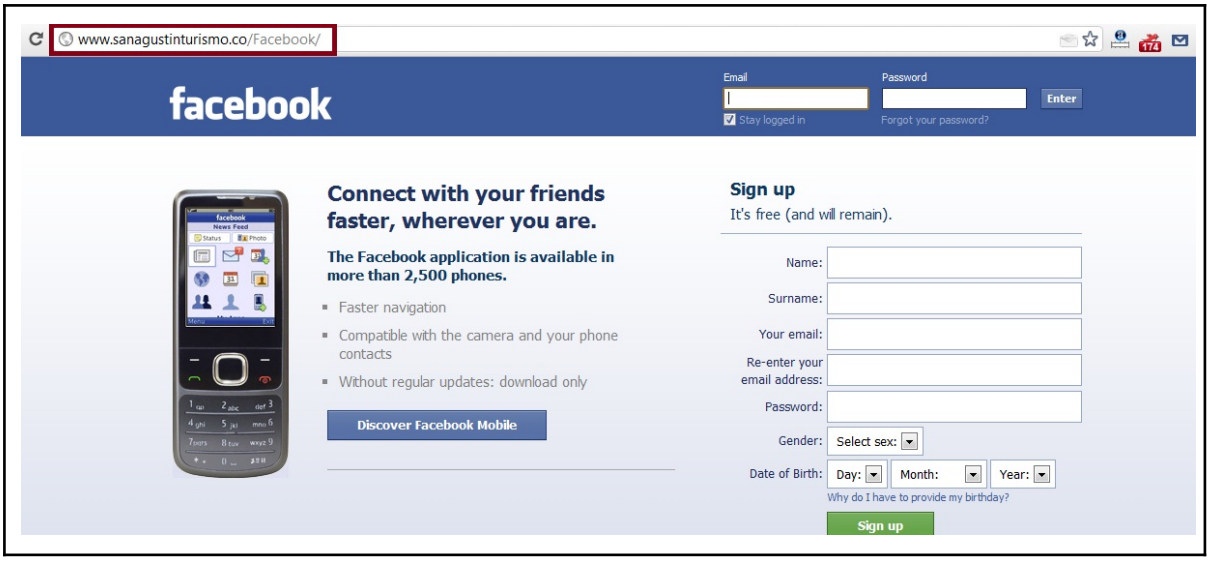

Figura 3. Web suplantada en un ataque de Phishing por TowardDataScience [33]

Por otro lado, busca generar en los usuarios la sensación de incertidumbre y de urgencia. Esto lo consiguen mediante correos electrónicos en los que se informa de un fallo en la seguridad de sus cuentas, un bloqueo de su acceso a ciertos servicios o cualquier otra situación que requiera una acción urgente. Además, para asegurarse que la víctima no tenga tiempo de pensar detenidamente o pueda consultar con alguien de su entorno, estos correos pueden llegar a mencionar un límite de horas (Figura 4) en las que la víctima pueda reaccionar, y de no ser así, la situación empeoraría.

Cabe mencionar que no siempre se utilizan webs falsificadas donde ingresar datos, también es común que los correos contengan archivos adjuntos. Estos, generalmente, están infectados con malware o incluyen un enlace a la web fraudulenta.

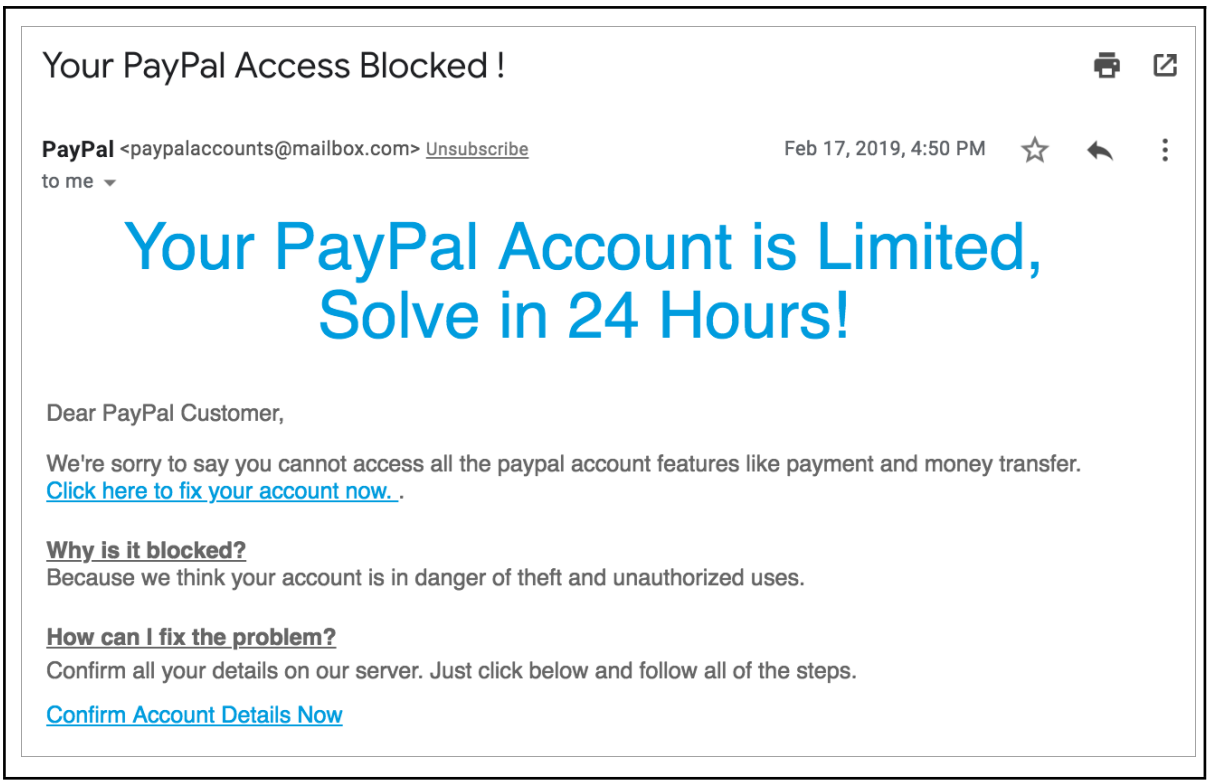

Figura 4. Presión ejercida por el uso de cuentas atrás por KhanAcademy [34]

Como resumen de definición, concluiré con la aportada en Léxico de Oxford [35]:
The fraudulent practice of sending emails purporting to be from reputable companies in order to induce individuals to reveal personal information, such as passwords and credit card numbers.
Traducción: La práctica fraudulenta de enviar correos electrónicos que pretenden ser de compañías acreditadas para inducir a las personas a revelar información personal, como contraseñas y números de tarjetas de crédito.

2.2 Tipos de Phishing

Si hablamos de los tipos de Phishing existentes, debemos hacer distinción entre una división según el tipo de víctima o una división según el medio que utilice. [36][37][38]

2.2.1 Según el tipo de víctima

Dependiendo de los objetivos del ataque realizado, podemos dividir el Phishing en:

Phishing (Spam Phishing)
Se envían los correos mediantes Spam a miles de correos electrónicos, generalmente obtenidos a través de repositorios. Estos repositorios se obtienen mediante fallos de seguridad en bases de datos que los hacen público o bien, comprándolos en la Dark Web. Generalmente se conoce simplemente como Phishing.

Spear Phishing
Se envían correos electrónicos a un objetivo focalizado intentado aprovechar la información conocida de la víctima. Se puede considerar Spear Phishing desde un sólo individuo a un grupo reducido.

En la Figura 5 podemos ver un ejemplo de cada uno.

Figura 5. Comparación entre Phishing y Spear Phishing [39]

Para el atacante, cada opción supone ventajas y desventajas respecto a la otra. Mientras que el Spam Phishing ataca a un mayor número de usuarios, las probabilidades de éxito (o incluso el beneficio que el atacante puede obtener de cada víctima) son menores que en un ataque dirigido.
En cuanto al coste de la campaña de Phishing, es difícil calcular qué opción es más barata de lanzar. Por un lado, el Spam sólo requiere de un repositorio de correos, que muchas veces se puede obtener de forma gratuita. Por otro lado, el spear Phishing requiere una investigación previa de la víctima, que puede ser más o menos exhaustiva y por tanto más o menos económica.

Dentro del Spear Phishing podemos encontrar otras variantes, las cuales tienen unas víctimas concretas:

- **CEO Fraud/Business Email Compromise :** Práctica consistente en engañar a la víctima haciéndole creer que el remitente del correo electrónico es un superior suyo dentro de la organización. Se solicita información de carácter confidencial con urgencia. Un ataque de este tipo bien elaborado puede llegar a copiar el formato y expresiones de la persona que trata de suplantar.
- **Whaling :** Ataque de Phishing donde la víctima no es un empleado cualquiera de la organización, sino un alto cargo. El nombre proviene de Whale (ballena) + Phishing, haciendo referencia a que busca pescar a los "peces gordos". Los ataques de este tipo suelen planearse detenidamente, buscando información sobre la víctima, sus contactos y cómo se comunica con ellos.

2.2.2 Según el medio

En función del medio utilizado para llevar a cabo el engaño, podemos distinguir entre 3 tipos principalmente:

Phishing (Spamming)

Cuando hablamos de Phishing generalmente nos referimos al que ocurre mediante correo electrónico, ya que es el medio más común.

Vishing (Voice Phishing)

Utiliza llamadas al teléfono de la víctima para obtener información personal o financiera. A menudo se utilizan llamadas automáticas que dirigen a los usuarios a otros medios u otras llamadas, llegando incluso a hablar con los propios atacantes, que se harán pasar por un empleado de una organización de confianza. Es frecuente ver cómo los atacantes utilizan aplicaciones u otras técnicas para falsificar o esconder sus números de teléfono.

Comúnmente, el motivo de estas llamadas fraudulentas es solucionar un problema de seguridad en un servicio. Los usuarios aportan la información necesaria para solucionar el supuesto problema por miedo a ser hackeados, irónicamente.

Smishing (SMS Phishing)

Utiliza mensajes de texto engañosos para estafar a sus víctimas (Figura 6). El objetivo es hacer creer que el mensaje proviene de una persona u organización de confianza, y así, lograr que la víctima realice una acción que de al atacante información vulnerable o que permita el acceso al dispositivo.

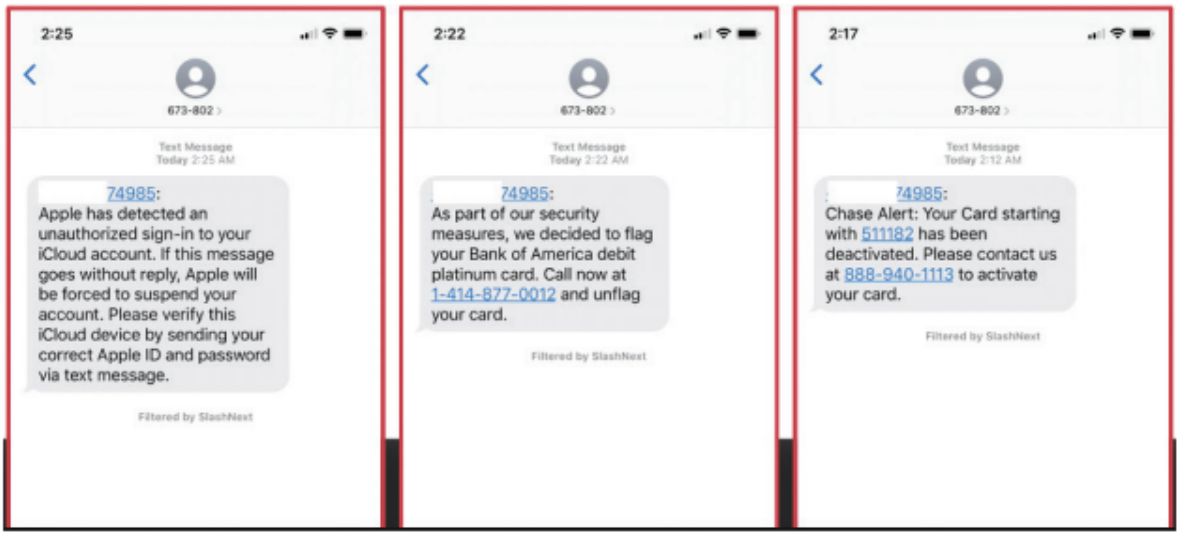

Figura 6. Uso de SMS fraudulentos en el Phishing por SlashNext [40]

Este tipo de ataque tiene éxito debido a dos factores:

1. Como se puede ver en la siguiente tabla (Tabla 1), la gente presta más atención a los SMS que a los correos.

	Leídos	**Respondidos**
Correos Electrónico	20%	6%
SMS	98%	45%

Tabla 1. Comparación entre Correos Electrónicos y SMS [41]

2. Los usuarios a menudo son más cuidadosos en sus ordenadores que en sus teléfonos móviles.

2.2.3 Otros tipos de Phishing

Ya hemos hablado de distintos ataques dependiendo del tipo de víctima al que fuera dirigido o del medio utilizado para extenderlos. Sin embargo, hay ciertos métodos que aún deben ser mencionados:

Clone Phishing

Para la realización de este ataque, el criminal diseñará un correo exactamente idéntico a otro que la víctima ya ha recibido. Todo será igual al mensaje previo, a diferencia del link adjunto en el mensaje, que en este caso dirigirá a la web fraudulenta del atacante. Es probable que el mensaje incluya alguna referencia a por qué se ha enviado el mismo mensaje repetido, explicando que se debe a una actualización o un fallo en el mensaje anterior.

Evil Twin

Si bien este tipo de ataques no es considerado por muchos como un método de Phishing, explicaré igualmente en qué consiste. Este ataque, a veces llamado la estafa de Starbucks (debido a la utilización en este tipo de establecimientos), consiste en la creación de una red WiFi que suplante a una red real. El criminal crea esta red usando el mismo SSID (el nombre de la red) o uno muy parecido al de la red WiFi original (aunque realmente no exista esa red en el establecimiento). De esta forma el atacante podrá capturar todo el tráfico de datos de los usuarios que se conecten a la red, incluyendo usuarios y contraseñas.

2.3 Vulnerabilidades aprovechadas

2.3.1 Malware adjunto

Pese a que al nombrar Phishing se tienda a visualizar como una página de inicio de sesión fraudulenta, también es frecuente encontrar malware en los correos electrónicos enviados. Mediante técnicas de ingeniería social se consigue que los usuarios descarguen archivos que, externamente, pueden parecer inofensivos. Como por ejemplo, un pdf informativo (como podría ser la factura de una supuesta compra realizada) o una actualización de una aplicación o software de confianza.

Cuando el usuario abra estos archivos, se ejecutará código malicioso en su dispositivo, pudiendo tener diversas consecuencias negativas. Una vez el Phishing ha conseguido que el usuario confíe en el archivo descargado, se llevarán a cabo otros ataques. Algunos de los más frecuentes son: la instalación de un keylogger, para registrar las teclas pulsadas; de un troyano, para poder tomar control del dispositivo del usuario; o un ataque de ransomware, donde toda la información de la víctima será “secuestrada” mediante encriptación para posteriormente pedir un rescate por ella. [30][38]

2.3.2 Domain Spoofing (idn homograph attack/unicode domain Phishing)

Una de las técnicas empleadas para que un dominio resulte aún más creíble es el uso del Domain Spoofing. [42][43] Este método se basa en replicar una web lo más detalladamente posible, haciendo que incluso la URL del dominio sea idéntica.

Pero, si los dominios son únicos, ¿Cómo pueden lograr que dos dominios sean idénticos?
Si bien muchas veces simplemente son parecidos (caixa-bank.es o caixabank.com en lugar del original caixabank.es), las técnicas de los criminales han avanzado hasta el punto que pueden ser visiblemente idénticas. Esto lo logran mediante el uso de IDN homograph attacks, también llamado Unicode Domain Phishing. Con este sistema, explotan el hecho de que dos caracteres distintos, de distintos alfabetos, sean idénticos. Por ejemplo el carácter latín 'a' es igual al carácter Cirílico 'а'.

Así, como se puede ver en la siguiente imagen (Figura 7), logran que dos URL que visualmente son iguales, redirigen a sitios distintos.

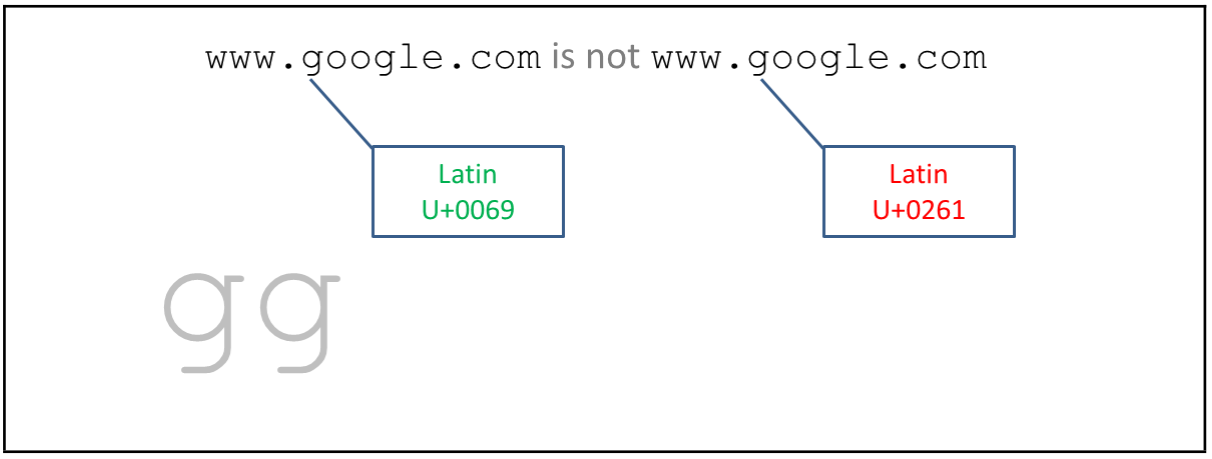

Figura 7. Uso de Unicode Domain Phishing [44]

Afortunadamente esta técnica cada día es menos usada, ya que la mayoría de los navegadores han implantado medidas de seguridad (Figura 8).
De esta forma, si intentamos acceder a *www.instagram.com* donde la segunda 'a' ha sido sustituida por el carácter cirílico 'а', el navegador nos mostrará lo siguiente:

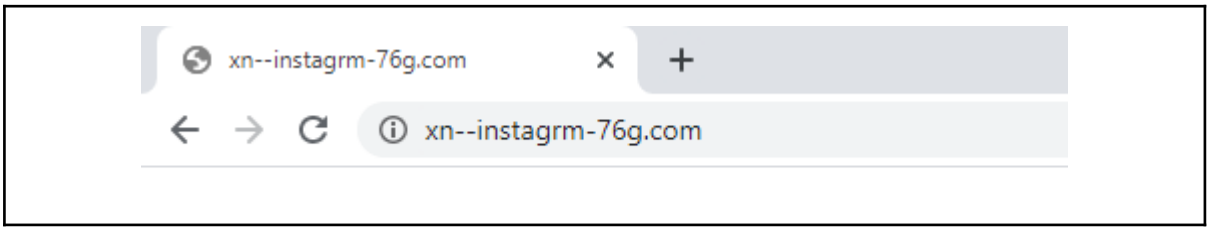

Figura 8. Comprobación de Unicode Domain Phishing en Google Chrome, elaboración propia.

2.3.3 HTTPS Phishing

Una de las recomendaciones más frecuentes para evitar los ataques de Phishing es comprobar que la web a la que accedes tiene el candado verde (indicando que cuenta con un certificado SSL válido) y que cuente con HTPPS.

A pesar de que es una valiosa recomendación, ya que nunca deberías introducir datos de valor si no se cumplen estas características, a día de hoy no es suficiente.
Como informa PhishLabs junto con APWG (Figura 9), en 2020 el 80% de las webs de Phishing contaban con HTTPS. [45]

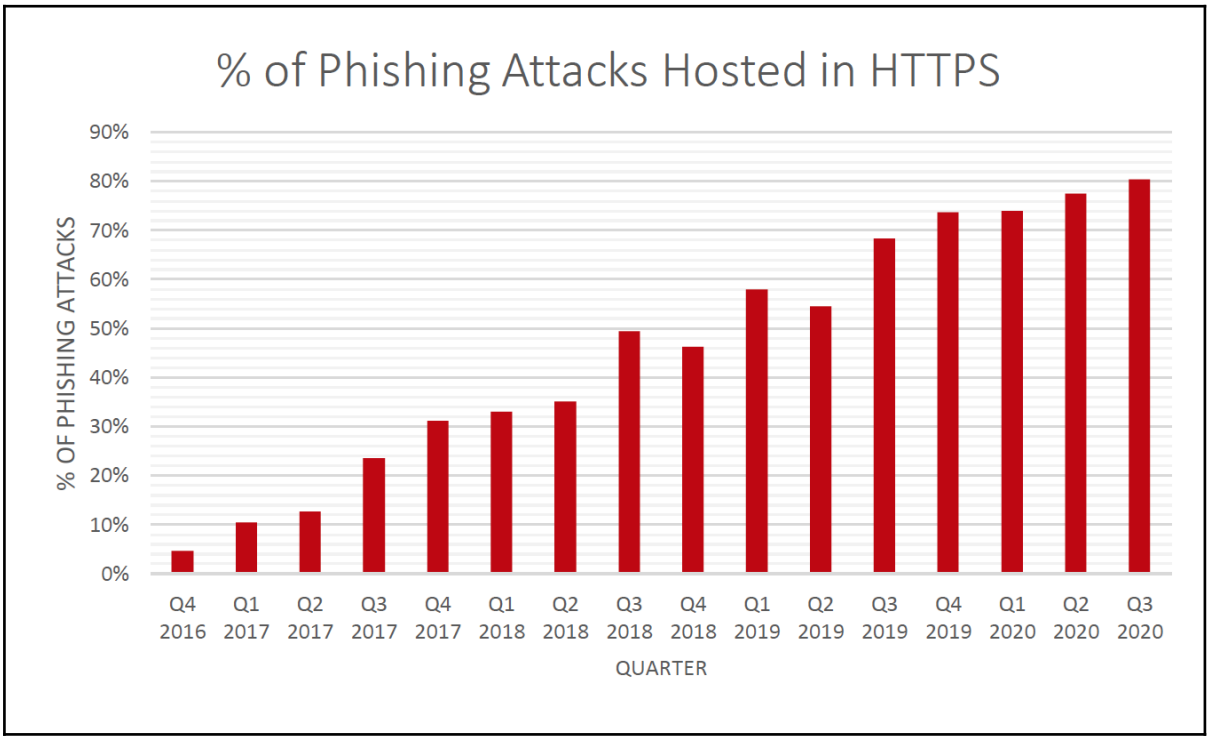

Figura 9. Porcentaje de ataques de Phishing realizados sobre HTTPS según PhishLabs. [45]

Además, el 40% de estos certificados fueron emitidos por Let's Encrypt, una de las autoridades de certificación más importantes, la cual permite la automatización de la emisión de certificados. [45]
Actualmente, hay 3 tipos de certificados SSL: *Extended Validation, Organization Validated* y *Domain Validated*. La mayoría de los atacantes (91.3%), usan este último tipo, el cual es el más débil, pero aun así, es suficiente para dar mayor credibilidad a la web fraudulenta y engañar a los usuarios.

2.3.4 Drive-by Phishing

Se basa en los Drive-by downloads. En estos ataques se aprovechan vulnerabilidades de páginas legítimas para introducir código malicioso en su interior. Este código (comúnmente en HTTP, PHP o Javascript) descarga de forma automática malware en el dispositivo de los usuarios que visiten la web. [46][47]
Esta técnica se utiliza tanto con webs legítimas infectadas, como en webs fraudulentas creadas por el atacante.

A día de hoy, según informa Eleven Research Team, se están empezando a utilizar los "drive-by virus", consistentes en un email HTML que podría descargar contenido en tu dispositivo simplemente por abrirlo (sin haber accedido a ningún enlace). [48]

2.4 Evolución en el tiempo

En este apartado haré un pequeño resumen de los orígenes del Phishing y cómo este ha evolucionado. [49]

En primer lugar, me gustaría hablar sobre el origen del término Phishing. Ya hemos mencionado que proviene del inglés "fishing", pero ¿por qué el uso de la 'ph' en lugar de la 'f'? Algunos de los primeros hackers que surgieron fueron denominados "phreaks". Por este motivo se utilizó la 'ph' para relacionar estos ataques con esa comunidad.

Se considera que la primera vez que se hace mención al término, es el 2 de Enero de 1996 en un grupo de noticias de Usenet llamado AOHell. Es además, en America Online (AOL) donde surgen los primeros ataques.
America Online, era por aquel entonces, el proveedor de acceso a Internet más usado. Los primeros ataques se basaban en el robo de contraseñas y, mediante algoritmos, creación de números de tarjetas de crédito hasta que alguno fuera correcto. Con estos números, registraban cuentas en AOL con las que Spammear a otros usuarios con diversos objetivos.
Estas prácticas se simplificaron con el uso de AOHell (Figura 10).

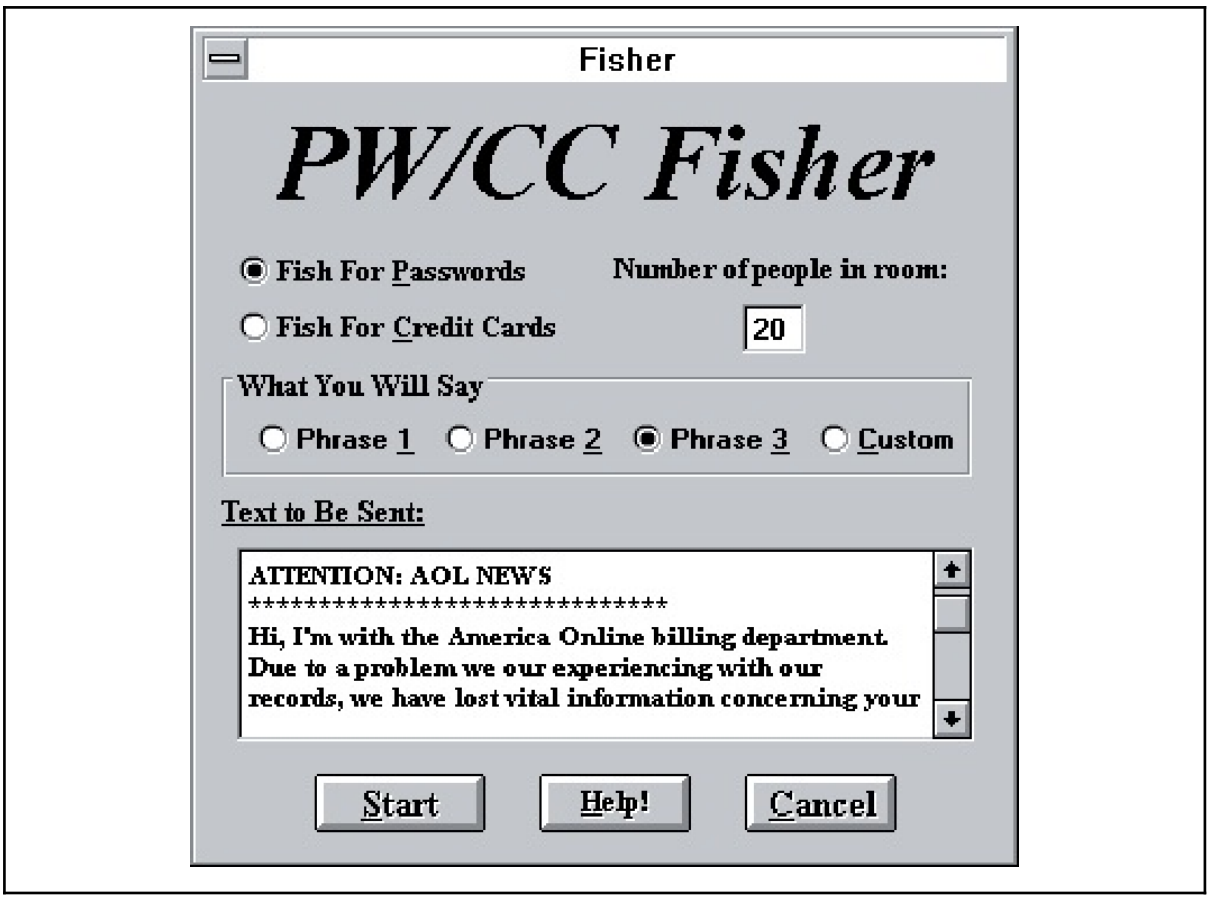

Figura 10. Interfaz de AOHELL [50]

En 1995, AOL implanta medidas de seguridad que haría que los hackers no pudieran generar números de tarjetas de crédito aleatorios.
Esto llevó a que los criminales refinaran sus técnicas: se harían pasar por empleados de AOL que necesitaban que se confirmara cierta información.

En 2001, el Phishing tomó una nueva dirección, los métodos de pago online.
Para finales de 2003 ya existían docenas de dominios que suplantaban las webs de Ebay y Paypal.
En 2004, los atacantes tuvieron una racha de éxitos atacando a bancos y sus clientes.

Acabando 2008, aparecen el Bitcoin y otras criptomonedas, lo cual permite a los atacantes que se realicen transacciones seguras y anónimas.

En Septiembre de 2013, el ransomware Cryptolocker infectó más de 250.000 ordenadores debido a descargas de webs comprometidas y a dos correos de Phishing distintos.

Se empiezan a utilizar ataques Phishing sobre HTTPS en 2017.

Durante 2018 y 2019, son muy comunes los ataques basados en Tarjetas de Regalo que ofrecían hasta 4.000$.

Comienza 2020 y la crisis del Covid-19. Los atacantes lo aprovechan usando temas como advertencias de Centros para el Control y Prevención de Enfermedades, teletrabajo, Netflix, etc.

A finales de 2020, la empresa más suplantada fue Microsoft, dada la importancia del teletrabajo (Microsoft Office y Azure son ampliamente usados por empresas).

2.5 Estadísticas

Por si a estas alturas aún hay alguien que no esté convencido de la importancia del Phishing, me dispongo a mostrar algunas estadísticas: [51][52][40][53][54]

-Según el FBI, el Phishing fue el cibercrimen más común en 2020.

-En el SonicWall Cyber Threat Report de 2021 se calculó que, en Junio de 2020, debido al teletrabajo, el Phishing en móviles aumentó un 37%.

-El Phishing ha crecido en más de 2 millones de casos de 2019 a 2020, principalmente por el covid-19 (Figura 11 y Figura 12).

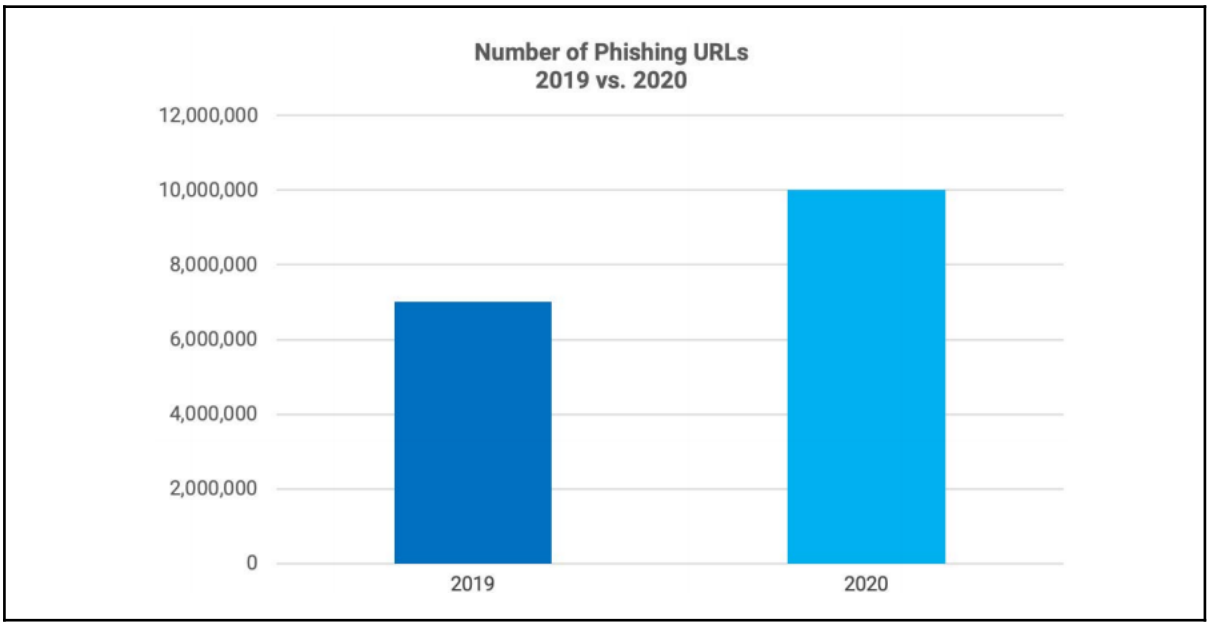

Figura 11. Comparación del número de URLs de Phishing de 2019 y 2020 por Slashnext [40]

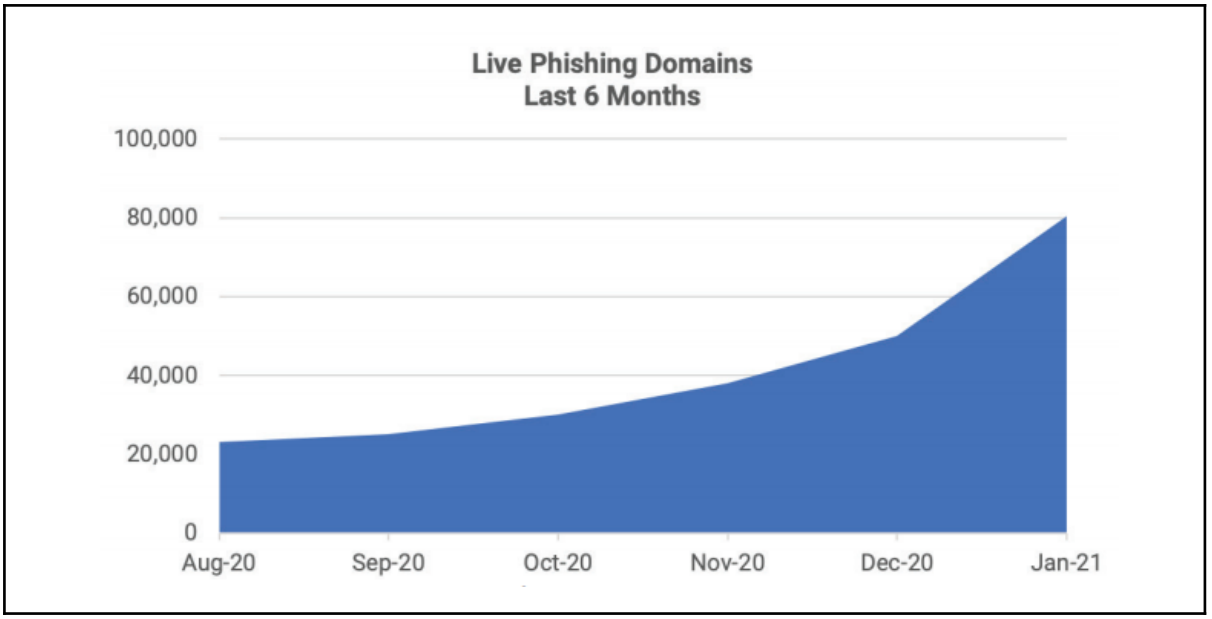

Figura 12. Dominios de Phishing activos en los últimos 6 meses de 2020 por Slashnext [40]

-El país que más Spam generó en 2020 fue Rusia (21.27%) (Figura 13)

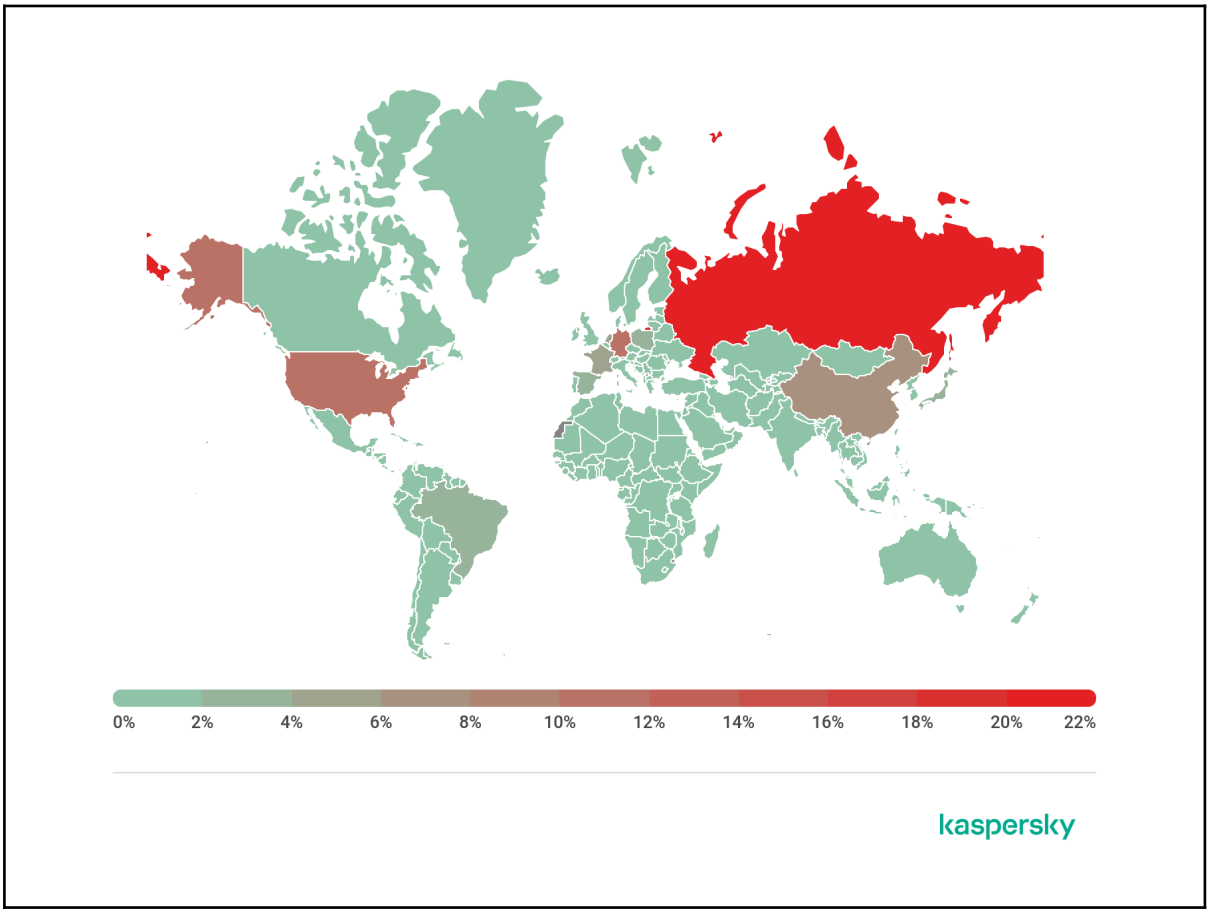

Figura 13. Porcentaje de Spam generado por países por Kaspersky [53]

-A nivel económico, el Phishing supone una gran cantidad de pérdidas. El FBI estima que sólo en 2018 se reportaron pérdidas de más de 1.200 millones de dólares.

-El último reporte del FBI estima 1.800 millones de dólares a lo largo de 2020, únicamente provenientes de BEC Phishing (Business Email Compromised).

-En 2018, recuperarse de un ataque de Phishing costaba a las empresas una media de 3.9 millones de dólares. Actualmente, esta cifra se ha mantenido, según informa IBM, siendo Estados Unidos el país donde a las empresas les sale más caro, con una media de 8.64 millones de dólares.

-Las marcas más suplantadas en lo que llevamos de 2021 aparecen en la siguiente imagen (Figura 14).

Figura 14. Marcas más suplantadas a principios de 2021 [51]

3 Simulación de un ataque

En la siguiente parte del trabajo se mostrará como se puede realizar un ataque de phising utilizando el framework de GoPhish [7][56][57][59][60]. Cabe resaltar que el objetivo del software es el de realizar test de penetración dentro de tu propia empresa, no el de realizar ataques reales.

Para nuestras pruebas, la simulación será realizada en un entorno de Windows 10.

3.1 Empezando en GoPhish

Como ya se ha comentado, GoPhish es un framework destinado comprobar cómo de expuesta está tu organización frente a un ataque de Phishing.

Esta poderosa herramienta de código abierto ofrece resultados en tiempo real a través de su intuitiva interfaz de usuario. Posee una instalación muy sencilla, compatible con Windows, Mac OSX y Linux.

Una vez descargado desde su github oficial[55], descomprimimos el .zip y nos encontraremos los siguientes archivos (Figura 15).

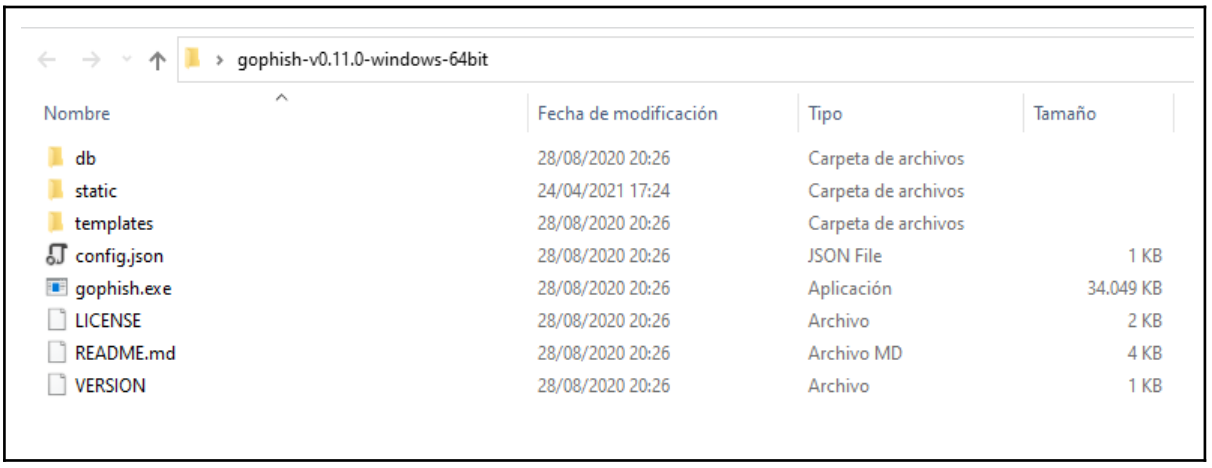

Figura 15. Contenido de la carpeta descargada de GoPhish. Elaboración propia

Para poder realizar nuestro ataque sólo tendremos que utilizar **config.json** y **gophish.exe**

Empezaremos por el config.json, que nos servirá para decidir la dirección en la que queremos que se ejecute tanto el admin server como el phish server. Además, podremos configurar qué certificado queremos que se use en estas direcciones.

Inicialmente el archivo contiene la información de la Figura 16:

Figura 16. Contenido del archivo config.json. Elaboración propia.

Como podemos ver en el apartado listen_url de cada servidor, las direcciones a las que deberemos acceder por defecto serán la 127.0.0.1:3333 para la página de administración y a la 0.0.0.0:80 para la página de Phishing.

Ejecutemos ahora gophish.exe. Debemos tener en cuenta que el servidor de Phishing se desplegará en el puerto 80, y como cualquier puerto inferior al 1024, necesitará privilegios de administrador para poder acceder.

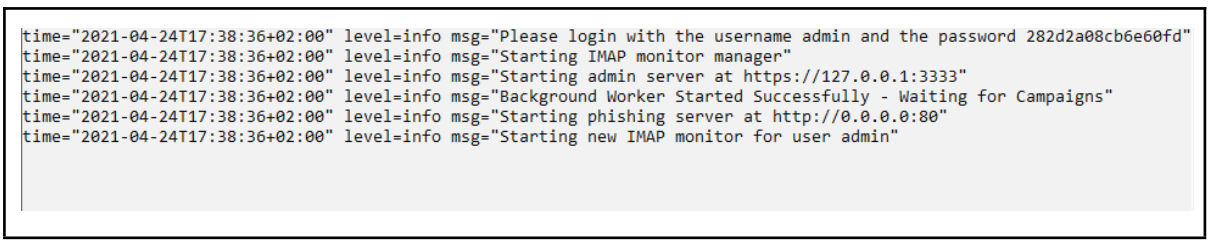

Figura 17. Respuesta del terminal al ejecutar gophish.exe. Elaboración propia.

En el terminal nos aparecerá información (Figura 17), indicándonos que los servicios se han iniciado con éxito (en caso de no mostrar ningún error) y mostrándonos en qué dirección IP lo han hecho.

NOTA 1: En el caso de que haya algún error es posible que la ventana del terminal abierta por el .exe se cierre inmediatamente. Para poder ver el error debemos abrir un Símbolo de Sistema (Tecla Windows + R, escribimos cmd y le damos a enviar), acceder a la ruta donde tenemos el ejecutable e iniciarlo desde ahí. [61]

NOTA 2: En el caso de que se muestre el siguiente error:
level=fatal msg="listen tcp 0.0.0.0:80: bind: An attempt was made to access a socket in a way forbidden by its access permissions."
Lo más probable sea que necesites permisos de administrador o que haya otro servicio usando el puerto 80, como puede ser una página de IIS o Skype. [62]

Si no ha habido ningún error, podemos acceder a https://127.0.0.1:3333 para logearnos en la página de administración. El usuario será admin, y la contraseña será la que os aparezca en el terminal (hasta que la cambiéis, una vez se inicie sesión por primera vez).

Cuando hayáis iniciado sesión os encontraréis con la siguiente página (Figura 18).

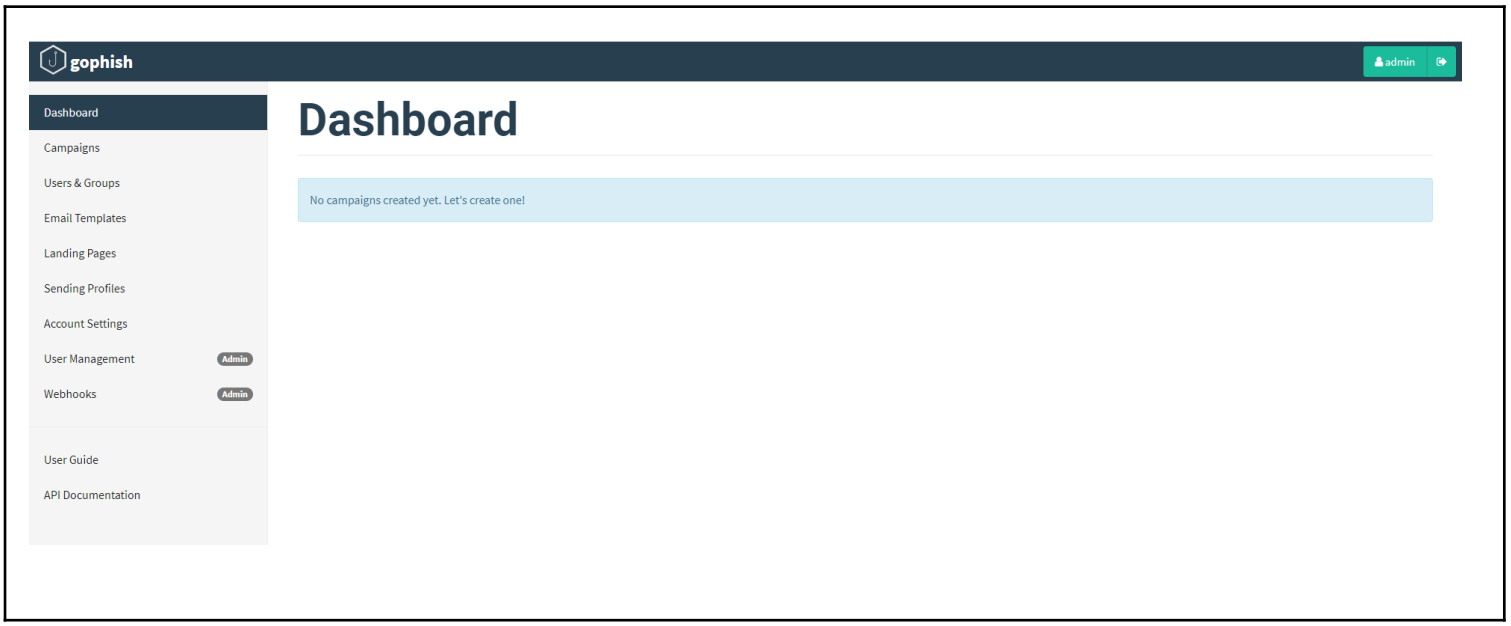

Figura 18. Página Dashboard de GoPhish. Elaboración propia.

Las 6 primeras opciones son con las que trabajaremos. Aún así, daré primero una breve definición de los otros apartados:

Account Settings: En este apartado podremos configurar opciones en los reportes, la interfaz o la cuenta.

User Management: Desde aquí, podremos gestionar los distintos usuarios que podrán acceder a la interfaz de administración. Por defecto, sólo existirá el usuario admin, pero podemos crear los que necesitemos.

Webhooks: En el caso de que queramos recibir notificaciones HTTP de los resultados de nuestras campañas en un endpoint de nuestra elección podremos hacerlo desde aquí. Será necesario poseer los permisos de administrador.

User Guide:Si tuviéramos dudas a la hora de trabajar con el framework, podremos acceder a esta guía para buscar información.

Api Documentation: Aquí podremos investigar cómo está desarrollada la Api de GoPhish. Para nuestro trabajo no será necesario consultarla.

En los siguientes puntos comenzaremos a configurar lo que necesitaremos para lanzar nuestra primera campaña.

3.2 Sending profiles

El primer paso a la hora de realizar nuestro ataque es decidir a quién queremos suplantar.

El campo name servirá sólo para que podamos distinguir los distintos perfiles que creemos.

En el apartado From colocaremos el remitente que queremos que se muestre en el correo que reciba la víctima. Además, añadiremos antes un nombre para hacerlo más creíble.

Para poder enviar los correos, es necesario utilizar un servidor SMTP. Dado que en nuestro caso se harán pruebas de tamaño reducido podemos optar por el servidor de google (smtp.gmail.com:587). [58]
En el caso de querer realizar pruebas a gran escala deberíamos utilizar otro. El propio framework en su documentación [60] recomienda el uso de MailHog.

Los últimos campos que harán falta rellenar son username y password. Debemos colocar los datos del correo electrónico real que se utilizará para la realización del envío. Esta dirección no debería ser mostrada, pero como se verá más adelante, las medidas de seguridad actuales, hacen que no se vea el utilizado en el campo From si no el real. Por este motivo, conviene crear un correo lo más parecido posible a uno del usuario real.
En nuestro caso utilizaremos secureinstagram34@gmail.com para suplantar a Instagram y amazonservices.secure@gmail.com para suplantar a Amazon.

Una vez esté listo podemos guardar el perfil y pasar al siguiente paso.

En nuestro caso (Figura 19) se utilizarán 2 perfiles. El primero suplantará a Instagram [63] y el segundo a Amazon. [64]

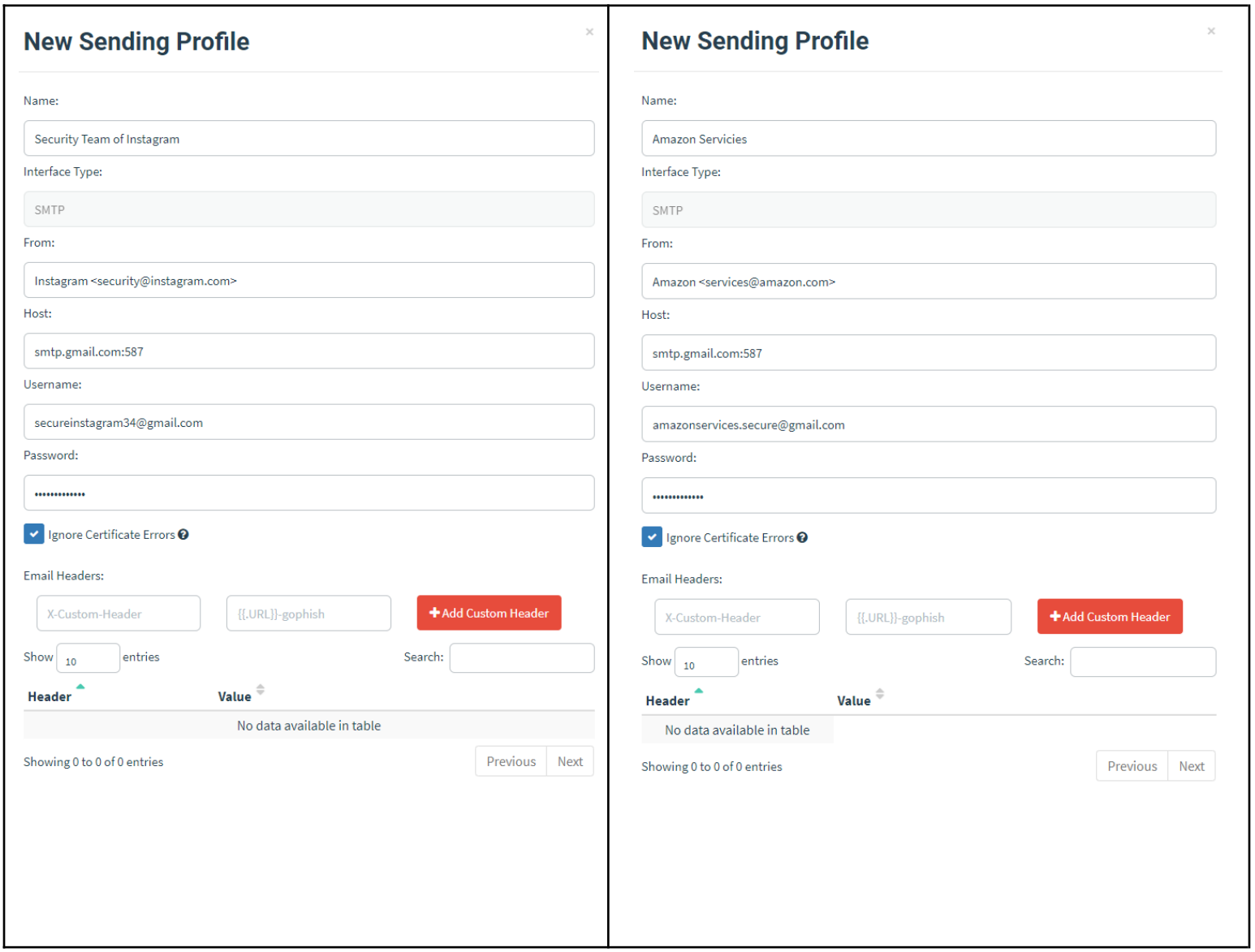

Figura 19. Perfiles utilizados en la simulación en GoPhish. Elaboración propia.

3.3 Email Templates

Ahora podremos crear la plantilla que usará el correo electrónico que enviemos.

Al igual que antes, el campo Name será únicamente para identificar la plantilla dentro de nuestra interfaz de GoPhish.

En el campo Subject escribiremos el asunto que queremos que se le muestre a nuestra víctima en el correo. Como se puede observar en la Figura 20, en este campo hemos utilizado la variable {{.Email}}. Esto nos permite que el mensaje parezca dirigido unicamente a la víctima, pues esta variable se sustituirá por su valor dentro de la definición del usuario. Más adelante, en el apartado 3.5 Users & Groups se detallará una lista con las posibles variables a utilizar.

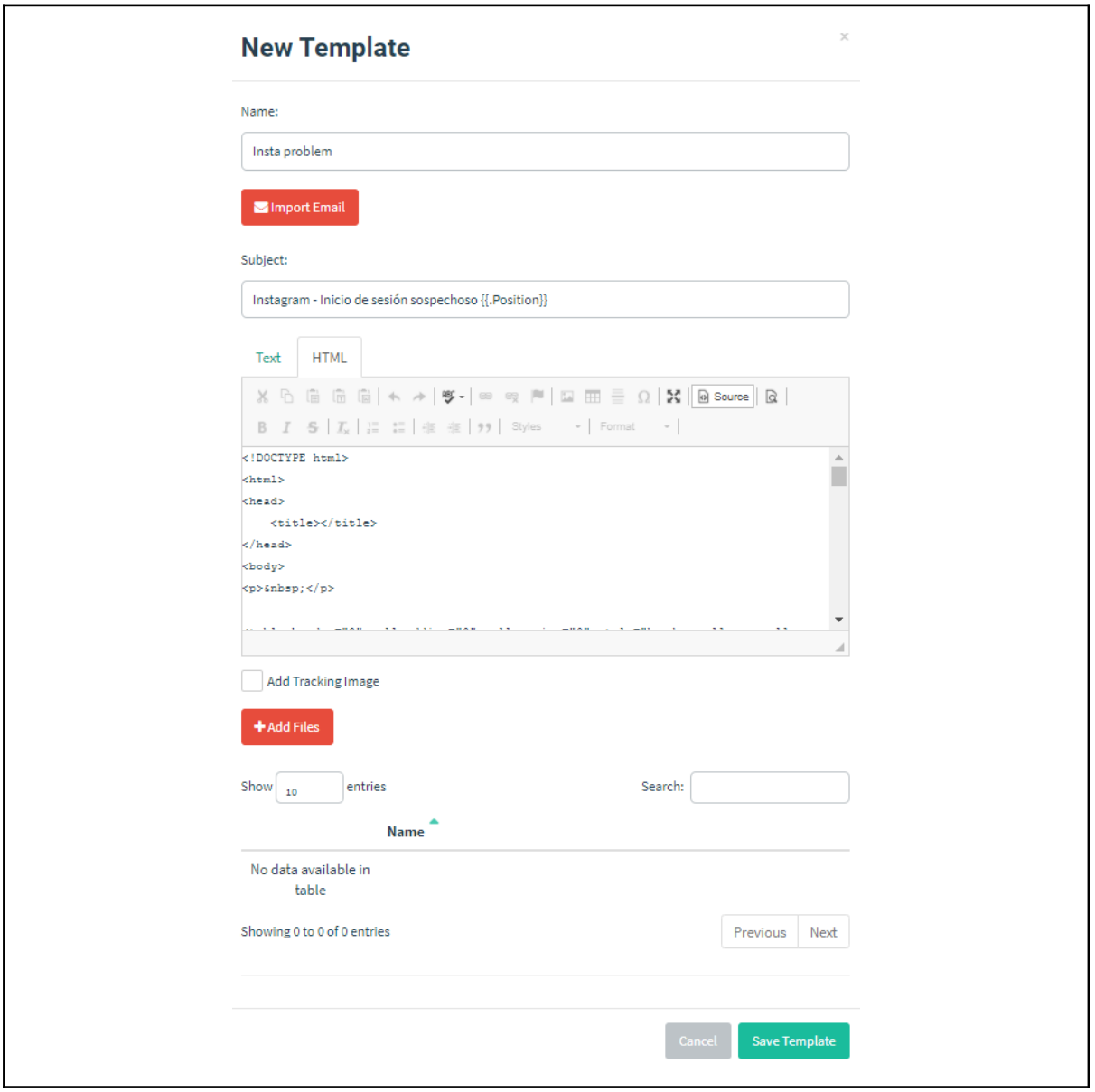

Figura 20. Ejemplo de plantilla de correo para la simulación en GoPhish. Elaboracion propia

A continuación, podremos definir cómo será el cuerpo de nuestro mensaje. Para ello contamos con una herramienta que nos permite diseñarlo en modo Texto y en modo HTML:

- El modo Texto nos servirá para realizar pequeñas pruebas en texto sin formatear.

- El modo HTML es el que utilizaremos para generar el mensaje lo más similar posible al original. Uno de los puntos fuertes de este framework es este editor de HTML, pues nos permite intercambiar entre código fuente y un modo visual donde podremos ir viendo el progreso y corregir los detalles que consideremos.

A la hora de generar una plantilla, una de las opciones más cómodas es importar un correo. Para hacer esto, pulsaremos sobre Import Email, localizaremos en nuestra bandeja de entrada el correo que queremos copiar y le daremos a Mostrar Original. Desde la ventana que se abrirá, podremos copiar el contenido en modo Raw y pegarlo en la casilla de GoPhish.

Otra opción, por la que yo he optado para la realización de mis correos, es mirar el código HTML del programa y modificarlo a nuestro gusto.
Para poder trabajar cómodamente, he modificado el código en Notepad++ [65], ya que uno de los problemas que tiene el editor de HTML en código fuente de GoPhish, es que elimina las indentaciones, haciendo que entender el código pueda resultar complejo.

Tras modificar los correos a nuestro gusto, quitando contenido innecesario y cambiando algunas cosas como los enlaces, podemos comprobar el resultado en las siguientes imágenes (Figura 21 y Figura 22).

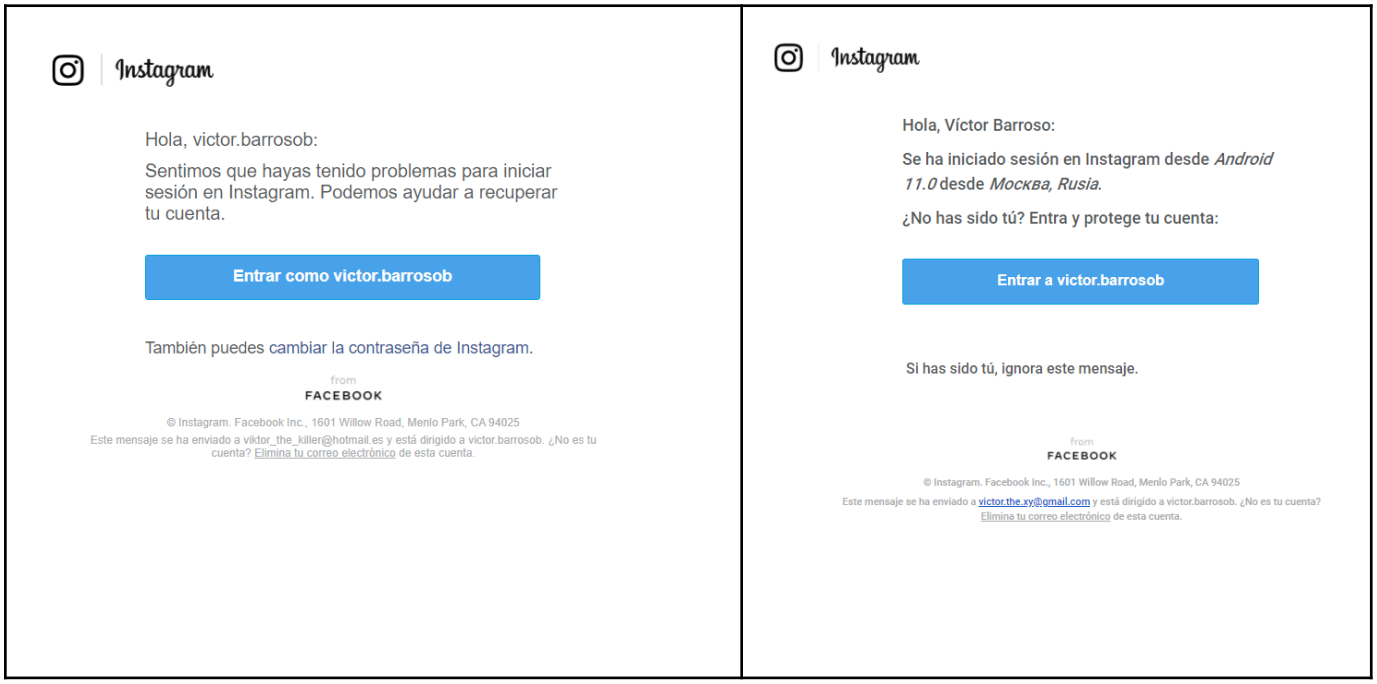

Figura 21. Mensaje original de Instagram (Izquierda) y el mensaje generado copiando el estilo (Derecha).

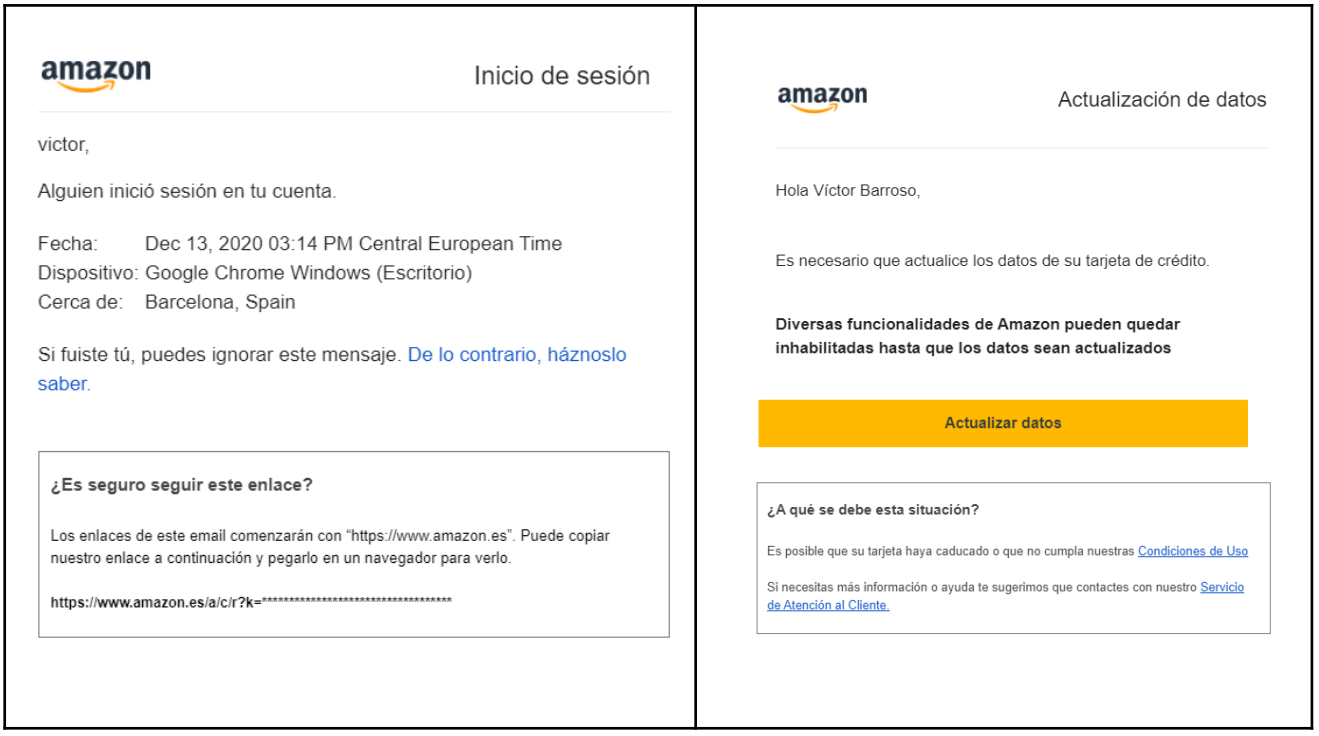

Figura 22. Mensaje original de Amazon (Izquierda) y el mensaje generado copiando el estilo (Derecha)

3.3.1 Ingeniería social de los correos.

Me gustaría aclarar por qué se han elegido estos diseños:

Instagram:
En el primer caso se ha decidido copiar un correo de recuperación de contraseña mandado por Instagram.
En nuestro versión hemos hecho creer a la víctima que su cuenta ha sido vulnerada por un usuario de Android desde Rusia, país que por su laxa legislación frente a los cibercrímenes suele relacionarse con estos delitos. Se le indica al usuario que si quiere proteger su cuenta debe seguir el enlace (el cual posiblemente bloquee futuros intentos de inicio de sesión sospechosos).
El correo se refiere a la víctima por su nombre y apellido, y muestra también su usuario de Instagram. Estos datos harán creer al usuario que es una fuente legítima quien manda el mensaje, ya que posee esa información. Sin embargo, este tipo de información es muy fácil de conseguir a día de hoy con un poco de investigación previa a través de Internet.

Amazon:
En este caso se ha usado un correo informativo de Amazon sobre un inicio de sesión. Los correos de Amazon son de diseños muy variados, no siguiendo un patrón muy claro. Se ha utilizado el correo de la Figura 22 porque representa bastante bien el estilo de la empresa, y puede ser más creíble que otros correos más sencillos que envían, como el siguiente (Figura 23).

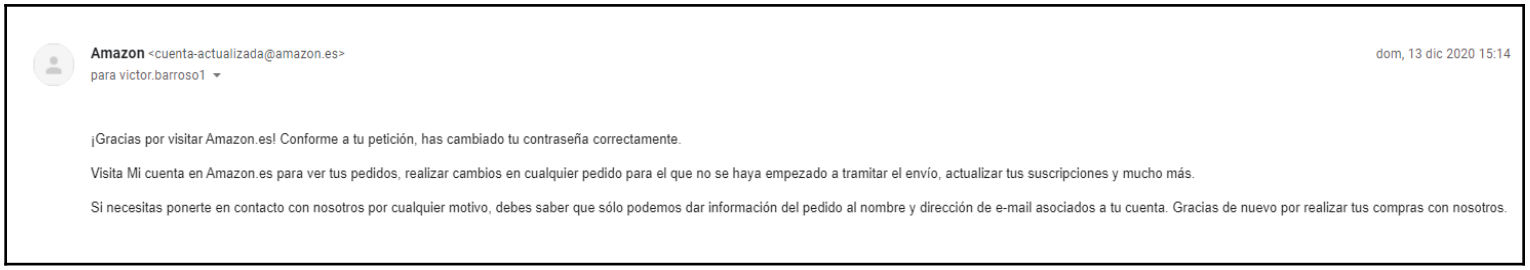

Figura 23. Correo real de Amazon de un diseño minimalista

Este ataque es más ambicioso que el anterior, pues ahora no se busca realmente los datos de sesión, lo que se está atacando es una tarjeta de crédito, utilizando a Amazon como empresa de confianza al que le darías dicha información.
En letra negrita se ha resaltado que algunos servicios de Amazon pueden desactivarse si no se arregla el problema, como por ejemplo, el servicio Amazon Prime Video. Causando así, que las víctimas tengan la necesidad de solucionarlo cuanto antes para evitar problemas.

Para finalizar, en el ataque se ha añadido un cartel indicando a qué se puede deber esta situación, para que los usuarios más desconfiados puedan entender por qué Amazon te está pidiendo esta información. Como posibles motivos se mencionan la caducidad de la tarjeta de crédito o un incumplimiento de las Condiciones de Uso, las cuales se adjuntan a través de un enlace (a sabiendas que serán muy pocos los usuarios que se lo leerán para comprobar si realmente se incumple algo).

Junto a esta explicación de los motivos, se ha añadido la recomendación de “ante cualquier duda contactar con el servicio técnico de Amazon”.

NOTA 3: Una posible mejora de cara al futuro consistiría en suplantar también el servicio técnico de la plataforma, para que si algún usuario duda de la veracidad del correo, el propio atacante pueda tratar de convencerlo u obtener más datos del usuario.

En ambos mensajes, el botón incluye un enlace a la variable {{.URL}}, que contiene la dirección del servidor de Phishing (0.0.0.0 en estos momentos) junto con un valor único que se generará para cada víctima (RId).

3.4 Landing Pages

Lo siguiente que debemos hacer es preparar las Landing Pages correspondientes a los correos. Es decir, las páginas a donde se enviará a las víctimas para obtener sus datos.

Recordemos que esta página será replegada por defecto en la dirección IP 0.0.0.0:80. De esta forma se podrá acceder a la web de forma local, pero si queremos que se pueda acceder desde Internet, tendremos que hacer algunos cambios que se explicarán más adelante (Apartado 3.7.1).
Cuando nos dispongamos a crear la página, al igual que antes, tendremos la opción de importar directamente el sitio web, esta vez mediante su URL.
Para el desarrollo de nuestra web, optaremos por copiar el código HTML de la página a replicar y modificarlo a nuestro gusto en Notepad++. Recomiendo además, tener el archivo HTML que estamos escribiendo abierto en el navegador, para poder ver el resultado que vamos obteniendo.

En el caso de Instagram, hemos obtenido distintos código HTML (Figura 24) y a partir de ahí, hemos generado una página que sigue el mismo estilo de diseño (Figura 25).

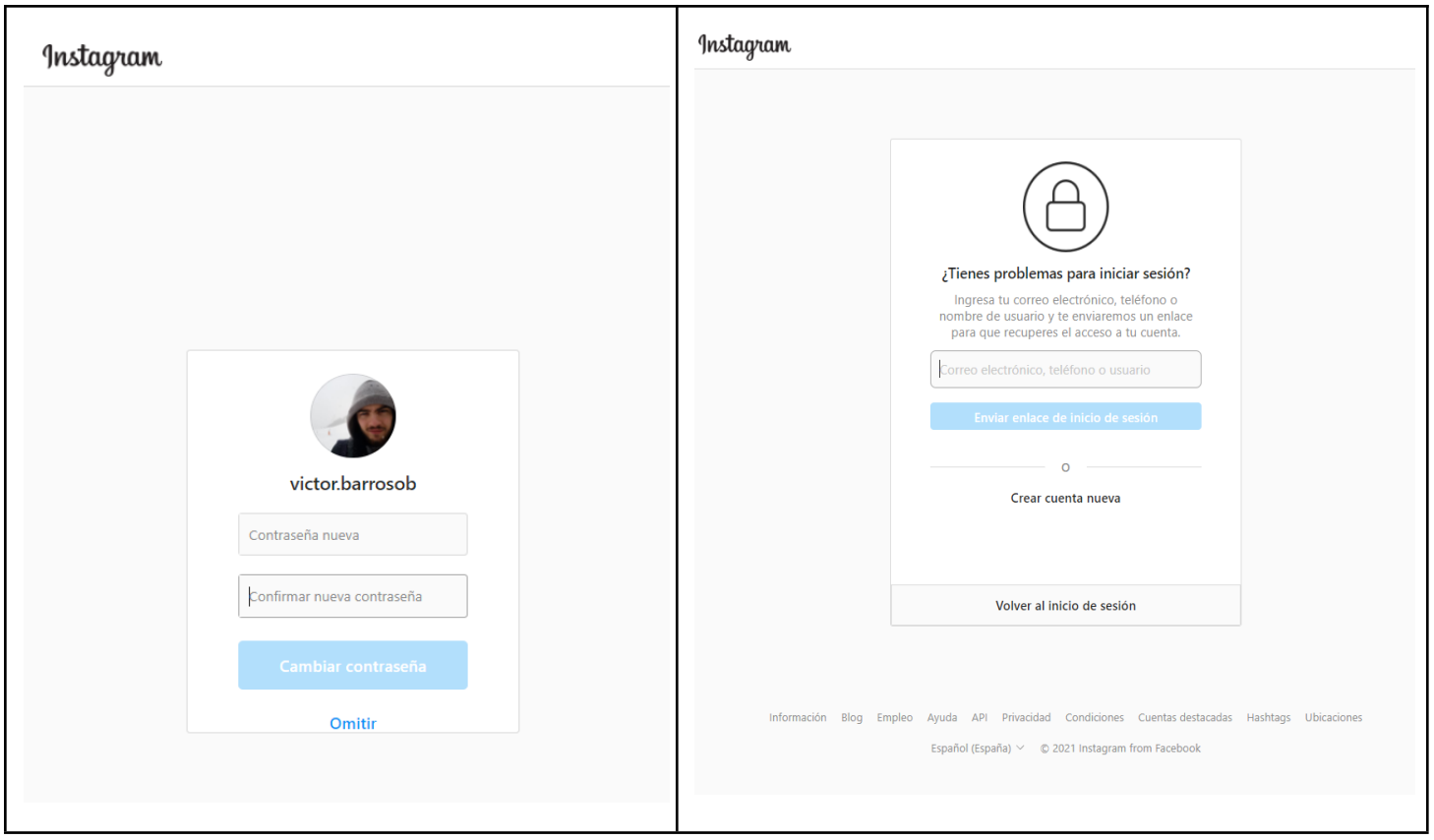

Figura 24. Páginas reales de restauración de contraseña

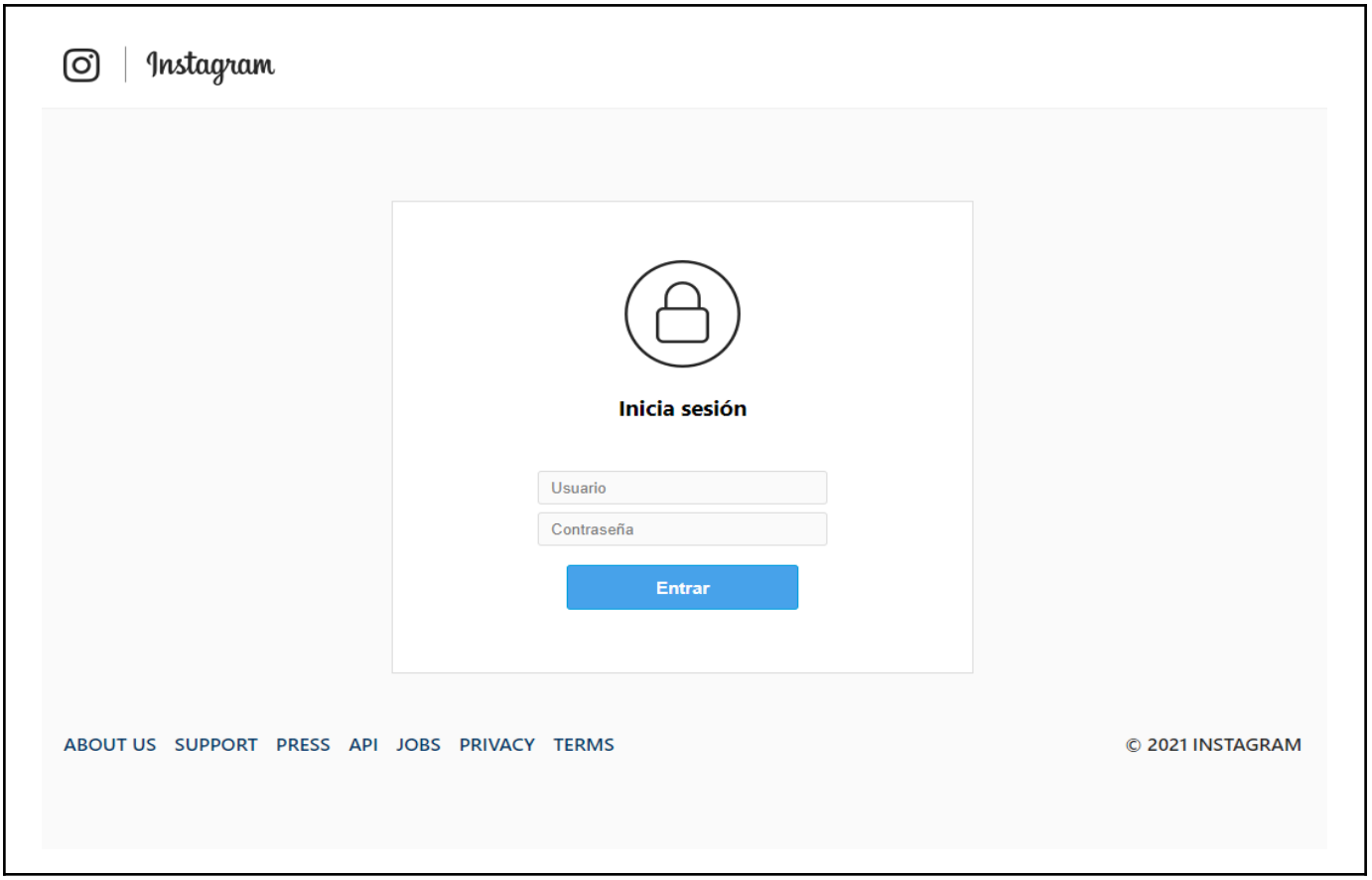

Figura 25. Página fraudulenta de inicio de sesión.

En el caso de Amazon (Figura 26) hemos copiado la página de Inicio de sesión, sólo que ahora te pedirá los datos de tu tarjeta.

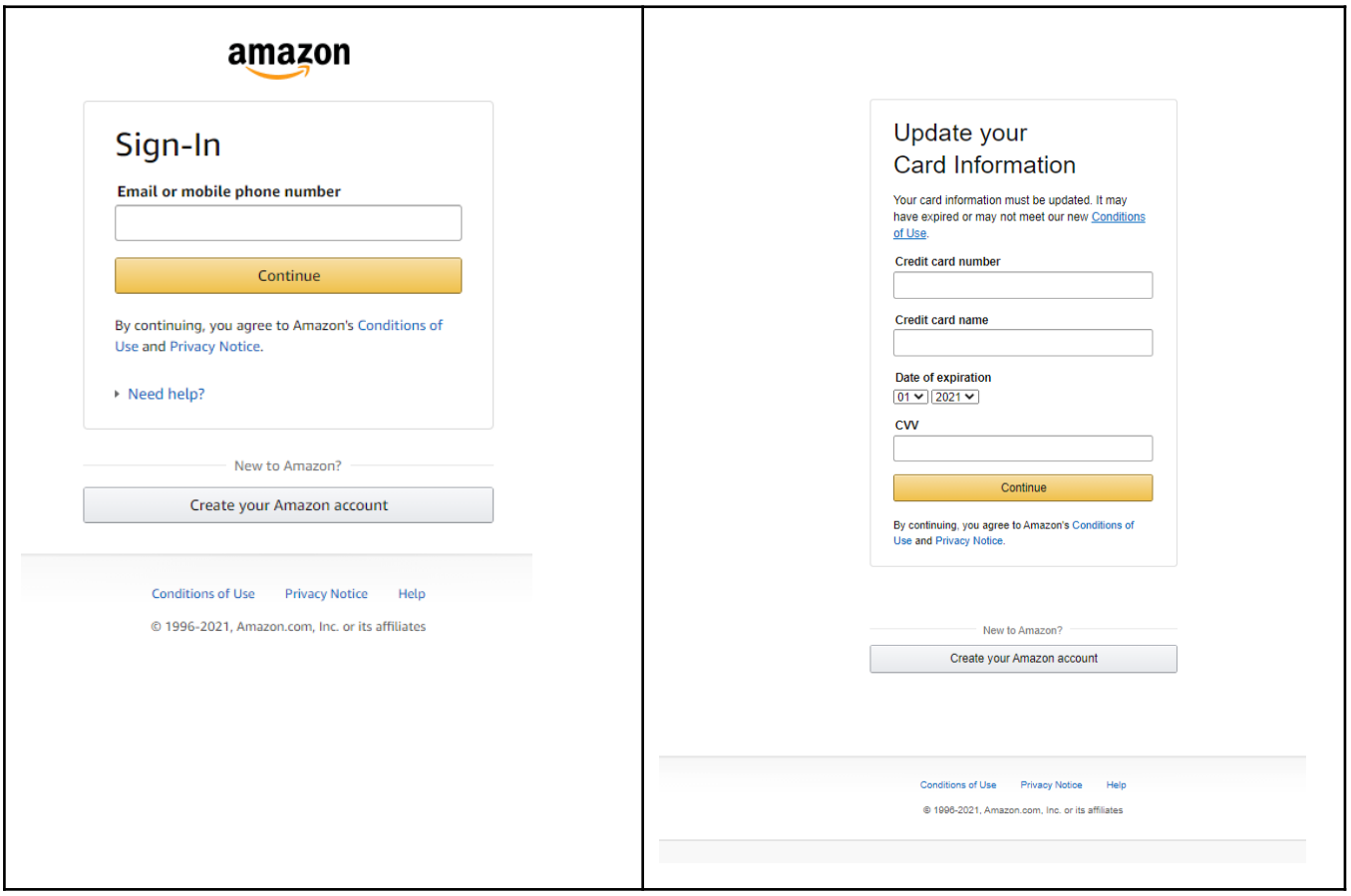

Figura 26. Página real de inicio de sesión en Amazon (Izquierda) y página fraudulenta de robo de datos bancarios (Derecha)

Una vez estemos satisfechos con nuestro código HTML, lo llevaremos a GoPhish y lo añadiremos como una nueva Landing Page.

Justo debajo del espacio para colocar nuestro código, disponemos de dos casillas.

- La primera nos permitirá capturar la información que contenga el formulario de nuestra página, a excepción de las contraseñas. Para hacer pruebas, con esta casilla, es más que suficiente.
- En el caso de que también queramos observar la contraseña (no entraré en el debate ético que esto supone) debemos marcar la segunda casilla también.

Es importante aclarar que para que GoPhish pueda capturar la información el formulario debe cumplir:

- El atributo *action* del *form* debe ser “”
- El atributo *method* del *form* debe ser POST
- Los inputs que desees capturar, deben de tener un atributo *name.*

Por último nos pedirá una web a la que redireccionar. Esto no es obligatorio, pero si redireccionamos al usuario a la web real (en la que posiblemente tenga la sesión iniciada), dará la sensación de que todo se ha tramitado con éxito.

Cuando tengamos la página lista, tendremos la tentación de acceder a ella a través de la dirección 0.0.0.0:80, pero aún nos falta un paso antes de poder visualizarla. Esto se debe a que GoPhish generará un RId único para cada objetivo como ya hemos mencionado, así que deberemos crear a nuestras víctimas para poder ver el resultado final.

3.5 Users & Groups

Lo último que debemos configurar antes de lanzar la campaña de Phishing, serán nuestras víctimas. Desde el panel de Users & Groups podremos crear distintos grupos, y en cada uno de ellos añadir las víctimas que consideremos. En nuestro caso habrá 2 grupos, uno con distintos correos a los que tengo acceso, y otro con correos de "víctimas", siendo estas últimas gente de mi entorno que me han concedido el permiso para la realización de estos ataques a modo de prueba.

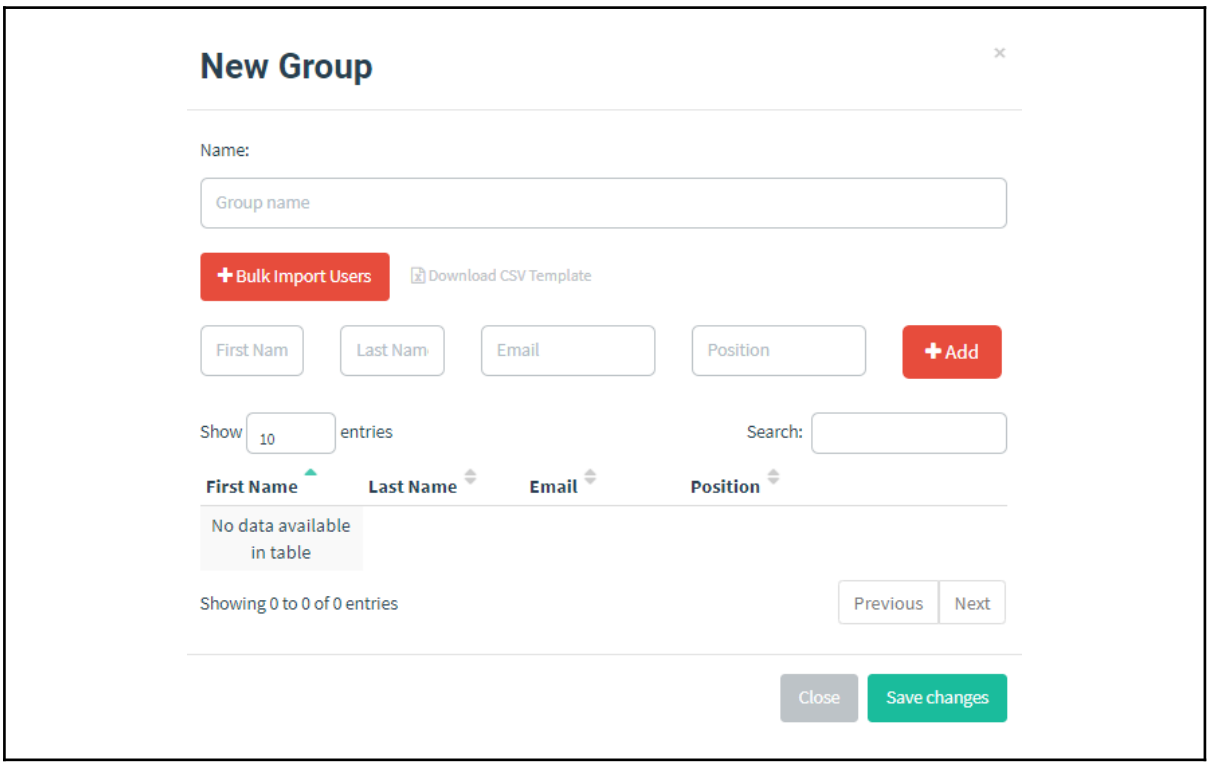

Figura 27. Creación de Grupos de usuarios en GoPhish

Como se puede ver en la Figura 27, dentro de cada grupo podremos añadir usuarios manualmente o bien importarlos de un archivo CSV. En el caso de añadirlos manualmente deberemos indicar diversos campos:

- First Name y Last Name: Nombre y Apellido de la víctima
- Email: El correo electrónico que se utilizará para el envío del ataque.
- Position: Tradicionalmente para indicar la posición dentro de la empresa. En nuestro caso hemos utilizado este campo para añadir el nombre de usuario de la red Instagram, pudiendo así añadirlo donde queramos usando la variable {{.Position}}.

A parte de la variable {{.Position}}, podemos usar también los otros campos, junto con algunas variables más. A continuación, muestro la lista completa (Tabla 2):

Variable	Descripción
{{.RId}}	Identificador de una víctima concreta en una campaña concreta
{{.FirstName}}	Nombre de la víctima
{{.LastName}}	Apellido de la víctima
{{.Position}}	Posición de la víctima
{{.Email}}	Email de la víctima
{{.From}}	Email suplantado
{{.TrackingURL}}	URL del controlador de seguimiento
{{.Tracker}}	Alias de
{{.URL}}	URL de la página de Phishing
{{.BaseURL}}	URL de la página de Phishing sin la ruta y sin el parámetro RId.

Tabla 2. Descripción de las variables existentes en GoPhish. [66]

Estas variables son de gran importancia a la hora de realizar un ataque. Podremos añadir tanto en el correo como en la propia página de Phishing, permitiéndonos que un ataque en vez de estar dirigido a “Querido cliente” esté dirigido a “Estimado Víctor Barroso”, haciéndolo mucho más personal y confiable.

3.6 Campaigns

Ya podemos comenzar a lanzar nuestras campañas y comprobar los resultados que estas tienen en tiempo real.

A la hora de crear la campaña nos pedirá un nombre, que seleccionemos las distintas plantillas que hemos creado, una URL (que será 0.0.0.0, la que seleccionamos previamente) y podremos elegir una fecha en la que lanzar la campaña.

NOTA 4: Antes de lanzar la campaña, y si el correo desde el que se enviarán los mensajes es Gmail, deberemos acceder a la configuración de seguridad de la cuenta y permitir el acceso a de aplicaciones poco seguras (Figura 28).

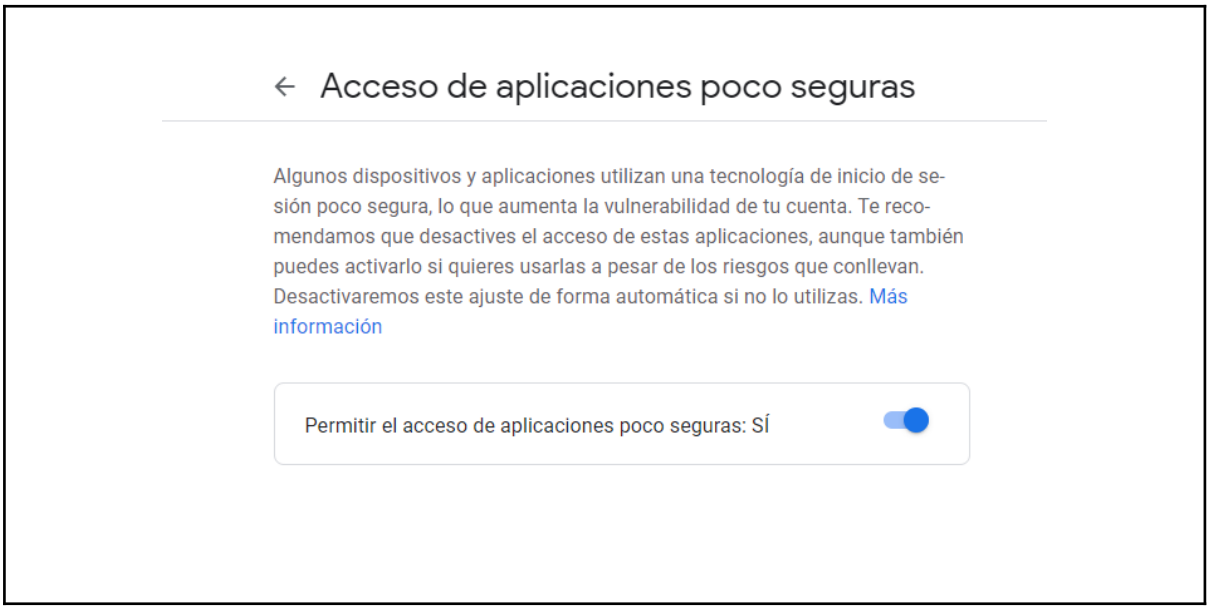

Figura 28. Configuración de Gmail sobre acceso a Aplicaciones poco seguras [67]

Una vez lanzada la campaña, nuestra víctima recibirá en su correo el mensaje que hemos diseñado. El enlace (URL + RId) le llevará a nuestro servidor de Phishing, donde tras rellenar el formulario y enviarlo, podremos visualizarlo en el panel de la campaña de GoPhish.

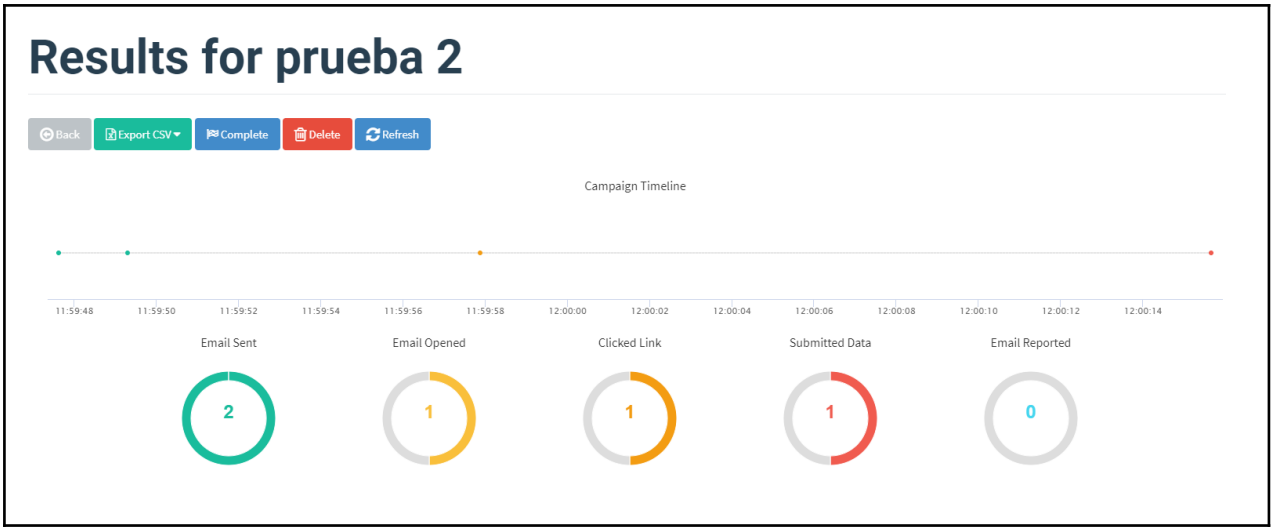

Figura 29. Resumen del resultado de una campaña en GoPhish

Por un lado podemos ver un breve resumen de cómo ha evolucionado la campaña a lo largo del tiempo (Figura 29), junto con los datos de cuanto ha conseguido penetrar nuestro ataque.

Más abajo tenemos los detalles de nuestra campaña, donde se detalla usuario por usuario la respuesta que ha tenido nuestro ataque (Figura 30). Aquí es donde podremos consultar los datos facilitados por la víctima.

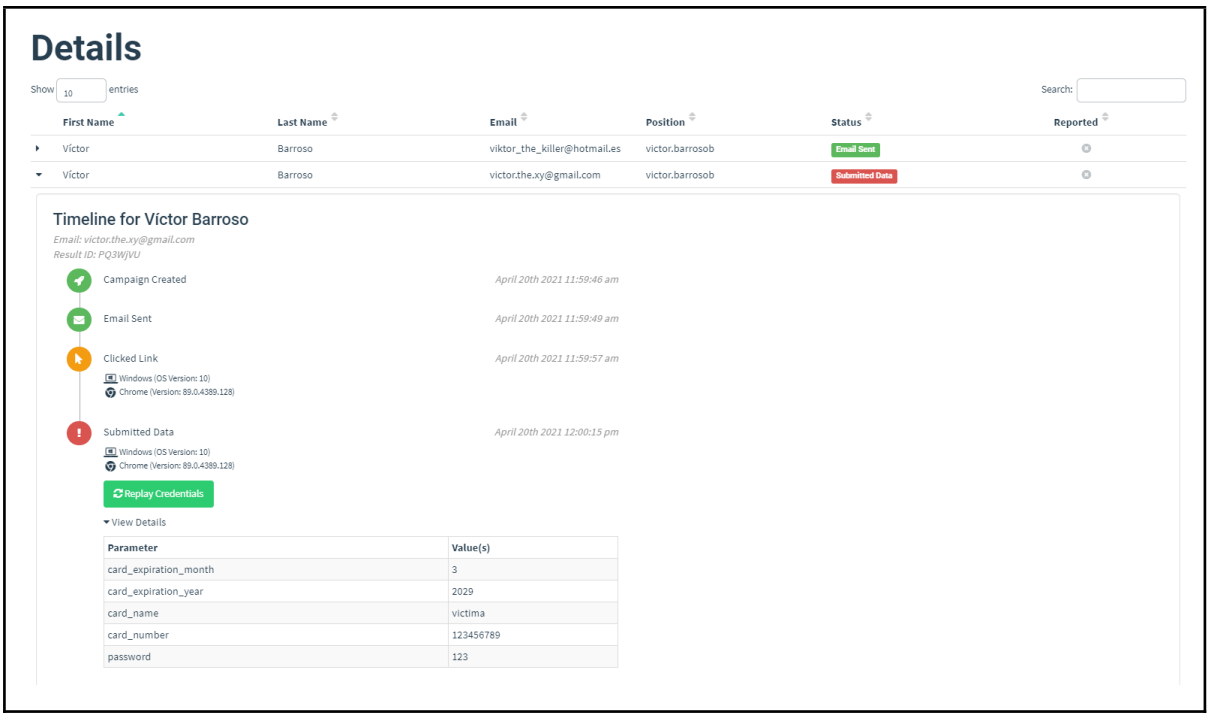

Figura 30. Detalles del resultado de una campaña en GoPhish

3.7 Mejorando el ataque

Con la configuración realizada ya tenemos montado un ataque funcional, pero aún pueden añadirse algunas modificaciones. Hay varias mejoras que se pueden realizar para hacerlo más realista:

3.7.1 De local a online

Ahora mismo nuestros servidores están funcionando en un área local. Con esto podremos hacer ciertas pruebas, pero para poder tener un entorno más realista, es necesario, como mínimo, que el servidor de Phishing esté conectado a Internet.
Hay diversas maneras de lograr esto. Por ejemplo, podríamos crear un servidor con Apache o utilizar el Internet Information Services (IIS) de Microsoft. [68][69]

En nuestro caso optaremos por crear una máquina virtual con Microsoft Azure[70], en la cual alojaremos nuestro servidor de Phishing. Se ha elegido Microsoft Azure ya que la UPC ofrece a los estudiantes 200$ de uso en sus servidores, más que suficiente para lo que necesitamos.

Aunque tengamos un margen amplio de consumo, crearemos una máquina virtual muy básica (Figura 31), ya que no tendrá una gran carga de trabajo.

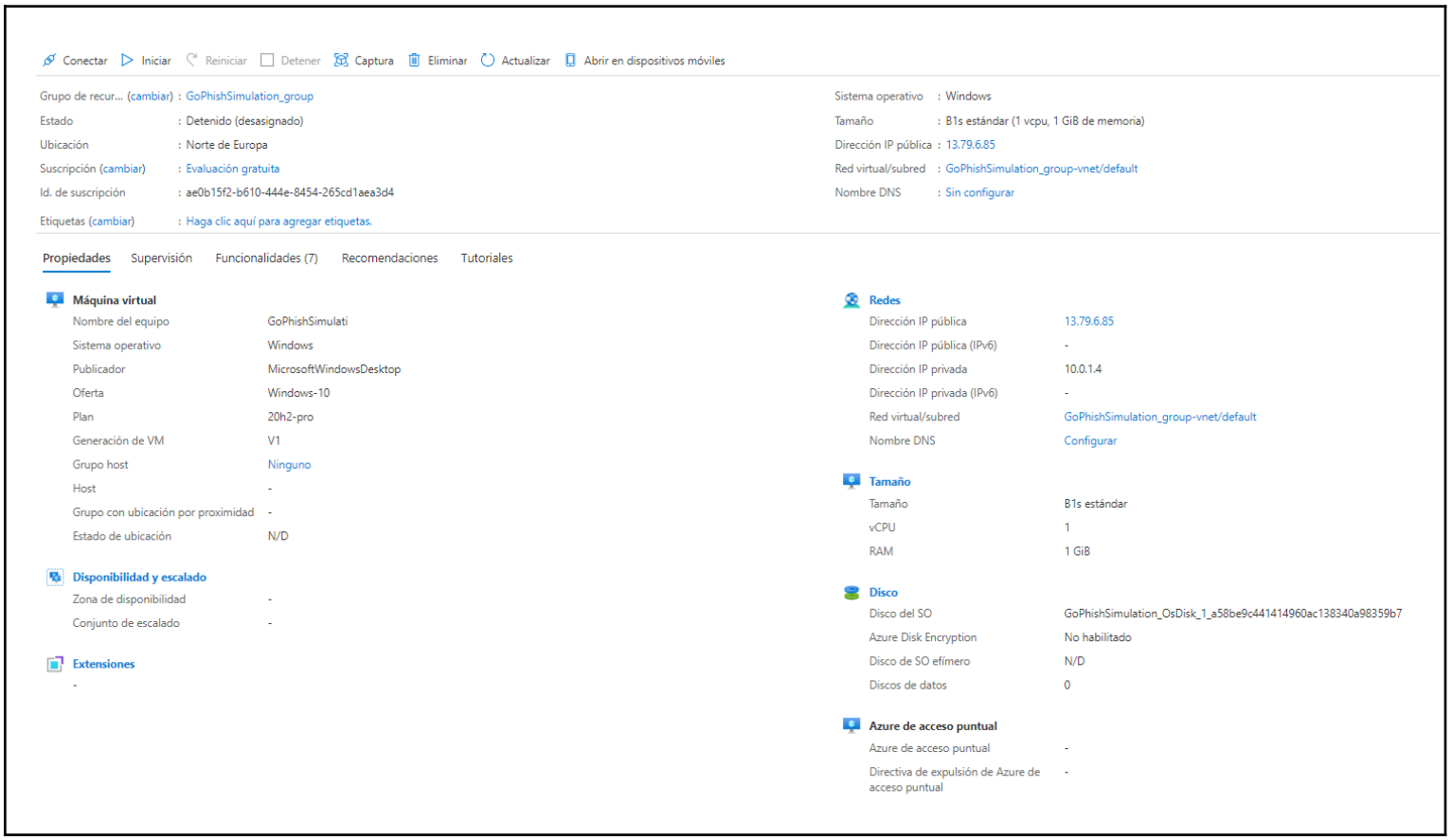

Figura 31. Panel de propiedades de máquina virtual creada con Microsoft Azure.

Para que funcione correctamente debemos configurar algunas cosas:

En primer lugar, debemos hacer que la IP pública de la que dispone nuestra máquina sea estática. Esto hará que el coste aumente un poco, pero es necesario para que podamos realizar nuestro ataque. Para ello debemos clicar en la IP pública que tenemos actualmente (en mi caso 13.79.6.85) y cambiar la asignación de dinámica a estática.

En segundo lugar, debemos dirigirnos al apartado Redes y crear unas reglas de puerto de entrada (Gophish_admin y Gophish_server en Figura 32) y unas reglas de puerto de salida (GMail_SMTP en Figura 33).

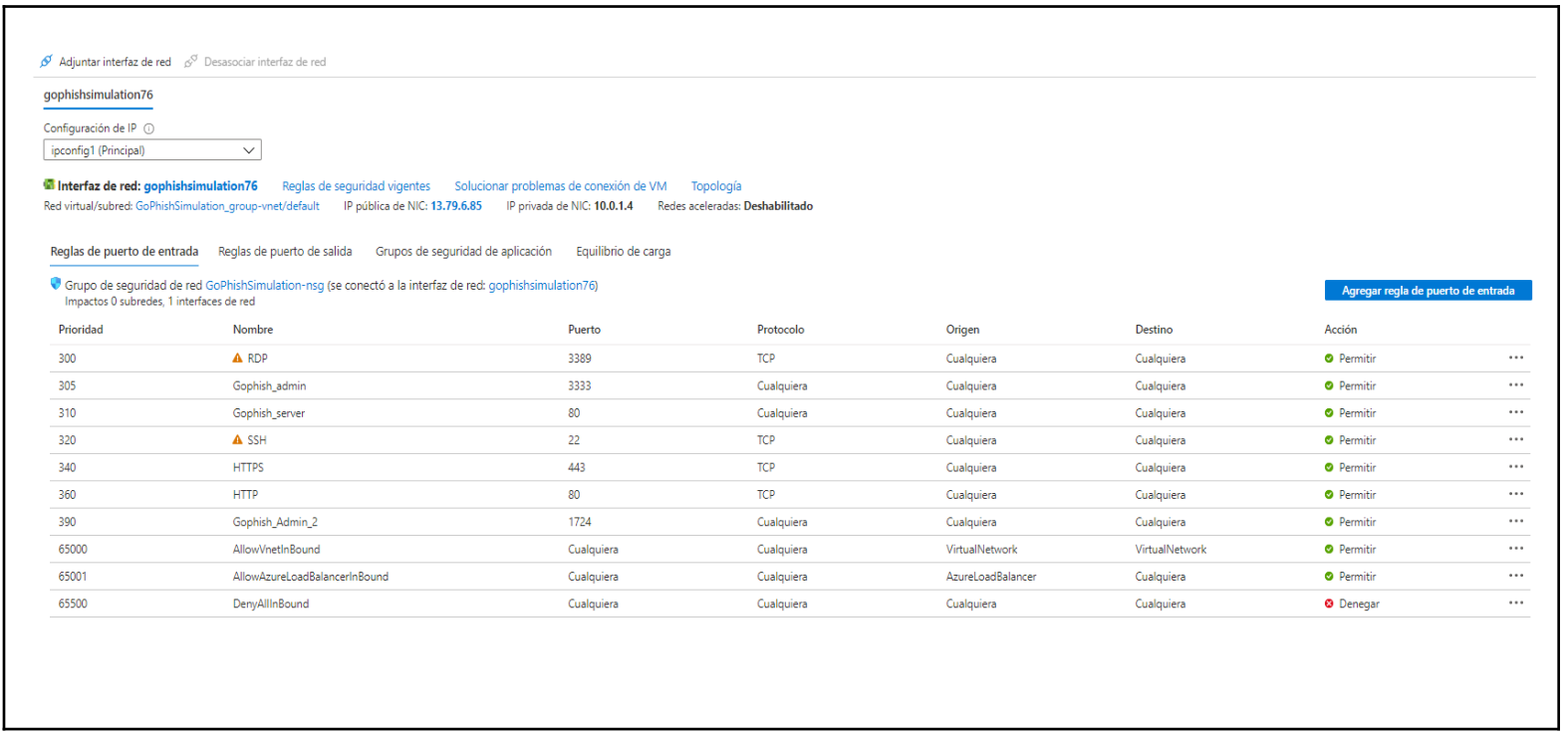

Figura 32. Reglas de puertos de entradas de la máquina virtual de Microsoft Azure

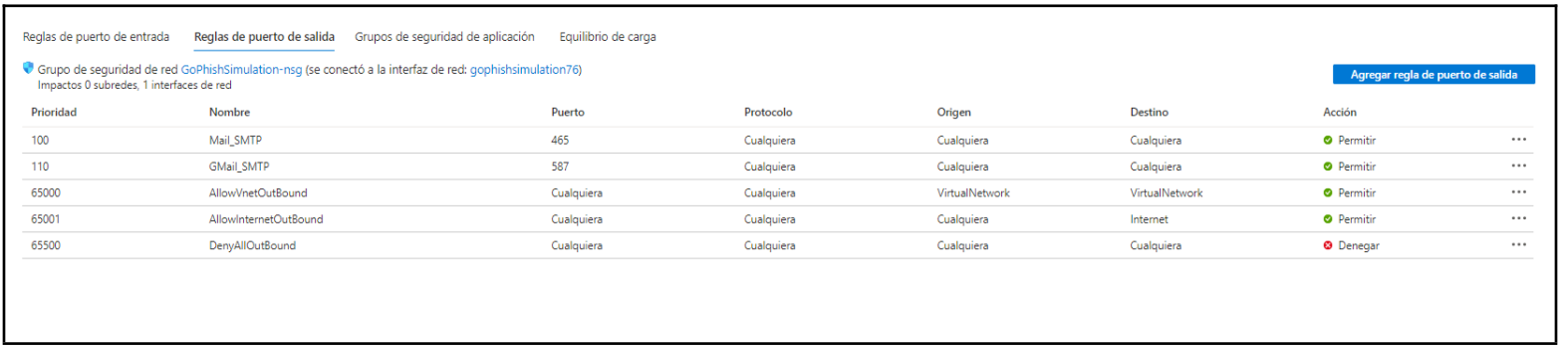

Figura 33. Reglas de puertos de salida de la máquina virtual de Microsoft Azure

Ahora tan sólo debemos acceder a la máquina como si fuera cualquier otra máquina Windows mediante Remote Desktop Protocol: Buscamos Conexión a Escritorio Remoto y como dirección ponemos la IP pública, nos pedirá la contraseña que establecimos cuando creamos la máquina virtual y ya podremos acceder.

Una vez dentro, tendremos que configurar Gophish en esa máquina, sólo que en el archivo config.json tendremos que cambiar la dirección por defecto por la de nuestra máquina virtual (en mi caso, cambiar 0.0.0.0 por 13.79.6.85).

Ahora nuestro ataque tendrá más alcance que antes, pues ya no se limita a un área local.

3.7.2 Añadir un dominio

Tras realizar los pasos anteriores se dispone de un ataque de Phishing bastante realista y que funciona de forma online. Aun así, si al acceder al enlace la víctima ve como URL una dirección IP puede generar desconfianza. Por ello es una buena idea conseguir un dominio para nuestra web.

Los dominios webs se pueden comprar en Internet por gran abanico de precios. Lo ideal sería tener un dominio .com o .es (en el caso de España) con una relación con los servicios que intentamos suplantar. Sin embargo, estos dominios pueden ser difíciles de encontrar o tener un coste más elevado.

En nuestro caso, y con el objetivo de reducir los costes lo máximo posible, utilizaremos uno de los dominios que podemos adquirir en Freenom.com [71] de forma gratuita. En mi caso los dominios que he podido adquirir son:

- loginweb-secure.tk
- loginamazon-secure.tk

La terminación .tk pertenece al archipiélago neozelandés de Tokelau.

Una vez tengamos los dominios a nuestra disposición, habrá que enlazarlos con nuestras direcciones IP.
En Freenom iremos a nuestros dominios y le daremos a “Manage Domain” (Gestionar dominio), y de aquí iremos a la pestaña “Gestionar DNS Freenom”. Aquí bastará con añadir dos entradas a nuestra tabla como se ve en la Figura 34.

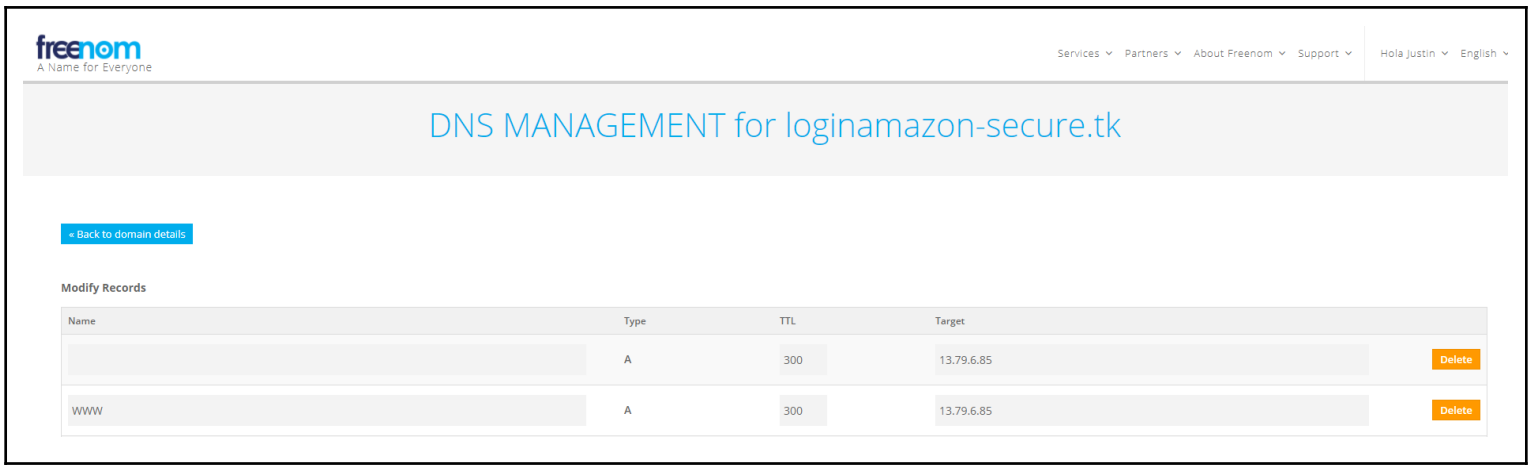

Figura 34. Configuración de DNS de un dominio en Freenom.

Una vez se actualicen los servidores de DNS, ya será equivalente acceder a 13.79.6.85 que a loginamazon-secure.tk.
Con este cambio realizado, tendremos que tener en cuenta que al añadir la URL en la configuración de la campaña en GoPhish deberemos poner el dominio adquirido y no la IP.

3.7.3 Certificado SSL

Tenemos una página web muy similar a la que suplantamos, podemos enviarla a través de Internet y hemos conseguido cambiar la URL de una IP a un dominio más creíble. Aun así, todos hemos escuchado alguna vez que no introduzcamos nuestros datos en una página que no tenga el candado de conexión segura, y nuestra página no lo tiene.

Como hemos visto al principio del trabajo, actualmente el 80% de los ataques de Phishing cuentan con un certificado de conexión segura, y vamos a hacer que el nuestro también.

Para conseguir este certificado iremos a ZeroSSL [72], aunque hay muchas otras páginas donde conseguir este tipo de certificados.

Una vez en la página introduciremos el dominio para el que queremos el certificado en el recuadro principal con el título “Create Free SSL Certificate”. Nos pedirá que nos registremos, y podremos certificar hasta 3 dominios de forma gratuita.
Debemos seleccionar ciertas características (en el caso de una cuenta gratuita no hay muchas opciones) hasta que nos aparezca el mensaje de la siguiente imagen (Figura 35):

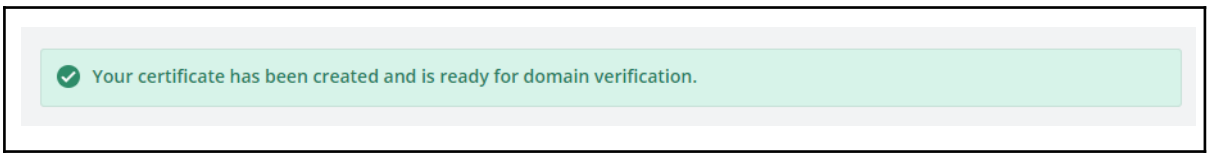

Figura 35. Mensaje de confirmación de creación de certificado SSL por ZeroSSL.com

Ya tenemos el certificado creado, pero hará falta verificarlo. Para ello seleccionamos la opción DNS (CNAME) (Figura 36) que nos indicará cómo debemos proseguir.

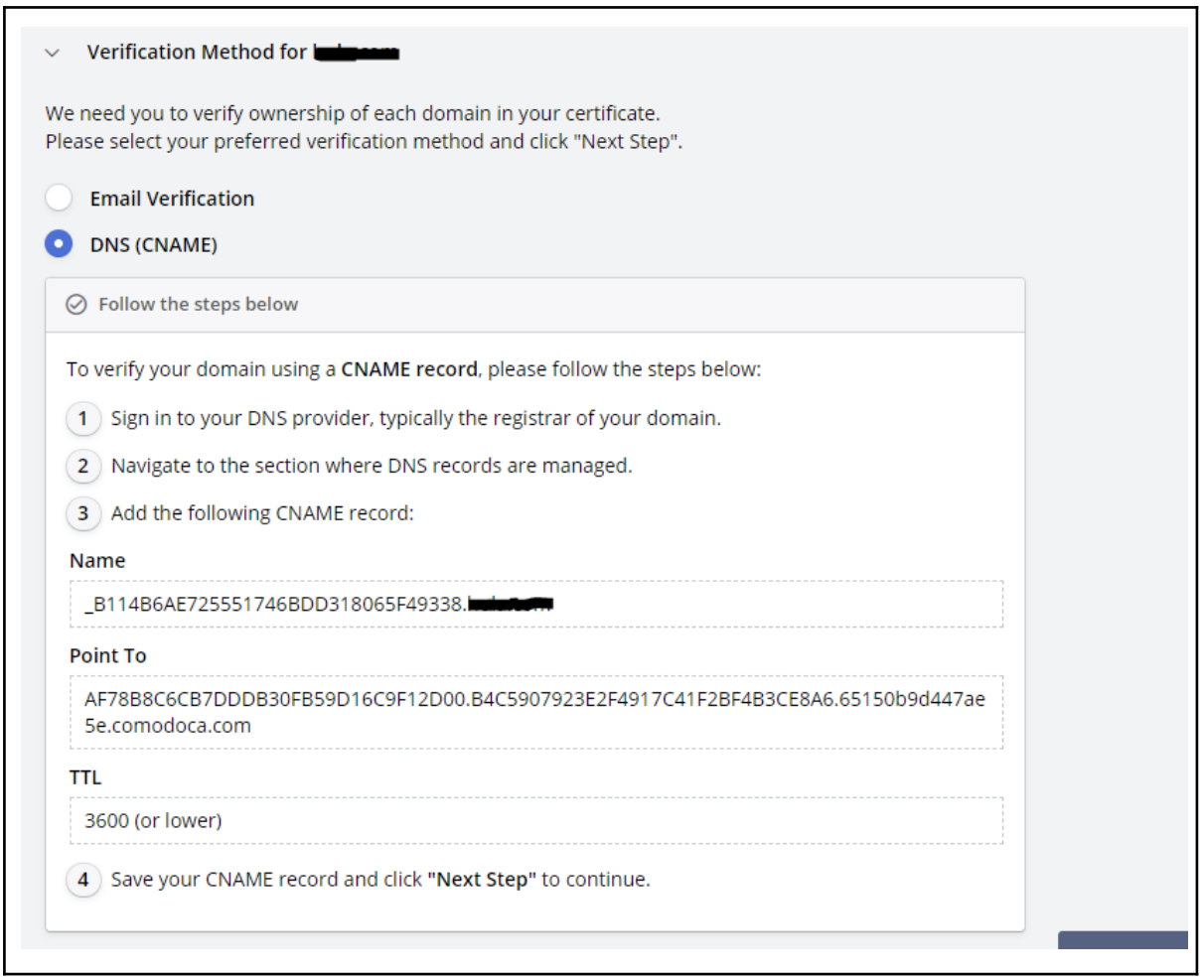

Figura 36. Pasos de verificación del certificado SSL por ZeroSSL.com

En nuestro caso iremos a Freenom.com y volveremos a la pantalla de gestión de dominios. Allí iremos a "Gestionar DNS Freenom" y añadiremos una entrada, que debe tener los valores que nos han indicado en ZeroSSL.com. Es importante que cambiemos el tipo de la entrada de "A" a "CNAME".
Cuando esté listo le damos al siguiente paso y nos pedirá que verifiquemos el dominio. Damos click al botón y esperamos a que nos verifique el certificado (es posible que haya que esperar unos 10 minutos antes de que lo verifique con éxito).

Una vez tengamos el certificado verificado nos permitirá descargar varios archivos en un .zip. Debemos descargarlos, descomprimirlos y llevarlos a nuestra carpeta de GoPhish.

Ahora necesitamos que GoPhish coja este certificado y su clave en lugar de los que elige por defecto. Para ello, abriremos el archivo config.json y realizaremos los siguientes cambios:

- Cambiaremos la listen_url de 0.0.0.0:80 a 0.0.0.0:443 (el predeterminado para HTTPS)
- Pondremos a "true" el valor de de "use_tls"
- En Cert_path pondremos el certificado que hemos descargado, el cual acaba en .crt.
- En Key_path pondremos la clave del certificado que hemos descargado, que acaba en .key.

Guardaremos este archivo y ya tendremos GoPhish configurado con HTTPS en nuestro dominio.

3.8 Problemas y limitaciones

Al trabajar buscando la simplicidad y la opción más económica pueden surgir algunos problemas y limitaciones. A continuación dispondré las más notables:

Freenom funciona mal
Si bien la página nos ofrece dominios gratuitos, conseguir registrar uno puede ser costoso en tiempo, pues la página muchas veces da errores. Algo que puede ayudar a corregir ciertos errores es hacer creer a la página que sois un usuario de EEUU. Para ello cambiad los datos de la cuenta por datos de EEUU (recomiendo usar un generador de direcciones como Fakeaddressgenerator.com[73]) y usad una VPN para cambiar vuestra dirección IP. En mi caso usé la extensión de Google Chrome de Windscribe. [74]

Dominios poco creíbles
Como ya se ha mencionado, siempre es mejor tener un dominio que una dirección IP. Sin embargo, cabe mencionar que los dominios que podremos conseguir de forma gratuita nunca serán tan creíbles como un dominio .com o .es.

Dominios certificados en GoPhish
Al añadir el certificado del dominio en GoPhishi cambiando el config.json, sólo podremos añadir un certificado a la vez. Esto significa que si queremos cambiar el dominio que utilizaremos de una campaña, deberemos cambiar el certificado que utilizará GoPhish.

SMTP de Google
Para que nuestro framework de GoPhish pueda enviar los correos, necesita un servidor de SMTP. Para realizar pequeñas pruebas podemos utilizar el que nos ofrece Google de forma gratuita. Si quisiéramos realizar campañas masivas, Google podría inhabilitarnos la cuenta al detectar que se está usando para Spam. En la documentación oficial de GoPhish se recomienda el uso de Mailhog. [75]

Filtros de Spam modernos
El principal problema de estos ataques, del cual nos daremos cuenta al intentar hacer pruebas, es que la seguridad actual en los servicios de correo electrónico, que catalogará nuestros correos como Spam. Esto hará que nuestro correo no se vea como debería (no cargará las imágenes) y desactivará los enlaces que contenga.

Más adelante se profundizará en cómo funcionan estos filtros de Spam, pero algo que podría ayudar a la hora de hacer pruebas de penetración dentro de una empresa, sería añadir la dirección IP del servidor de GoPhish a una whitelist temporalmente.

3.9 Conclusión - Ataques de Phishing

Cómo se puede apreciar a lo largo del Apartado 3 de este trabajo, cada vez resulta más sencillo poder realizar ataques de Phishing.
Actualmente, existen multitud de herramientas que los cibercriminales pueden utilizar para automatizar el proceso de creación y distribución de este tipo de ataques, y muchas de ellas, ni siquiera requieren tener ningún conocimiento informático elevado.

En este caso, hemos usado GoPhish para generar 2 ataques sencillos y comprobar cómo de legítimos llegaban a parecer. Pese a que finalmente no se pudo hacer pruebas de los ataques realizados en el público general (dado la invasión de privacidad que esto supondría), colegas y amistades cercanas me han reconocido que les costaría identificar cuál es el legítimo y además, varios de ellos caerían en el ataque.

A falta de realizar pruebas con los ataques generados, se ha difundido una encuesta creada por Jigsaw [93], usando la herramienta Quiz de Google. Los resultados de esta se estudian en el siguiente apartado.

NOTA 5: Cabe resaltar que el objetivo principal del framework GoPhish es realizar auditorías en empresas y así poder estudiar el grado de seguridad dentro de los distintos departamentos. El uso de este software con intenciones maliciosas, no será apoyado por la empresa.

4 Cuestionario sobre Phishing

Para comprobar la concienciación y la capacidad de los usuarios de distinguir el Phishing de los correos legítimos, la mejor opción (sin recurrir a prácticas que invadan su privacidad) es la realización de un cuestionario y la posterior evaluación de sus resultados.

Dada la gran calidad del test realizado por Jigsaw de Google [93], abarcando distintos tipos de ataques y personalizando los "ataques" al usuario, se ha decidido hacer uso de él para mis estudios (Figura 37).

Figura 37. Cuestionario sobre Phishing por Jigsaw, de Google [93].

El cuestionario de Google pone a prueba a los usuarios enseñándoles 8 correos personalizados los cuales deben responder si son Phishing o si son legítimos.

Antes de comenzar les pide un nombre y un correo, para que los ejercicios siguientes estén dirigidos a ellos.

A continuación, le pregunta por 8 casos distintos, en los cuales pueden interactuar con el correo, mirando a donde redirigen los enlaces, desplegando la cabecera para más información, etc.
Una vez seleccionen si un caso es legítimo o es Phishing, el cuestionario les dirá si han acertado y les indicará dónde estaba el fallo del que deberían haberse percatado.

Como no tengo acceso a los resultados de este test, se ha añadido un cuestionario previo diseñado con Formularios de Google [94] donde se añaden algunos datos para poder comprobar el número de aciertos en función de otros parámetros.

El cuestionario presenta las siguientes preguntas:

Nombre: Simplemente para identificar a las personas.
Género: A elegir entre:
Mujer
Hombre
Otro
Franja de Edad: A elegir entre:
<18
18-25
26-45
46-59
>59
Conocimiento previo del Phishing: A elegir entre:
Sí
Sí, pero no con ese nombre
Algo había oído
No conocía la existencia de esos ataques
Nivel de estudios: A elegir entre:
Obligatorios
Bachiller
Formación Profesional
Estudios Universitarios
Aciertos: A elegir entre:
0
1
2
3
4
5
6
7
8

Una vez diseñado el cuestionario, y distribuido a través de redes sociales, deberemos esperar hasta conseguir una muestra con un tamaño representativo, para así poder estudiar los resultados obtenidos.

4.1 Estudio de resultados

Tras varias semanas de distribución del cuestionario, y a fin de poder estudiarlo a tiempo, he logrado obtener 111 respuestas.
Usando una calculadora de tamaño de muestras [95] podemos ver que con nuestra cantidad de respuestas, obtenemos un 10% de margen de error y un intervalo de confianza del 96%.

Esto significa que los datos obtenidos en la encuesta en el 96% de los casos no se apartará más de un 10% del resultado real.

Las cuestiones que pretendemos investigar con nuestro test son las siguientes:

A. *¿Cuántos casos es capaz de identificar la población?*
B. *¿Qué porcentaje de la población conocía el Phishing previamente?*
C. *¿Existe una diferencia notable en los resultados según la edad?*
D. *¿Existe una diferencia notable en los resultados según el nivel de estudios?*
E. *¿Existe una diferencia notable en los resultados según el género?*

Para todos los cálculos y gráficas de los siguientes apartados he utilizado la herramienta R Studio [92], así como Hojas de Cálculo de Google [14] para generar el archivo csv que analizaremos.

4.1.A Aciertos población general

El primer punto que me gustaría tratar del estudio es el resultado obtenido en las 8 preguntas del test de Google. En los siguientes apartados se mostrará cómo se distribuyen estos resultados según los distintos atributos.

Como se puede ver en las Figuras 38 y 39, la mayoría de los aciertos se encuentran en el rango de 4 a 6 aciertos.

Figura 38. Gráfico circular de Aciertos [Elaboración propia mediante R Studio]

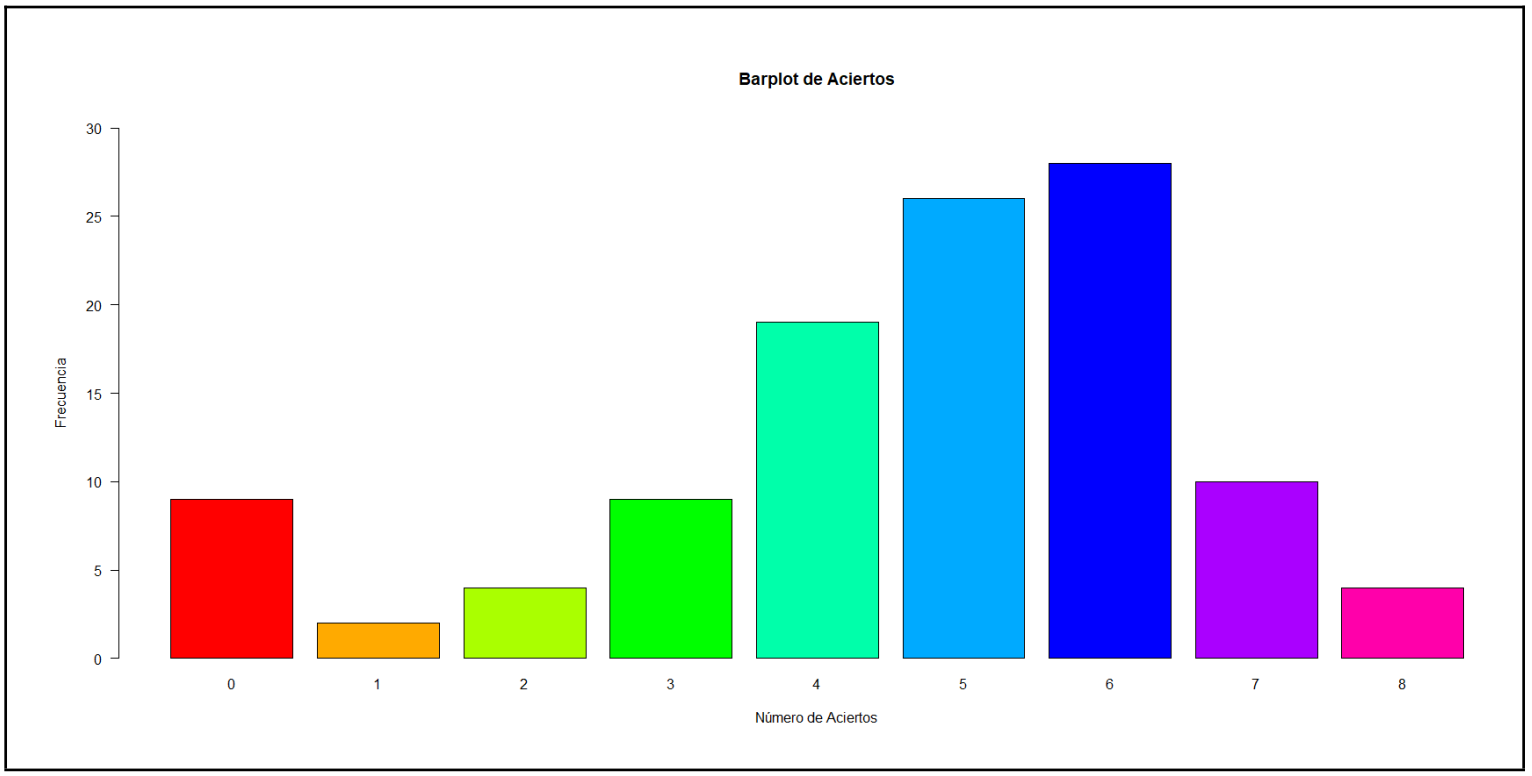

Figura 39. Barplot de Aciertos. [Elaboración propia mediante R Studio]

Si observamos a nivel numérico tenemos lo siguiente:

Nº de Aciertos	Frecuencia de Respuesta	Porcentaje de Respuestas
0 Aciertos	9	8.11%
1 Acierto	2	1.80%
2 Aciertos	4	3.60%
3 Aciertos	9	8.11%
4 Aciertos	19	17.12%
5 Aciertos	26	23.42%
6 Aciertos	28	25.23%
7 Aciertos	10	9.01%
8 Aciertos	4	3.60%

Tabla 3. Frecuencia y porcentaje de cada valor de Aciertos

A simple vista, puede parecer que los resultados son buenos, dado que la **media es de 4.62**, lo cual implica un acierto de más de la mitad. Aun así, hay que tener en cuenta que sólo el 3.6% de la gente ha sido capaz de distinguir el Phishing de los correos legítimos en las 8 ocasiones.

Además, me gustaría aclarar que no considero que estos resultados sean positivos. El 96.4% de los entrevistados caerían en este tipo de ataques, la mayoría de ellos incluso en varios. Aunque hayas identificado el ataque en 7 de las 8 ocasiones, si esa octava vez fallas, tus datos se verán comprometidos igualmente.

4.1.B Aciertos según Conocimiento Previo

Otro de los objetivos que tenía este estudio era saber cuánta gente conocía el Phishing antes de la realización de la encuesta (Figura 40).

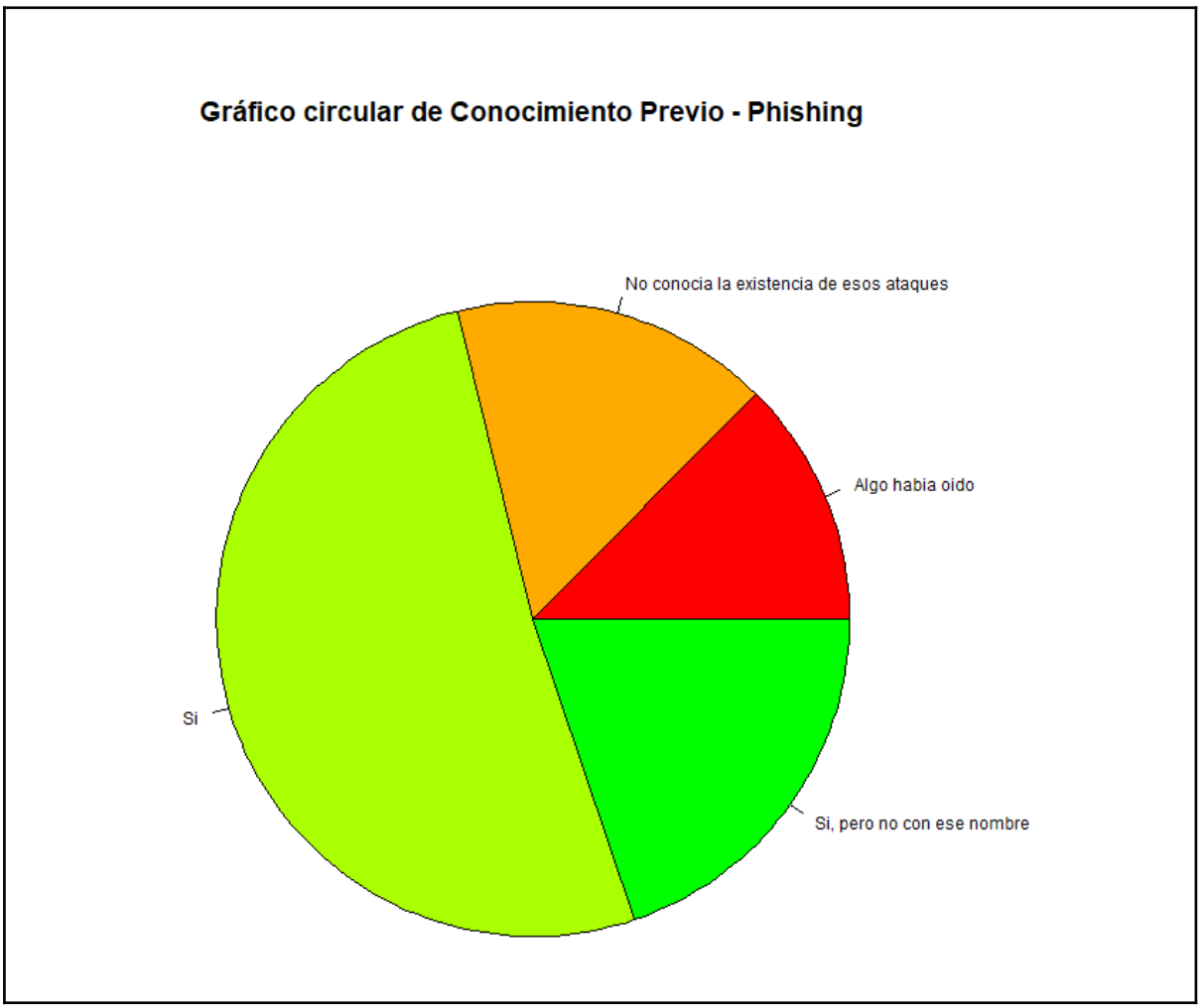

Figura 40. Gráfico circular de Conocimiento Previo del Phishing [Elaboración propia mediante R Studio]

El 51.35% de la gente conocía el Phishing, y el 19.82% lo conocía pero no sabía cómo se llamaba.
Por otro lado, tenemos que el 16.22% no sabía nada de su existencia y el 12.61% había oído algo al respecto.

Pese a que más de la mitad de los entrevistados conoce estos ataques, es necesario que se siga dando a conocer a los usuarios, pues como podemos ver en la siguiente tabla (Tabla 4), a mayor conocimiento tenga la gente de este tipo de ataques, mayor es su capacidad de aciertos.

Grado de Conocimiento Previo	Media de respuestas acertadas
No conocía la existencia de esos ataques	3.11
Algo había oído	4.21
Sí, pero no con ese nombre	4.45
Sí	5.26

Tabla 4. Media de Aciertos para cada valor de Conocimiento Previo

Como se aprecia, la diferencia entre no conocer el Phishing y conocerlo, puede suponer una media de 2 aciertos más por cada 8 correos.

4.1.C Aciertos según Edad

En el estudio de la edad (Figura 41), podemos observar cómo la frecuencia de cada uno de los rangos de edad está muy influenciada por los círculos en los que, personalmente, me puedo mover con más facilidad, y en los que distribuir una encuesta de este tipo resulta más sencillo.

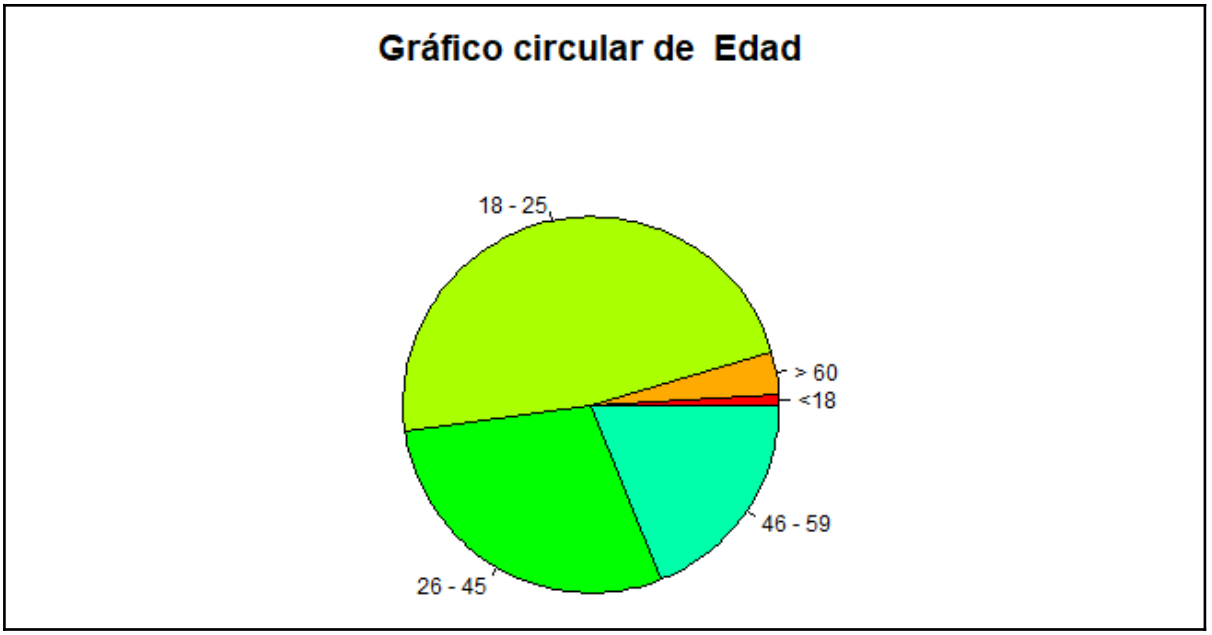

Figura 41. Gráfico circular de Edad [Elaboración propia mediante R Studio]

En los grupos más extremos, según la edad (<18/>60), he conseguido muy pocos resultados, por lo que los datos obtenidos respecto a ellos no son fiables. Por ese motivo, no consideraré dichas respuestas en los cálculos posteriores.

A continuación podemos ver los datos relativos a los tres grupos intermedios (Tabla 5).

Rango de Edad	Media de Aciertos
18-25	4.75
26-45	4.25
46-59	4.86

Tabla 5. Media de Aciertos para cada valor de Edad

Como se puede ver, las medias por edades no distan tanto, con una ligera diferencia en el rango intermedio.

4.1.D Aciertos según Nivel de estudios

Si comprobamos el nivel de estudios en relación a los aciertos obtenidos, sí que podemos ver una correlación (Figura 42), como pasaba con el grado de conocimiento previo.

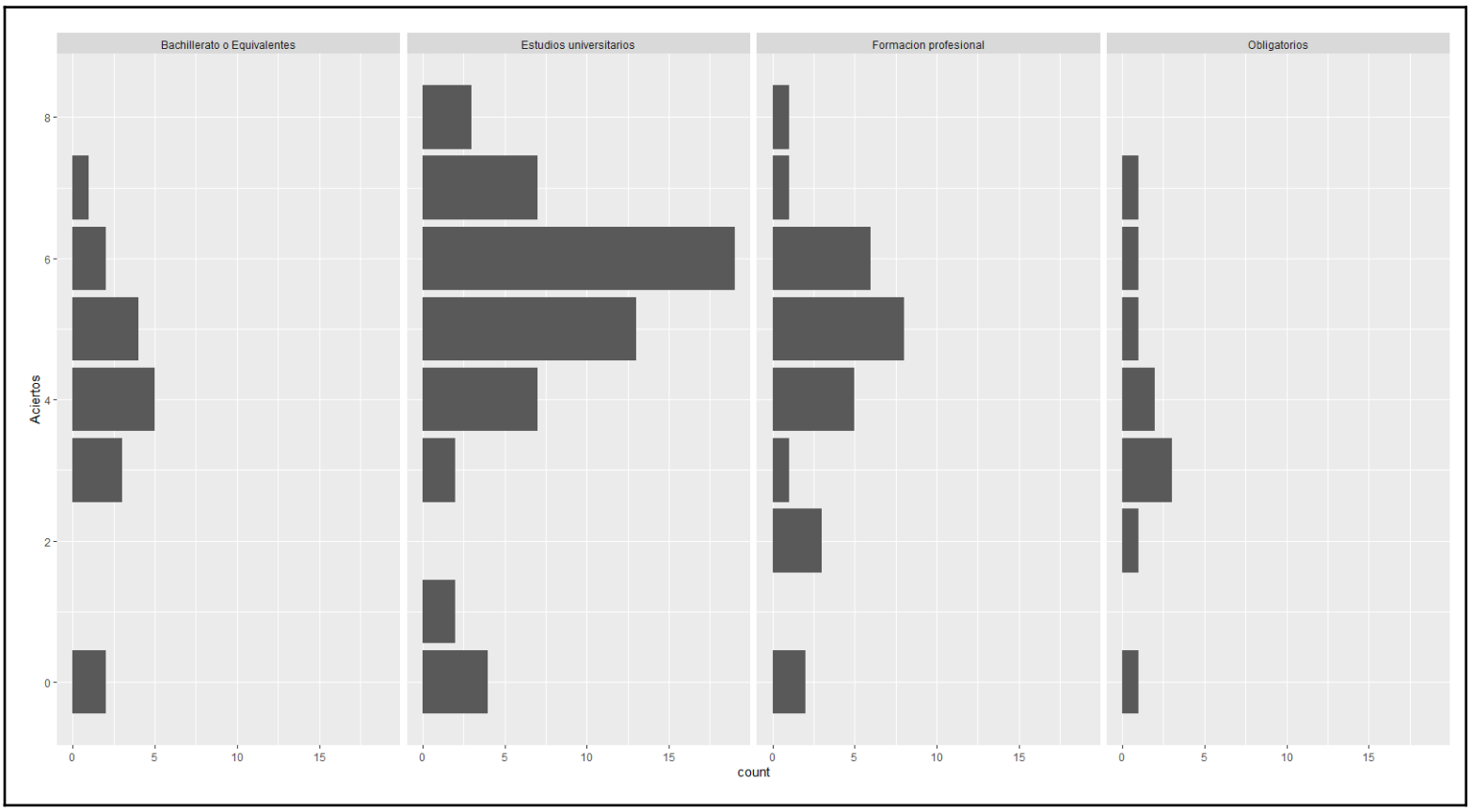

Figura 42. Barplot de Nivel de Estudios frente a los Aciertos [Elaboración propia mediante R Studio]

A simple vista, ya se puede apreciar como los picos de las curvas corresponden al nivel de estudios. Es decir, obtendrán mayor número de aciertos cuanto mayor sea el grado de estudios del que disponga el usuario.

Podemos hacer lo mismo que con otras variables y calcular numéricamente. Es tan sencillo como calcular la media de cada grupo (Tabla 6).

Nivel de Estudios	Media de Aciertos
Obligatorios	3.7
Bachillerato o Equivalentes	4
Formación Profesional	4.44
Estudios Universitarios	5.05

Tabla 6. Media de Aciertos para cada valor de Nivel de Estudios

4.1.E Aciertos según Género

Por último, quería comprobar cómo afecta en los resultados el género del usuario (Figura 43).

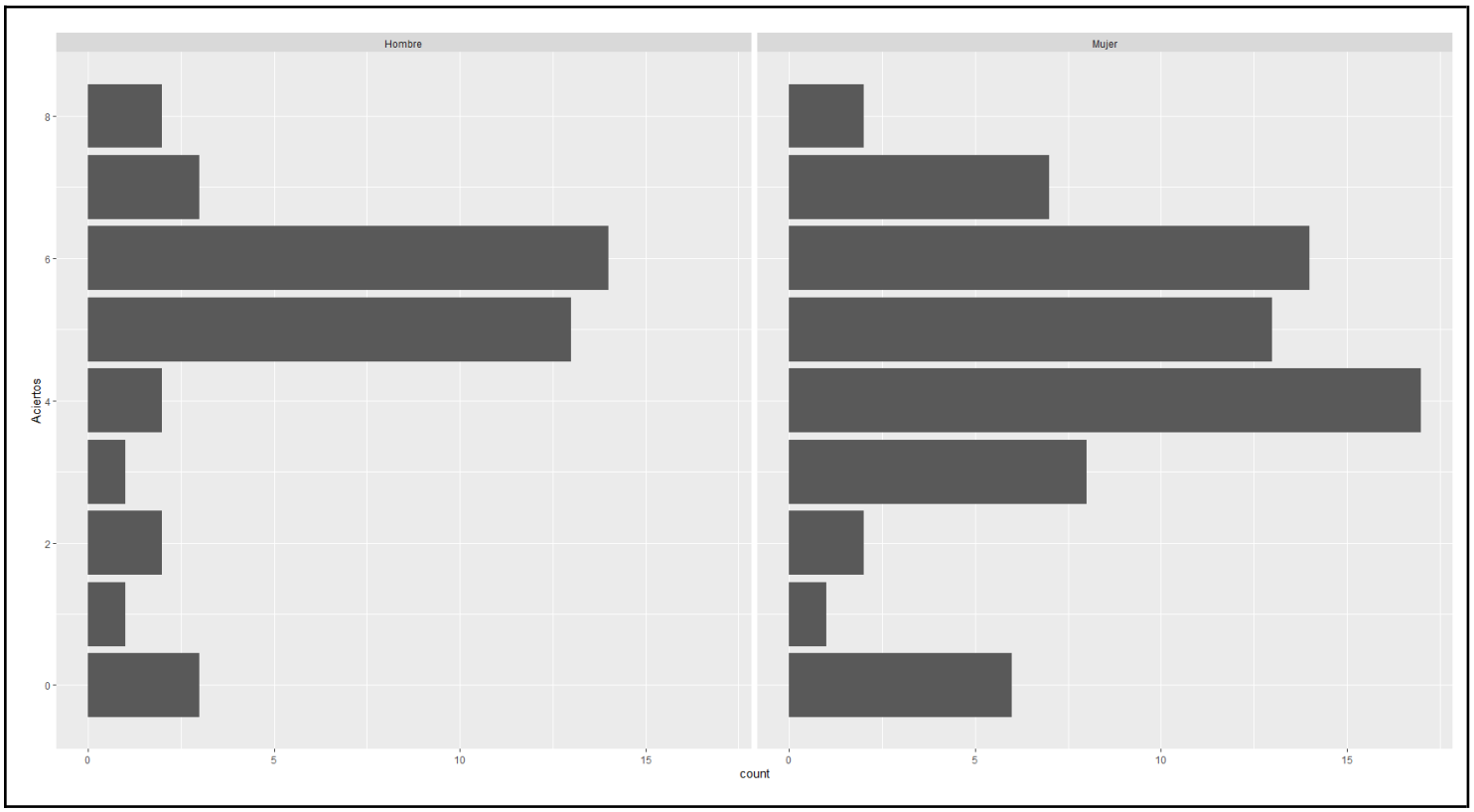

Figura 43. Barplot de Género frente a los Aciertos [Elaboración propia mediante R Studio]

Antes de comenzar, cabe detallar que a la encuesta respondieron casi el doble de mujeres que de hombres (70 frente a 41). La opción de género Otro no obtuvo ninguna respuesta.

Curiosamente, pese a que los resultados son similares en ambos casos, encontramos que el caso más frecuente de las mujeres es de 4 Aciertos. Mientras, que si nos fijamos en el caso de los hombres, apenas nadie ha obtenido ese resultado.

Esto provoca que a la hora de calcular la media obtengamos una diferencia (Tabla 7).

Género	**Media de Aciertos**
Mujeres	4.44
Hombres	4.93

Tabla 7. Media de Aciertos para cada valor de Género

Dado que no veía ningún motivo para que existiera esta diferencia basada en el género, decidí comparar la variable género con las 2 variables que han mostrado una correlación clara con el número de aciertos, el Conocimiento Previo y el Nivel de Estudios.

Estudio de Conocimiento Previo según Género

Mujeres		
Valor	Conocimiento Previo	Porcentaje
1	No conocía la existencia de esos ataques	21.43%
2	Algo había oído	11.43%
3	Sí, pero no con ese nombre	20%
4	Sí	47.14%
	MEDIA	**2.9285**

Tabla 8. Cálculo de la media de Conocimiento Previo en Mujeres

Hombres		
Valor	Conocimiento Previo	Porcentaje
1	No conocía la existencia de esos ataques	7.32%%
2	Algo había oído	14.63%
3	Sí, pero no con ese nombre	19.51%
4	Sí	58.54%
	MEDIA	**3.2927**

Tabla 9. Cálculo de la media de Conocimiento Previo en Hombres

–

Estudio de Nivel de Estudios según Género

Mujeres		
Valor	Nivel de Estudios	Porcentaje
1	Obligatorios	10%
2	Bachillerato o Equivalentes	15.71%
3	Formación Profesional	24.29%
4	Estudios universitarios	50%
	MEDIA	**3.14**

Tabla 10. Cálculo de la media de Nivel de Estudios en Mujeres

–

Hombres		
Valor	Nivel de Estudios	Porcentaje
1	Obligatorios	7.32%%
2	Bachillerato o Equivalentes	14.63%
3	Formación Profesional	24.39%
4	Estudios universitarios	53.66%
	MEDIA	**3.24**

Tabla 11. Cálculo de la media de Nivel de Estudios en Hombres

–

Como se puede apreciar en las tablas (Tablas 8, 9, 10 y 11), existe una diferencia en la media del nivel de Estudios y del Conocimiento previo de las respuestas recogidas de ambos géneros.
Probablemente este es el motivo por el que existe esa diferencia en la media de ambos géneros.

4.2 Conclusión - Cuestionario sobre Phishing

Alguien podría pensar que dado que la media de aciertos es un 4.26, siendo la nota máxima es un 8, la población está aprobada con una nota justa. Nada más lejos de la realidad. Esa nota significa que por cada 8 correos fraudulentos que una persona media recibe, cae en 4 de ellos.
Se calcula que actualmente se envían alrededor de 3.000 millones de correos fraudulentos al día, de los cuales 1.500 millones tendrían éxito si dependiera sólo de los usuarios.

Si bien en nuestro estudio no podemos concluir que la edad sea un factor relevante a la hora de identificar estos ataques. Lo que sí podemos asegurar, con una alta probabilidad de no equivocarnos, es que la gente que ya había oído sobre estos ataques y aquella con un nivel de estudios más elevado, tienen una tasa de acierto mayor.
Por este motivo, es de vital importancia que los usuarios reciban formación y advertencias sobre los posibles peligros de Internet, que sepan cómo los criminales tratan de robar sus datos y que se les enseñe que pueden hacer para tratar de mantenerse seguros.

Como veremos en detalle más adelante, los 1.500 millones de correos fraudulentos con éxito que mencionaba anteriormente, no son reales. No son solo los usuarios los que deben actuar para evitar estos correos, las empresas tienen un papel fundamental.

Cabe resaltar la gran utilidad del Quiz diseñado por Google. El entorno que crean genera unos emails muy convincentes, donde puedes poner a prueba tus capacidades ante estos ataques. Además, el propio ejercicio te muestra correo por correo donde están las características que lo delatan, sirviendo así a su vez para que en próximas ocasiones los participantes estén más preparados y puedan evitar estos ataques.

5. Filtros de Spam

En este apartado del proyecto analizaremos cómo funciona un filtro de Spam cómo el utilizado por Gmail [76], Outlook [77] o cualquier otro servicio de correo electrónico.

Como se ha visto a la hora de realizar la simulación del punto 3, muchos de los consejos que se dan actualmente (buscar faltas de ortografía, comprobar que tiene el candado de conexión segura, etc), pese a ser útiles, no son suficientes.

En la actualidad, se calcula que el 85% del tráfico de correo diario es Spam [78]. El tiempo que debe emplear una persona para separar los correos deseados de los no solicitados, junto con los problemas de seguridad que puede suponer que el usuario vea todos los correos sin una criba previa, hacen de vital importancia la figura de los filtros de Spam.

Pero, ¿Qué es exactamente un filtro de Spam?[79] Como su nombre indica, un filtro de Spam es una tecnología capaz de diferenciar cuales de los correos recibidos son Spam, y separar estos de los correos solicitados. Este filtro puede presentarse de diversas maneras: puede estar implementado por el propio proveedor de correo, por una empresa o incluso por el propio usuario. Así mismo, la forma en la que se implementa también varía, pudiendo ser dispositivos configurables, softwares, algoritmos o, hacía donde parece apuntar el futuro de esta tecnología, usando machine learning.

5.1 Tipos de Filtros de Spam

No podemos considerar todos los filtros de Spam iguales, pese a que el objetivo final sea el mismo. En los siguientes subapartados se mencionarán las principales categorías en las que podemos dividirlos[79], primero teniendo en cuenta su localización y luego su funcionamiento.

5.1.1 Según su localización

Dependiendo de en qué parte del proceso de envío de correos encontremos el filtro, podemos separarlos en los siguientes tipos, entre otros:

- **Filtro de Spam Gateway**

  Los filtros de Spam de tipo Gateway se encuentran detrás del Firewall de la red. Generalmente, este tipo de filtros se encuentran instalados on premise (en local) en los servidores, y no en la nube. La idea de este filtro es que sea la "puerta de seguridad" de la empresa, haciendo que sólo entren en la red de la compañía aquellos correos que no se consideren Spam.

  Un ejemplo conocido de este tipo de filtro es el Barracuda Email Security Gateway.

- **Filtro de Spam Hosted**

  Los filtros de Spam de tipo Hosted se pueden encontrar antes de que el correo acceda a la red (como los Gateway) o una vez ya dentro de ella. En este caso, el filtro se encuentra en la nube, permitiendo que pueda ser actualizado de forma sencilla teniendo las versiones más recientes. Estos filtros suelen pertenecer a terceros y cuentan con servicios de suscripción. SpamTitan es un ejemplo de este método (aunque también ofrecen un filtro gateway).

- **Filtro de Spam Desktop**

  Los filtros de Spam de tipo Desktop son instalados por el propio usuario en su ordenador. Su principal beneficio es que permite configurar el filtro de manera que mejor se adapte a sus propias necesidades.

5.1.2 Según su funcionamiento

Por otro lado, podemos separar los filtros según que es lo que se observa para realizar la criba. Algunos de ellos son:

- **Filtro de Contenido**

  Un filtro de contenido comprobará la cabecera y el cuerpo del mensaje para tomar la decisión.

  Dentro de la cabecera mirará entre otras cosas:

  - La dirección de correo de origen
  - La dirección de correo de destino
  - Las paradas que el correo ha hecho en distintos servidores

  El filtro utilizará la cabecera para comprobar si el origen del correo está en alguna blacklist o si ha parado en algún servidor sospechoso.

  Además de la cabecera, el filtro comprobará el cuerpo del mensaje, buscando palabras o imágenes comunes en el Spam. Este tipo de filtros son muy sensibles a ciertas palabras, pudiendo ocasionar falsos positivos de Spam.

- **Filtro Basado en Reglas**

  Un filtro basado en reglas determinará si el mensaje es Spam o no comprobando si cumple alguna de las reglas creadas previamente. Este tipo de reglas pueden ser: bloquear los correos de una dirección o aquellos que contengan unas ciertas palabras.

- **Filtro Bayesiano**

  Un filtro bayesiano aprenderá las preferencias de Spam del usuario. Cuando el usuario marque un determinado correo como Spam, el filtro analizará sus características y buscará correos similares. Lo mismo sucederá si le decimos que un correo marcado como Spam no es realmente Spam. Este tipo de filtros inteligente es capaz de adaptarse al usuario.

A pesar de que hay multitud de tipos de filtros distintos, no hay un filtro predilecto.
A día de hoy, lo normal es que las empresas combinen distintos tipos de filtros para mayor seguridad y adaptabilidad.

5.2 Filtro de Spam de Gmail

Según los cálculos de Litmus Email Analytics[80], Gmail es el segundo cliente de correo electrónico más usado con un 27.2%, solo siendo superado por Apple iPhone (38.9%). Dado su importancia, tanto en ordenador como en móvil, será el filtro que tomaremos de ejemplo para ver qué tecnologías se están llevando a cabo en la actualidad.

5.2.1 Comienzos de Gmail

Gmail fue lanzado el 1 de abril de 2004 y rápidamente pasó a ser uno de los clientes de correo más utilizados en el mundo.

Por aquel entonces, los clientes de correo electrónico ya eran capaces de distinguir el Spam leyendo el contenido del mensaje. Ante esta situación surge el Image Spam, el cual sustituye el texto de los correos por una simple imagen que contenga dicho texto, imposibilitando que los clientes de correo pudieran leer el contenido.
Gmail disponía de tecnologías avanzadas como OCR (Optical Character Recognition) que permitía leer el contenido de estas imágenes, y filtrar así los mensajes.
Desde el comienzo, Gmail ha basado parte de su filtrado en la respuesta de los usuarios, haciendo que si suficientes personas marcaban un correo como Spam, pasará a ser considerado como tal en otros usuarios.

Actualmente, el sistema basado en reglas que tenían, ha evolucionado a clasificadores lineales de Machine Learning, algoritmos de Deep Learning (Aprendizaje profundo), etc. [81]

5.2.2 Actualidad de Gmail

El sistema actual de filtrado de Gmail es capaz de detectar Spam con un 99.9% de precisión. Este sistema se basa principalmente en 4 tecnologías: Clasificadores lineales de *Machine Learning, Reputation, Deep Learning* y *Tensor Flow.* [82]
La importancia de estas tecnologías se puede apreciar en la Figura 44.

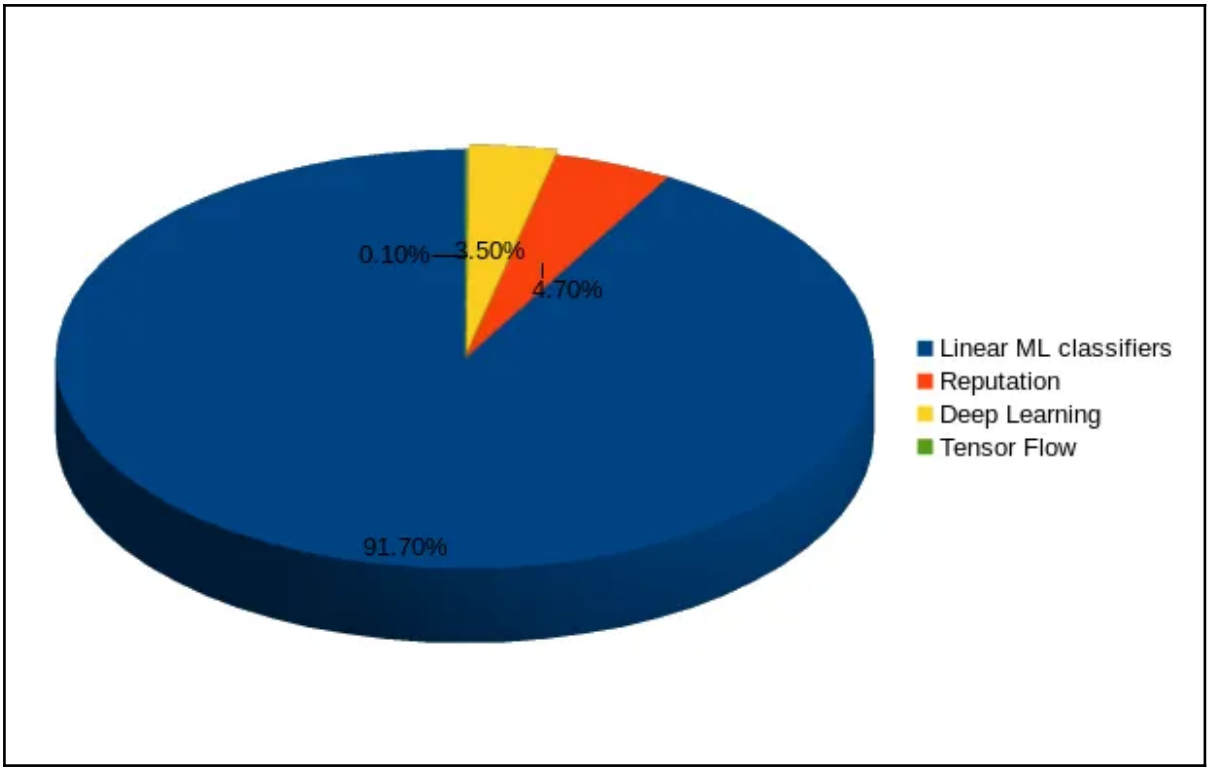

Figura 44. Porcentaje de utilización de las diversas tecnologías en el diseño de un filtro de Spam [82].

Clasificadores Lineales de Machine Learning

Un clasificador nos ayuda a separar datos en distintas clases basándose en las características que compartan. En nuestro caso, los datos serán obtenidos de los correos electrónicos y el objetivo será separar los correos legítimos del Spam. [84]

Gracias a la IA actual, es posible identificar los conceptos del Spam y conseguir así Extrapolación Temporal.[83] Esta extrapolación permite conocer la tendencia del Spam y, mediante los datos del pasado, conocer como evolucionará en el futuro.
Según Google, Gmail es capaz de predecir el 3.5% de todo el Phishing avanzado y del Spam gracias a la predicción de tendencias.

Estos clasificadores necesitan de un dataset lo más amplio y actualizado posible, para poder seguir mejorando el filtro con los nuevos ataques que puedan surgir.

Cómo determinar "qué es Spam y qué no" es relativo (puesto que para alguien puede ser Spam un mensaje que yo considero de interés) las IA trabajan en modelos propios para cada usuario, basándose por ejemplo en la tolerancia del usuario frente al Spam.

Reputation
Desde la aparición de la herramienta Google Postmaster, la reputación de dominio ha tomado mucha importancia.

Con esta herramienta se pretende ayudar a las empresas a analizar las campañas de emails masivos que realizan, descubriendo los problemas que surjan y aprendiendo mejores prácticas de email marketing para que sus correos no se cataloguen como Spam.

Por su popularización, la reputación que tu dominio tenga, ha adquirido gran importancia en los últimos años. Así, si tu reputación es alta, indicará que el patrón de tus correos es correcto y estás enviando contenido relevante a tus usuarios, los cuales realmente quieren recibirlos. Por otro lado, si tu reputación es baja, significa que deberás mejorar tu sistema de envío. Estos correos de dominios con baja reputación serán catalogados como Spam.

Deep Learning
Deep Learning es un subcampo del Machine Learning interesado en algoritmos inspirados en la estructura y funcionamiento del cerebro, llamados Artificial Neural Networks (Red Neuronal Artificial). [85]

Google utiliza grandes cantidades de datos junto con potentes supercomputadoras para entrenar estas redes para "pensar" y tomar decisiones inteligentes. Este tipo de redes se utiliza tanto en Gmail como en otros de sus servicios.

Gmail utiliza esta red para analizar patrones de Spam, Phishing, spoofing u otros tipos de correos fraudulentos y, poder así, predecir los próximos casos. En el caso de detectar posible Spam, Gmail optará por limitar la velocidad a la que envía dichos mensajes o, simplemente, bloquear la entrada de los correos a sus servidores.

Tensor Flow
Otro de los mecanismos utilizados por Google, para hacer que su sistema de correo electrónico esté libre de Spam, es TensorFlow. [86] Esta herramienta de Machine Learning es un framework de código abierto desarrollado en Google. En 2019, Gmail consiguió bloquear 100 millones de mensajes de Spam adicionales al día gracias a TensorFlow.

Esta herramienta ha ayudado a bloquear mensajes basado en imágenes, correos con contenido incrustado oculto y mensajes de dominios recientes que tratan de ocultar pequeños volúmenes de Spam dentro de tráfico legítimo.
TensorFlow permite trabajar con Machine Learning de una forma más sencilla, usando herramientas que hacen el proceso más eficiente y acelerando la velocidad a la que se puede iterar. Este framework también da la posibilidad de entrenar usando diferentes modelos en paralelo para buscar la solución más efectiva.

El carácter de código abierto ha generado una comunidad que para 2019 ya había desarrollado más de 71.000 modificaciones del código, haciendo que surjan nuevas ideas que pueden ser aplicadas rápidamente.

5.3 Cibercriminales y filtros de Spam

Como hemos visto, los servicios de correos tratan de reducir el número de mensajes de Spam que llegan a tu bandeja de entrada. Sin embargo, los pocos que logren pasar estos filtros pueden causar mucho daño a las empresas.
A continuación, expondré algunos de los métodos que los cibercriminales utilizan para tratar de evitar estos filtros [87]:

5.3.1 Links maliciosos en documentos adjuntos

A día de hoy, es prácticamente inviable que una empresa no comparta mediante servicios de correo documentos en formato de Office (.docx, .xlsx, .pptx, etc). Estos documentos consisten en diversos archivos XML que describen su contenido y su formato.

La mayoría de herramientas de escaneo comprueban el archivo xml.rels (archivo de metadatos) para buscar enlaces externos sospechosos. Los cibercriminales pueden llegar a introducir estos enlaces en el documento, pero a su vez, conseguir eliminarlos del archivo xml.rels. De esta forma, podría pasar ciertos filtros.

5.3.2 Lenguaje o Links confusos

Según Bleeping Computer, una de las técnicas que se están empleando a día de hoy para tratar de saltar estos filtros es el uso de lenguaje confuso. [88]

Algunas de las palabras más usadas en los correos de Spam son “sex” y “free cash”. Los filtros actuales pueden con facilidad analizar las palabras del correo y determinar si es Spam o no. El problema viene cuando los atacantes utilizan lenguaje muy poco común o en otros lenguajes, provocando que sea algo más confuso para la víctima, pero también para el filtro.

5.3.3 Texto oculto

Otra de las técnicas que se han reportado en los últimos años es el uso de texto oculto. [89][90]

Como ya hemos comentado, un filtro es capaz de clasificar el correo según el contenido del mismo. Pero, ¿y si lo que ve la víctima y lo que ve el filtro son cosas distintas? Mediante el uso de la complejidad de HTML los atacantes añaden texto invisible entre ciertos caracteres. De esta forma, donde el usuario lee “change your password.”, el filtro leería el texto completo, como por ejemplo “c-h-a-n-g-e- -y-o-u-r- -p-a-s-s-w-o-r-d-.”, lo cual no detectaría como sospechoso.

Este tipo de técnica se usó en 2020 para enviar Spam a través de Office 365.

5.3.4 HTML <font>

Junto con el ataque de texto oculto, también se usó este método para introducir Spam en Office 365.

Se basa en el hecho de utilizar una fuente diseñada para recrear un logo en concreto. De esta forma, donde el filtro está leyendo un texto cualquiera, la fuente está mostrando al usuario el logo de una empresa.

Figura 45. Correo fraudulento mostrado como lo ve el usuario. [91]

Figura 46. Correo fraudulento mostrando el contenido real del mensaje. *[91]*

En las Figuras 45 y 46 se pueden apreciar varias de estas técnicas.[91]

Se puede ver lo mencionado en el apartado 4.3.3, fijándonos en como en el texto real hay "j" entre cada carácter, para que el filtro no identifique la frase real.
Por otro lado, tenemos una variante del apartado 4.3.4 la cual ha utilizado tablas de HTML para generar un logo.
Finalmente, podemos ver como ha añadido otro texto oculto al comienzo del mensaje. Este texto son enlaces a páginas de alta confianza, como Google.com o Amazon.com. De esta forma el atacante espera ser valorado más positivamente por parte del filtro.

5.4 Implementando filtro de Spam con Machine Learning

De cara a demostrar la alta eficiencia, de los filtros de Spam, me dispongo a realizar uno usando Machine Learning. Si bien esta elaboración será sencilla, comparada con las que pueden estar utilizando ahora los servicios de correo electrónico, espero poder obtener resultados satisfactorios.

El objetivo del modelo es que sea capaz de distinguir un correo legítimo de uno de Spam, y para ello lo que hará será analizar qué palabras son más frecuentes en cada categoría, y poder así, a posteriori, distinguir los nuevos correos. [96]

5.4.1 Búsqueda de datos

Lo primero que haremos será encontrar un conjunto de correos ya catalogados previamente como Spam o legítimos, para así probar el modelo que desarrollaremos.

Tras buscar en distintas páginas de distribución de Datasets (Conjuntos de Datos) como puede ser Kaggle [97], finalmente he optado por usar los datos que utiliza Chirag Jain en su repositorio de Github [98] para la realización de un script de un carácter similar.

5.4.2 Preparación de datos

Una vez elegido los datos, comenzaremos a trabajar con ellos utilizando Jupyter Notebook [99]. Desde este programa podremos trabajar cómodamente en el formato IPYNB (IPython Notebook), en el que usaremos Python de una manera interactiva.

Una vez leamos los datos, podremos ver cómo están dispuestos. (Figura 47).

	text	spam
0	Subject: naturally irresistible your corporate...	1
1	Subject: the stock trading gunslinger fanny i...	1
2	Subject: unbelievable new homes made easy im ...	1
3	Subject: 4 color printing special request add...	1
4	Subject: do not have money , get software cds ...	1
...	...	...
5723	Subject: re : research and development charges...	0
5724	Subject: re : receipts from visit jim , than...	0
5725	Subject: re : enron case study update wow ! a...	0
5726	Subject: re : interest david , please , call...	0
5727	Subject: news : aurora 5 . 2 update aurora ve...	0

Figura 47. Dataset de correos catalogados.

Como se puede apreciar, disponemos de más de 5.000 correos, identificados como Spam (atributo Spam igual a 1), o como correos legítimos (atributo Spam igual a 0). El atributo text contiene el texto del correo.

A continuación, deberemos comprobar si hay duplicados (y eliminarlos) o si falta algún valor (y decidir que hacer al respecto).
El siguiente paso es crear una función a la que le pasaremos un texto y realizará dos acciones:

1. Eliminará los signos de puntuación que encuentre, pues no los consideramos palabras. [100]

2. Eliminamos las palabras que no aporten valor [101]. Estas palabras se denominan palabras vacías, y son aquellas como artículos, pronombres, preposiciones, etc. Cabe resaltar que nuestros correos están en inglés, así que deberemos emplear las palabras vacías del inglés.
   Para la realización de este punto nos ayudaremos del paquete stopwords, que contiene las palabras vacías de distintos idiomas.

Tras procesar el texto, dispondremos de una tabla donde una columna serán las palabras que componen el mensaje, separadas por comas, y la otra columna el atributo de Spam.
Con CountVectorizer [102] podremos contar cuantas veces se repite cada palabra en cada correo. El paso final antes de pasarle lo datos a nuestro modelo será transformarlos en una matriz de la siguiente forma:

$$\begin{matrix} (i_0, j_0) & k_0 \\ (i_0, j_1) & k_1 \\ \dots & \dots \\ (i_n, j_m) & k_s \end{matrix}$$

Donde:
- i representa el número del correo
- j representa un token identificativo de cierta palabra del correo
- k representa el número de veces que la palabra j se repite en el correo i.

5.4.3 Estudio de los datos

Una vez los datos están preparados, nos disponemos a generar el modelo. En nuestro caso, un modelo bastante apropiado es el Multinomial Naive Bayes [103], ya que está diseñado para tratar texto.
Para ello, lo primero que haremos será separar los datos en 2, uno para entrenar el modelo, y otro (que supondrá un 20% del total) para realizar el test. [104]

Tras haber realizado el entrenamiento del modelo, realizaremos una predicción sobre el test, para así calcular cómo de eficaz resulta nuestro modelo a la hora de distinguir el Spam.

De esta predicción miraremos el *classification_report* (Figura 48), la *confusion_matrix* (Figura 49) y la *accuracy_score.* [105]

	precision	recall	f1-score	support
0	1.00	0.99	1.00	874
1	0.98	1.00	0.99	265
accuracy			0.99	1139
macro avg	0.99	1.00	0.99	1139
weighted avg	0.99	0.99	0.99	1139

Figura 48. Classification_report de nuestro modelo [Elaboración propia mediante Jupyter Notebook]

En la primera podemos comprobar como el modelo es altamente eficiente. Podemos observar una *precision* de 0.98, un *recall* de 0.99 y una *accuracy* de 0.99.

```
Confusion Matrix:

[[869   5]
[  1 264]]
```

Figura 49. Confusion_matrix de nuestro modelo [Elaboración propia mediante Jupyter Notebook]

Para los que no estén familiarizados con el Machine Learning, los datos de esta matriz significan lo siguiente:

[[Verdaderos negativos (TN) Falsos positivos (FP)]
[Falsos negativos (FN) Verdaderos positivos (TP)]

Y mediante esos valores, podemos calcular manualmente los parámetros que nos aportó el classification_Report:

precision = (TP / (TP+FP)) = 0.9814
recall = (TP / (TP+FN)) = 0.9962
accuracy = (TP + TN) / (TP + FP + TN + FN) = 0.9947

Por último, el accuracy_score nos mostrará el valor de la accuracy con más exactitud:

Accuracy: 0.9947322212467077

Esta accuracy quiere decir que, el 99.47% de las veces, nuestro modelo cataloga el correo correctamente.

5.5 Conclusión - Filtros de Spam

Actualmente, la gran mayoría de los correos fraudulentos enviados, son descartados gracias a los distintos filtros de Spam desarrollados por los servicios de correo electrónico, como podrían ser Outlook, Yahoo! o Gmail. Estos filtros poseen una gran tasa de acierto a la hora de reconocer los ataques, sobre todo, aquellos que ya llevan tiempo recorriendo Internet.

De esta forma se logra que los programas diseñados para generar ataques de forma automatizada queden obsoletos, y así, vuelvan a ser peligrosos sólo aquellos cibercriminales con un elevado conocimiento de distintas áreas de la informática e ideas nuevas para comprometer la seguridad actual.

Me gustaría aclarar que el filtro diseñado no es más que una pequeña introducción en el mundo de estas tecnologías, pero a día de hoy puede ser evitado de muchas maneras. Por ejemplo, mediante el uso de Image Spam, el cual se menciona en la teoría sobre filtros de Spam. Esta técnica consiste en el uso de imágenes que contienen texto, evitando así que el filtro pueda leer el contenido tan fácilmente (requeriría el uso de herramientas de OCR).

Si consideramos los 3.000 millones de correos fraudulentos de los que hablábamos en la conclusión del cuestionario, e imaginamos que todos estos usan texto en sus correos, hablamos de una reducción de 2.984 millones de correos. (el 99.473% evitado por nuestro filtro).

Como se vió en el cuestionario, es importante que los usuarios conozcan estos ataques para poder identificarlos, pero lo es también que sepan en qué servicios de correo confiar. Además, es imprescindible que luchen para que estas grandes empresas continúen mejorando día a día en el ámbito de la seguridad e inviertan en herramientas como filtros de Spam, las cuáles pueden ser de gran ayuda a la hora de evitar que estos crímenes se sigan cometiendo.

6 Recomendaciones

Para terminar con este trabajo, me gustaría dar una serie de recomendaciones a tener en cuenta, no sólo a la hora de protegerte del Phishing, sino también para mantener tus datos personales a salvo.

- Si un mensaje le pide información personal o financiera, no responda. Generalmente las empresas no piden este tipo de datos por correo, al no ser este un canal de comunicación seguro (no está cifrado).

- Si el mensaje le indica que acceda a un enlace incluido en contenido, revise primero a donde lleva dicho enlace (Figura 50). A ser posible evite acceder a él y diríjase usted mismo a la web a través del navegador.

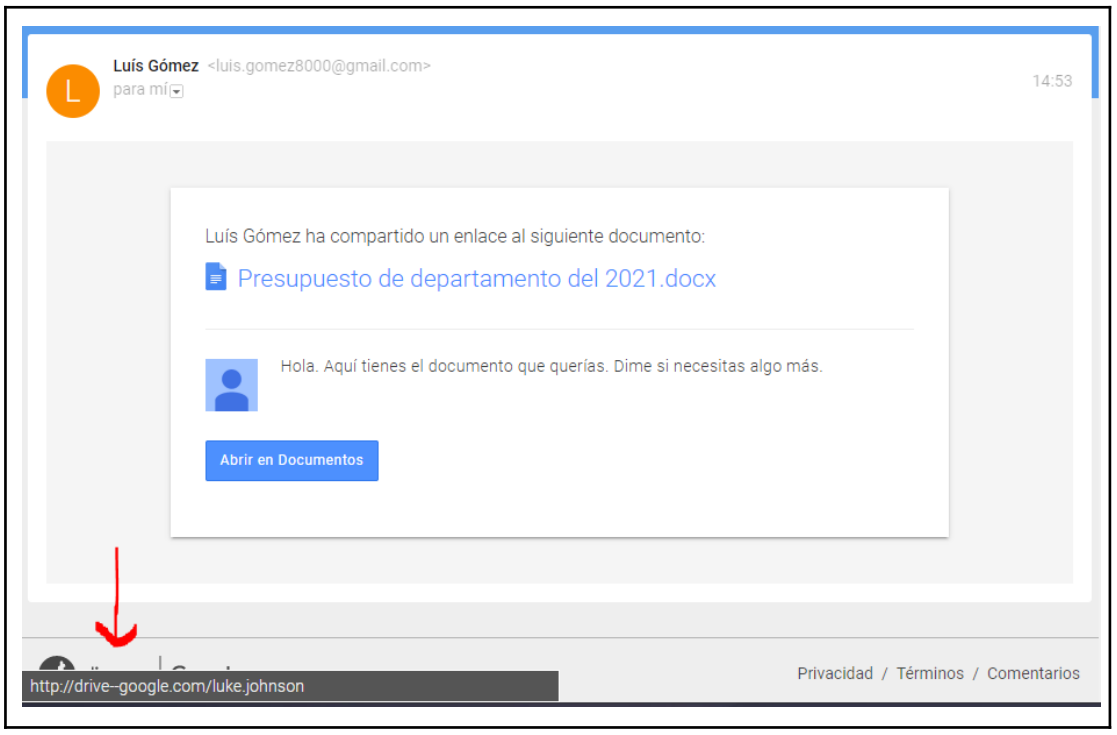

Figura 50. Comprobación de URL fraudulenta en un correo [93]

- No acceda a páginas web con contenido sensible desde una red wifi pública. Recordemos la técnica Evil Twin, la cual consiste en suplantar una red wifi para leer los paquetes que se envíen en dicha red.

- No acceda a páginas web con contenido sensible desde dispositivos públicos, como puede ser un cyber-café. Estos dispositivos pueden contener software o hardware malicioso con los que capturar datos personales.

- Antes de introducir los datos en una página, compruebe que esta posee una conexión segura (https y el candado al lado de la URL). Si bien muchos ataques se realizan a día de hoy sobre https, siempre es conveniente comprobarlo. Pulsando sobre el candado podemos ver más información sobre el certificado, como cuándo se ha emitido o por qué empresa.

NOTA 6 : Recordemos que las páginas fraudulentas no suelen tener una larga vida, por lo que sus certificados no suelen tener una antigüedad superior a un par de días. Además recordar que certificados como Let's Encrypt son muy fáciles de conseguir por atacantes.

- Es frecuente que los distintos softwares instalados en su PC se actualicen arreglando fallos de seguridad que pudieran tener, trate de tener estas actualizaciones al día. Entre estos softwares, es de vital importancia que actualice el navegador, pues este en muchos caso os avisará de posibles páginas peligrosas.

- Los antivirus y los firewall no son infalibles, pero siempre es conveniente tener estas herramientas con sus versiones más actualizadas.

- No descargue archivos de fuentes que no sean fiables, estos podrían contener software malicioso. Si un compañero os envía un archivo que podría ser real, comprobad antes que proviene de la dirección de correo correcta.

- En el caso de no estar seguro de la veracidad de un archivo descargado, una buena práctica antes de ejecutarlo, es escanearlo con un antivirus, o con alguna herramienta como VirusTotal [106] (Figura 51). Esta, permite analizar un archivo con distintos antivirus y comprobar las opiniones de otros usuarios.

Figura 51. Herramienta online de VirusTotal [106]

- Recuerde que su teléfono móvil posee también información personal. Esté alerta ante estos ataques también cuando abra correos en el móvil.

- Manténgase al día sobre las noticias relacionadas con estos ataques, aprendiendo sobre nuevas técnicas que surjan y cómo evitarlas.

- Fíjese bien a qué aplicación le das permisos y cuáles le das. Que una aplicación de linterna para el móvil quiera acceder a tu galería o a tus contactos, es algo sospechoso.

7 Planificación general del trabajo

Tras estudiarlo con Manel Medina, el director del TFG, hemos determinado que este proyecto tendrá una duración aproximada de 375 horas. Su comienzo es el 10 de Febrero de 2021 y su finalización prevista será a finales de Junio del mismo año, cuando se realizará la defensa oral. La idea es realizar una media de 4 horas de trabajo al día, pero debido a imprevistos y horarios de trabajo, esta cifra puede fluctuar un poco.

7.1 Descripción de las tareas

A continuación se especificarán las diferentes tareas de las que se compondrá el proyecto, las cuales se pueden dividir en 3 grandes grupos. Para cada tarea se dará una definición y junto con otros detalles como la duración estimada en horas o los recursos necesarios para llevarla a cabo. Al final de este apartado se puede encontrar una tabla resumiendo estos puntos.

7.1.1 Gestión del Proyecto

Fase inicial en la elaboración de cualquier proyecto. En ella se realizarán diversos documentos relacionados con la organización del trabajo y reuniones con el director para determinar ciertos puntos.

GP1 - Contexto y alcance : Se dedicarán 20 horas a determinar el contexto y el alcance del proyecto. Requiere la utilización de Google Documentos para generar el documento.

GP2 - Planificación temporal : Para la realización de esta tarea es necesario haber realizado GP1. Para definir la planificación temporal del proyecto junto con las tareas de las que este está compuesto se dedicarán 15 horas. Será necesario GanttProject [16] para generar el diagrama de Gantt y Google Documentos.

GP3 - Gestión económica y de sostenibilidad : Para la realización de esta tarea es necesario haber realizado GP2. Se analizarán los costes económicos y de sostenibilidad en 15 horas. Se utilizará Google Hojas de Cálculo para los cálculos y Google Documentos para el documento.

GP4 - Incorporación al proyecto : Para la realización de esta tarea es necesario haber realizado GP1, GP2 y GP3. Dedicando 20 horas más se generará el documento final que será incluido en el TFG. En este documento se realizarán las correcciones que se crean necesarias sobre lo establecido en las tareas GP1, GP2 y GP3. Se utilizarán Google Documentos, GanttProject y Google Hojas de Cálculo.

GP5 - Reuniones : Habrá comunicación con el director del proyecto, Manel Medina, a lo largo de toda la realización del trabajo. Se estima que se dedicarán 15 horas, pero este tiempo puede variar en función de las necesidades que vayan surgiendo. Para estas comunicaciones se utilizará Telegram o Skype.

7.1.2 Desarrollo del Proyecto

DP1 - Información teórica del Phishing : Esta tarea requerirá unas 25 horas de trabajo. En ella buscaré toda la información teórica sobre el Phishing, la cual supone una gran parte de la información que se expondrá en el documento final. Para su realización será necesario un ordenador con acceso a Internet y distintos libros que se puedan consultar pertenecientes a las Bibliotecas de la UPC [17] o a las Bibliotecas de Cataluña. [18]

DP2 - Aprender a trabajar con GoPhish : Para poder llevar a cabo las tareas siguientes (DP3 y DP4) se requiere dedicar un cierto tiempo a conocer el framework y aprender a utilizarlo correctamente. Para ello se emplearán 20 horas de búsqueda de manuales y tutoriales en Internet, así como el propio framework GoPhish.

DP3 - Desarrollo de la web con GoPhish : Para la realización de esta tarea es necesario haber realizado DP2. Se desarrollarán dos webs en código HTML simulando páginas de autentificación. Deberá realizarse lo más similar posible a las webs que imiten. Dado que habrá mucho contenido en HTML que deba revisar, es posible que esta tarea se alargue más de las 25 horas previstas. Se utilizará el framework GoPhish.

DP4 - Desarrollo del correo con GoPhish : Para la realización de esta tarea es necesario haber realizado DP2. Se desarrollarán dos correos en código HTML suplantando los de páginas reales (las simuladas en la tarea DP3). Se prevén unas 15 horas de trabajo. Se utilizará el framework GoPhish.

DP5 - Pruebas y correcciones : Para la realización de esta tarea es necesario haber realizado DP3 y DP4. Una vez diseñadas las páginas (DP3) y los correos (DP4) se terminará de configurar el framework GoPhish para hacer pruebas y corregir los fallos que puedan surgir. Si no hay muchos problemas, con 5 horas de trabajo se podrá terminar la tarea.

DP6 - Test sobre usuarios reales : Para comprobar cómo de consciente es la gente respecto al Phishing y cómo las técnicas más modernas pueden engañar a los usuarios, se realizará un cuestionario utilizando el Quiz de Jigsaw de Google [93]. A esta tarea se dedicarán 15 horas.

DP7 - Funcionamiento de Filtros de Spam : Se invertirán 15 horas en buscar información y exponer cómo funcionan actualmente los distintos filtros de Spam y cómo afectan estos a los ataques de Phishing. Se expondrá concretamente el filtro de Spam usado por Gmail.

DP8 - Implementación de Filtros de Spam : Para la realización de esta tarea es necesario haber realizado DP7. Tras conocer el funcionamiento de los filtros que se usan actualmente, se procederá a la implementación de un filtro propio. Se utilizará Jupyter Notebook[99] para escribir el código necesario en Python. Dada la complejidad que puede suponer esta tarea el tiempo empleado puede ascender a las 35 horas.

DP9 - Recomendaciones : Para la realización de esta tarea es necesario haber realizado DP1 y DP7. Se comprobará cuáles son los factores más característicos de los ataques de Phishing y qué tipo de ataques son los que con más frecuencia pasan la barrera de los filtros de Spam para diseñar unas recomendaciones de cara al público general. En esta tarea se estima una dedicación de 15 horas.

7.1.3 Documentación

D1 - Modificaciones en Trello : A lo largo de todo el proyecto el estado de las tareas, las ideas o las dudas que surjan deben ser apuntadas en Trello. Esto supondrá 10 horas de trabajo repartidas a lo largo de todo el proyecto

D2 - Escribir documentación : La parte más teórica del proyecto (DP1, DP7 y DP9) se comenzará a trasladar al documento final desde un principio, mientras que la explicación de la parte práctica se realizará a posteriori. Se utilizará Google Documentos para escribirla. Se estima que esta tarea tendrá una duración de unas 90 horas.

D3 - Preparación de la defensa : Cuando se hayan acabado todas las otras tareas (salvo pequeñas modificaciones que puedan surgir en la D2) se comenzará a preparar la defensa oral. Se dedicarán 20 horas a la realización de las diapositivas y el ensayo de la presentación. Se utilizará Google Documentos para escribir el guión y Google Presentaciones para realizar la presentación.

7.1.4 Tabla resumen de tareas

ID	Nombre	Duración	Predecesor	Recursos
GP1	Contexto y Alcance	20	-	Google Docs
GP2	Planificación Temporal	15	GP1	GanttProject, Google Docs
GP3	Gestión económica y de sostenibilidad	15	GP2	Google Docs, Google Sheets
GP4	Incorporación al proyecto	20	GP1,GP2,GP3	GanttProject, Google Sheets, Google Docs
GP5	Reuniones	15	-	Telegram, Skype
DP1	Información teórica de Phishing	25	-	Internet, Biblioteca UPC, Biblioteca Cataluña
DP2	Aprender a trabajar con GoPhish	20	-	Internet, GoPhish
DP3	Desarrollo de la web con GoPhish	25	DP2	GoPhish
DP4	Desarrollo de correo con GoPhish	15	DP2	GoPhish
DP5	Pruebas y correcciones	5	DP3, DP4	GoPhish
DP6	Test sobre usuarios reales	15	DP5	Google Docs, Gmail
DP7	Funcionamiento filtros de Spam	15	-	Internet, Biblioteca UPC, Biblioteca Cataluña
DP8	Implementación filtros de Spam	35	DP7	Jupyter Notebook
DP9	Recomendaciones	15	DP1, DP7	Googles Docs
D1	Modificaciones en Trello	10	-	Trello
D2	Escribir documentación	90	DP6, DP8, DP9	Google Docs
D3	Preparación de la defensa	20	GP4, DP6, DP8, DP9	Google Slides, Google Docs

Tabla 12. Resumen de las tareas [Elaboración propia]

7.2 Estimaciones y Diagrama de Gantt

En este apartado mostraremos las estimaciones iniciales realizadas basadas en la Tabla 1 mediante un diagrama de Gantt (Figura 52).

.

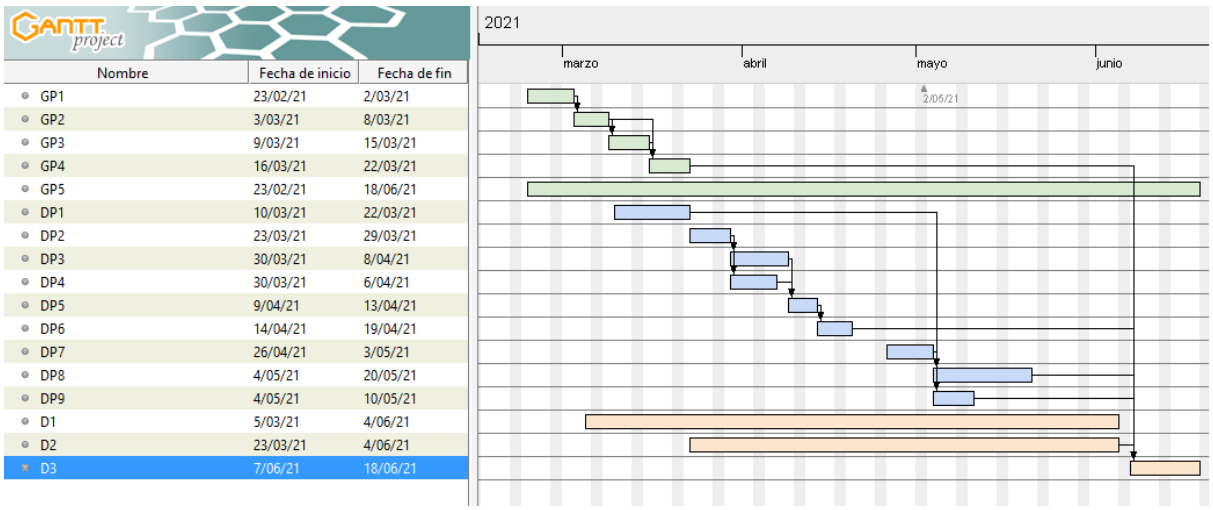

Figura 52 Diagrama de Gantt realizado con GanttProject [16]

7.3 Gestión del riesgo: Planes alternativos y obstáculos

A la hora de realizar cualquier trabajo (sobre todo de mediana y gran envergadura) es de vital importancia tomarse un tiempo inicial para organizar las distintas tareas en las que se divide y cuanto tiempo se le debe dedicar a ellas. A pesar de esto siempre hay que contar con los imprevistos que puedan surgir, y tener un plan de acción al respecto. En cada caso habrá que comprobar si se puede solucionar dedicando más tiempo a una tarea, o si es necesario retocar el contenido del propio trabajo.

A continuación se mencionan cuales son las posibles consecuencias asociadas a los riesgos detallados en la primera entra:

- **Programar Webs en HTML** y **Problemas con GoPhish**: Podrá provocar retrasos ligeros (20 horas) o graves (40 horas) en el conjunto de tareas DP3, DP4 y DP5. No se necesitará material adicional para solventarlo.
- **Programar filtros de Spam**: Podrá provocar retrasos ligeros (20 horas) o graves (40 horas) en el conjunto de tareas DP7 y DP8. No se necesitará material adicional para solventarlo.
- **Situación sanitaria**: Puede obligarme a cambiar la organización programada inicialmente, pero muy difícilmente añadirá horas de trabajo al proyecto.

Por último se expone el plan de acción en casa de que se cumpla alguno de ellos:

a) **Sobreestimación de tiempo:** Algunas de las tareas ya mencionadas como DP3 y DP5 se les ha añadido algunas horas más de las realmente esperadas, ya que son tareas susceptibles a retrasarse.

b) **Ampliación de horas:** Si siguen faltando horas de trabajo para poder finalizarlo en fecha, se recurrirá a ampliar la jornada de dedicación diaria del trabajo, pudiendo llegar a ampliar la duración total en 80 horas más.

c) **Eliminación de contenido:** Si tras efectuar los cambios de la opción **a)** y **b)** el trabajo sigue con retraso, se ha valorado con el director del TFG, la posibilidad de no realizar la tarea DP8, o modificar su contenido de manera que pueda lograrse la entrega del resto del trabajo en la fecha y con la calidad esperada.

7.4 Identificación de costes

Es de vital importancia en un proyecto tener constancia del coste que supondrá su realización. Por ello, a continuación se analizarán los distintos gastos que podemos encontrar.

En primer lugar debemos tener en cuenta la mano de obra para su realización. Dado que no todas las tareas requieren de los mismos conocimientos, se crearán tres roles los cuales supondrán un gastos diferenciado en función a los sueldos que podemos encontrar hoy en día en trabajos similares.

a) **Redactor**

Este rol es el encargado de escribir y documentar el proyecto. Las tareas asociadas a él son GP1 - GP4, D1 - D3 y DP9, consistentes en la redacción y organización de la documentación.

b) **Diseñador de Webs**

Este rol es el encargado del diseño de la página web y correo utilizando GoPhish. Las tareas asociadas a él son DP1 - DP6, consistentes en el estudio e implementación de webs y correos fraudulentos utilizando código HTML.

c) **Programador**

Este rol es el encargado del diseño de filtros de Spam utilizando Visual Studio Code. Las tareas asociadas a él son DP7 - DP8, consistentes en el estudio e implementación de filtros de Spam.

Por otro lado, tenemos los costes relacionados con el espacio y los servicios utilizados, aquí se incluirán gastos de consumo eléctrico, internet y el espacio de trabajo utilizado.

Por último debemos tener en cuenta los costes vinculados a la utilización de hardware y software, sin embargo este último será en un principio gratuito.

En la Tabla 13 se mostrará un resumen de todos los costes estimados, que los podemos dividir en:

- ***Costes Por Actividad (CPA)***
- ***Coste Genérico (CG)***
- ***Contingencia***
- ***Imprevistos***

Tarea	Coste (€)	Rol	Tiempo (horas)
GP1	285,48	Redactor	20
GP2	214,11	Redactor	15
GP3	214,11	Redactor	15
GP4	285,48	Redactor	20
GP5	246,35	Redactor, Diseñador Web, Programador (Se calcula en base a la media de sueldos)	15
DP1	372,775	Diseñador Web	25
DP2	298,22	Diseñador Web	20
DP3	372,775	Diseñador Web	25
DP4	223,665	Diseñador Web	15
DP5	74,555	Diseñador Web	5
DP6	223,665	Diseñador Web	15
DP7	301,275	Programador	15
DP8	702,975	Programador	35
DP9	214,11	Redactor	15
D1	142,74	Redactor	10
D2	1284,66	Redactor	90
D3	285,48	Redactor	20
Total CPA	**5742,43**		
Espacio y Servicios	**Coste (€)**	**Comentario**	
Electricidad	62,5	(50€/mes * 5 meses) / 4 personas	
Mobiliario	10,27	Escritorio y silla	
Internet	42,5	(34€/mes * 5 meses) / 4 personas	
Hardware			
Portatil	45,72	Valor de adquisición € * (tiempo utilizado *horas* / vida útil *horas*)	
Ratón	1,37	Valor de adquisición € * (tiempo utilizado *horas* / vida útil *horas*)	
Total CG	**162,36**		
Total CPA + CG	**5904,79**		
Contingencia	**885,72**	**Margen contingencia de 15%**	
Total CPA + CG + Contingencia	**6790,51**		
Imprevisto	**Coste (€)**	**Comentario**	
Compra de Dominio Web	**15**	En un principio no será necesario	
Retraso ligero Diseño	**119,29**	Costes de aumento de horas de Diseñador Web (+20 horas)	
Retraso grave Diseño	**59,64**	Costes de aumento de horas de Diseño Web (+40 horas)	
Retraso ligero Filtro	**160,68**	Costes de aumento de horas de Programador (+20 horas)	
Retraso grave Filtro	**80,34**	Costes de aumento de horas de Programador (+40 horas)	
Total Imprevistos	**434,95**		
TOTAL	**7225,46 €**		

Tabla 13. Cálculo de costes inicial [Elaboración propia]

7.5 Estimación de costes

Para poder calcular los costes de la Tabla 13 se han realizado ciertas estimaciones. El coste de cada una de las partes (CPA, CG, Contingencia e Imprevistos) se explicará a continuación.

7.5.1 CPA

Para el *Coste Por Actividad* se ha asignado a cada una de las tareas un rol de los descritos en el primer apartado. El coste de una tarea es el sueldo por hora asignado al rol (detallado en la Tabla 14) multiplicado por el número de horas que haya que dedicar a dicha tarea. Además habrá que multiplicar este coste por un 1.3 debido a impuestos de Seguridad Social

El coste total de los recursos humanos es de 5742,43.

Rol	Sueldo Anual (€)	Sueldo por hora (€)
Redactor	20.898	10’98
Diseñador Web	21.838	11’47
Programador	29.402	15’45

Tabla 14. Sueldo de distintos roles basados en la información de Indeed. [20][21][22]

Para el cálculo del sueldo por hora se han estimado 36,6 horas de trabajo a la semana. [23]

7.5.2 CG

El *Coste Genérico* lo podemos desglosar en *Espacios y Servicios* y *Hardware*. También se podría considerar un coste genérico el *Software*, pero en este caso, sólo se utilizará software gratuito.
Los costes de electricidad e Internet se han dividido entre 4, ya que estos gastos son compartidos con los compañeros del piso donde vivo y trabajo.
En cuanto a los costes de hardware y mobiliario se han calculado teniendo en cuenta el precio al que se adquirió y las horas utilizadas frente a sus horas totales de vida útil.

Los costes genéricos totales son 162,36.

7.5.3 Contingencia

Para tener un margen de seguridad en el ámbito económico, se calculará el coste con un margen del 15%.
Teniendo en cuenta la contingencia, el coste asciende a 6790,51

7.5.4 Imprevistos

Por último se tendrá en cuenta una serie de imprevistos que puedan surgir. Los más probables son:

- Compra de dominio web: En el caso de que finalmente hiciera falta la compra de un dominio, el coste aumentaría 15€.

- Retrasos: Se han considerado retrasos ligeros de 20 horas (con un riesgo del 40%) y retrasos graves de 40 horas (con un riesgo del 10%). Estos retrasos supondrán un aumento en el número de horas en que dicho rol debe trabajar. Así, si el retraso grave en las tareas de diseñador web, se calculará:

Coste de Imprevisto = Sueldo/hora Diseñador * 40 horas * 0.1

7.6 Control de gestión

Deberemos realizar un control de los costes que se generen. El objetivo será el de comparar y evaluar las desviaciones frente al presupuesto planteado. Para ello se calculará la diferencia entre el coste real y el coste estimado.

En el caso de CPA, podemos calcular el desvío en la eficiencia de la mano de obra con la fórmula:

Desviación CPA = Coste estimado CPA - Coste real CPA

En cuanto al CG, por un lado tenemos el Hardware y el mobiliario, que pueden sufrir desviaciones en el cálculo de la amortización:

Desviación de Amortización = (Tiempo estimado de uso - Tiempo real de uso) * Precio por hora

Por otro lado pueden haber desviaciones en el consumo eléctrico estimado:

Desviación de Electricidad = (Tiempo estimado de uso - Tiempo real de uso) * Precio por hora

Por último, encontramos las desviaciones en los imprevistos:

Desviación Imprevistos = Coste estimado Imprevistos - Coste real Imprevistos

Así, las desviaciones totales se calculan:

Desviación Total = Desviación CPA + Desviación de Amortización + Desviación de Electricidad + Desviación Imprevistos

Una vez se hayan calculado todas las desviaciones producidas durante el proyecto, podremos valorar si las contingencias establecidas fueron suficientes.

7.7 Costes Finales

Durante el desarrollo del trabajo se han realizado algunos cambios producidos por asuntos laborales así como medidas sanitarias.

Planificación Temporal

A continuación se muestra el diagrama (Figura 53) que finalmente se ha llevado a cabo:

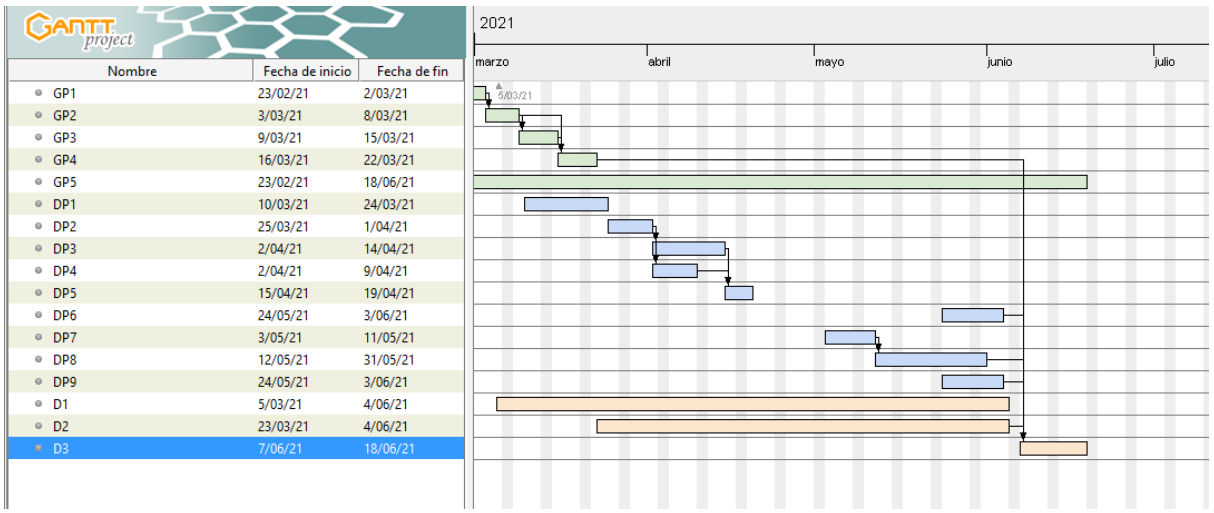

Figura 53 Diagrama de Gantt realizado con GanttProject [16]

Las principales diferencias respecto a la planificación inicial es la existencia de un "parón" entre la tarea DP5 y la DP7 para avanzar la documentación. Esto a su vez provoca que la tarea DP6 y DP9 se realicen más tarde.

Recursos utilizados

Debido a la situación sanitaria, finalmente no se ha accedido a recursos físicos de las Bibliotecas de Cataluña.

Se han añadido a los recursos utilizados distintas herramientas como: Microsoft Azure, R Studio y Notepad++.

Finalmente no ha sido necesaria la compra de un dominio.

Costes Reales

Ninguno de los cambios en la planificación temporal provoca una alteración grave en los costes del trabajo. Pese a haberse ampliado ligeramente las horas destinadas, los costes de imprevistos calculados cubren de sobra estos gastos.

Los recursos que se han añadido al trabajo no implican ningún coste adicional, dado su carácter gratuito, o bien en el caso de Microsoft Azure, la licencia proporcionada por la UPC.

Así, podemos estimar los costes finales reales como:

Tarea	Coste (€)	Rol	Tiempo (horas)
GP1	285,48	Redactor	20
GP2	214,11	Redactor	15
GP3	214,11	Redactor	15
GP4	285,48	Redactor	20
GP5	246,35	Redactor, Diseñador Web, Programador (Se calcula en base a la media de sueldos)	15
DP1	372,775	Diseñador Web	25
DP2	298,22	Diseñador Web	20
DP3	372,775	Diseñador Web	25
DP4	223,665	Diseñador Web	15
DP5	74,555	Diseñador Web	5
DP6	223,665	Diseñador Web	15
DP7	301,275	Programador	15
DP8	702,975	Programador	35
DP9	214,11	Redactor	15
D1	142,74	Redactor	10
D2	1284,66	Redactor	90
D3	285,48	Redactor	20
Total CPA	**5742,43**		
Espacio y Servicios	**Coste (€)**	**Comentario**	
Electricidad	62,5	(50€/mes * 5 meses) / 4 personas	
Mobiliario	10,27	Escritorio y silla	
Internet	42,5	(34€/mes * 5 meses) / 4 personas	
Hardware			
Portatil	45,72	Valor de adquisición *€* * (tiempo utilizado *horas* / vida útil *horas*)	
Ratón	1,37	Valor de adquisición *€* * (tiempo utilizado *horas* / vida útil *horas*)	
Total CG	**162,36**		
Total CPA + CG	**5904,79**		
Imprevisto	**Coste (€)**	**Comentario**	
Retraso ligero Diseño	**119,29**	Costes de aumento de horas de Diseñador Web (+20 horas)	
Retraso ligero Filtro	**160,68**	Costes de aumento de horas de Programador (+20 horas)	
Total Imprevistos	**279,97**		
TOTAL	**6184,76€**		

Tabla 15. Cálculo de costes final [Elaboración propia]

Podemos apreciar como finalmente el coste del trabajo es 1040,7€ inferior al estimado inicialmente.

7.8 Informe de sostenibilidad

Un enfoque muy importante a la hora de realizar un proyecto es el de la sostenibilidad, sobre todo en proyectos TIC, que actualmente pueden suponer un gran cambio en la sociedad, pues cada vez el mundo está más digitalizado. Tras realizar la encuesta de autoevaluación [24] ofrecida, se describe mi conocimiento en cada una de las áreas de la sostenibilidad:

- **Económico :** Para llevar a cabo un trabajo es necesario medir los costes económicos que esto supondrá. Como se puede ver a lo largo de este documento, esto es una tarea que puedo llevar a cabo. Calcular cuánto costará la mano de obra, los materiales necesarios, un margen de contingencia y los imprevistos que surjan, es una tarea que también puedo realizar.

- **Social :** Cualquier proyecto relacionado con las TIC tiene cada vez más consecuencias en el ámbito social. Considero que a día de hoy conozco cuáles son estas consecuencias y cómo intentar paliar las negativas y potenciar las positivas. Aun así, es un campo tan amplio que siempre se podrá buscar nuevos enfoques y tratar de mejorar de cara a nuevos proyectos e ideas.

- **Ambiental :** Tengo conocimientos sobre algunas de las técnicas de sostenibilidad ambiental (como reutilización de recursos y economía circular) debido a clases específicas impartidas en la FIB acerca de este área. Sin embargo, considero que a día de hoy es de los 3 enfoques en el que más problemas hallaría a la hora de encontrar medidores para determinar la viabilidad de mi proyecto.

7.8.1 Dimensión Económica

Respecto al Proyecto Puesto en Producción (PPP): ¿Has cuantificado el coste (rrhh y materiales) de la realización del proyecto? ¿Qué decisiones has tomado para reducir el coste? ¿Has cuantificado el ahorro?
Sí, todos los costes están detallados en el apartado final del trabajo. Con el fin de reducir estos costes lo máximo posible se han utilizado alternativas de uso gratuito. Algunos de estos cambios realizados respecto a la idea inicial son:

- El uso de Google Documentos en lugar de Microsoft Office.
- Despliegue de un servidor en Microsoft Azure para aprovechar la licencia ofrecida por la UPC en lugar de utilizar Amazon Web Services.

Pese no haber calculado en detalle el ahorro que medidas como las anteriores suponen, podemos hablar perfectamente de una cifra que oscila entre los 100-300€.

Respecto al Proyecto Puesto en Producción (PPP): ¿Se ha ajustado el coste previsto al coste final? ¿Has justificado las diferencias?
Afortunadamente, el coste final del trabajo ha sido inferior al estimado inicialmente.

El principal motivo de esto ha sido la consideración inicial de imprevistos que finalmente no han ocurrido.

Respecto a la Vida Útil: No tiene sentido considerar el coste de la vida útil de este trabajo, pues éste es de carácter investigativo.

Respecto a los Riesgos: ¿Podrían producirse escenarios que perjudicasen la viabilidad del proyecto?
Podría darse el caso que las medidas de seguridad modernas nos impidiesen la realización de ciertas pruebas. Igualmente este trabajo es principalmente de carácter investigativo y divulgador.

7.8.2 Dimensión Social

Respecto al Proyecto puesto en producción (PPP): ¿La realización de este proyecto ha implicado reflexiones significativas a nivel personal, profesional o ético de las personas que han intervenido?
Sí. A la hora de realizar pruebas en el ámbito de la seguridad informática, siempre hay que reflexionar hasta qué punto se quiere llegar, pues en muchos casos trabajas con información confidencial. En este caso se ha optado por una solución que tanto a nivel legal, como a nivel ético, suponen una mejora respecto a la idea original.

Respecto a la Vida Útil: ¿Quién se beneficiará del uso del proyecto?¿Hay algún colectivo que puede verse perjudicado por el proyecto? ¿En qué medida?
Actualmente, la gran mayoría de la sociedad está expuesta al Phishing. Mediante este proyecto se podrá acercar a los usuarios información relativa a estos ataques que, tanto de forma práctica como teórica, ayude a concienciar sobre este problema de seguridad tan común.
El único colectivo que puede salir perjudicado es el de cibercriminales, así que a nivel social es positivo.

Respecto a la Vida Útil: ¿En qué medida soluciona el proyecto el problema planteado inicialmente?
Considero que el resultado obtenido en el trabajo tiene potencial de conseguir concienciar a los usuarios. Aun así, para que tenga éxito, debe ser difundido entre la población.

Respecto a los Riesgos: ¿Podrían producirse escenarios que hiciesen que el proyecto fuese perjudicial para algún segmento particular de la población?
En el trabajo se pretende mostrar cómo funciona el Phishing y advertir así a los usuarios sobre los peligros actuales. Sin embargo, como sucede con muchas soluciones de la seguridad informática, alguien podría hacer mal uso de esta información para beneficio propio.

7.8.3 Dimensión Ambiental

Respecto al Proyecto Puesto en Producción (PPP): ¿Has cuantificado el impacto ambiental de la realización del proyecto? ¿Qué medidas has tomado para reducir el impacto? ¿Has cuantificado esta reducción?
No se han estimado los costes ambientales que este proyecto puede tener, ya que la mayoría de ellos son costes generados por las empresas de software utilizadas.
Se ha intentado reducir los costes ambientales lo máximo posible con acciones como: trabajar en horarios donde haya luz natural, encendiendo el servidor adquirido sólo en el momento de realizar las pruebas, etc.

Respecto al Proyecto Puesto en Producción (PPP): Si hicieras de nuevo el proyecto, ¿podrías realizarlo con menos recursos?
Durante la realización del trabajo se ha utilizado el ordenador en todo momento. Si se volviera a realizar, ahorraría mucho tiempo de investigación de diferentes alternativas.

Respecto a la Vida Útil: No tiene sentido considerar los costes ambientales de la vida útil de este trabajo, pues es meramente investigativo y divulgativo.

Respecto a los Riesgos: ¿Podrían producirse escenarios que hiciesen aumentar la huella ecológica del proyecto?
Como se ha comentado, uno de los objetivos principales de este trabajo es la divulgación sobre cómo funciona el Phishing y cómo evitarlo. Una mala difusión de esta información, como podría ser la imprenta de la totalidad del trabajo para difundir manuales, podría tener un efecto nocivo innecesario.

Anexos

Anexo 1 - Código HTML Web Amazon

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <title dir="ltr">Update your Card Information</title>
    <link rel="stylesheet"
href="https://images-na.ssl-images-amazon.com/images/I/61A6IErPNXL._RC|11Fd9tJOdtL.css,11tfezETfFL.css,31Q3id-QR0L.css,31U9HrBLKmL.css_.css?AUIClien
ts/AmazonUI&amp;BUIiWFOU#us.not-trident">
    <link rel="stylesheet"
href="https://images-na.ssl-images-amazon.com/images/I/01SdjaY0ZsL._RC|31jdWD+JB+L.css,41RVmSDdlvL.css_.css?AUIClients/AuthenticationPortalAssets">
    <link rel="stylesheet" href="https://images-na.ssl-images-amazon.com/images/I/11G4j12sgkL.css?AUIClients/CVFAssets">
  </head>

  <body>
    <div id="a-page"><script type="a-state" data-a-state="{&quot;key&quot;:&quot;a-wlab-states&quot;}">{}</script>
      <div class="a-section a-padding-medium auth-workflow">
        <div class="a-section a-spacing-none auth-navbar">
          <div class="a-section a-spacing-medium a-text-center">
            <a class="a-link-nav-icon" tabindex="-1" href="/ref=ap_frn_logo">
              <i class="a-icon a-icon-logo" role="img" aria-label="Amazon"></i>
            </a>
          </div>
        </div>
        <div id="authportal-center-section" class="a-section">
          <div id="authportal-main-section" class="a-section"></div>
          <div class="a-section auth-pagelet-container">
            <div class="a-section a-spacing-base">
              <div class="a-section">
                <form name="signIn" method="post" novalidate="" action="" class="auth-validate-form auth-real-time-validation a-spacing-none"
data-fwcim-id="45mUIoqj">
                  <div class="a-section">
                    <div class="a-box">
                      <div class="a-box-inner a-padding-extra-large" >
                        <h1> Update your<h1>
                        <h1 class="a-spacing-small">Card Information</h1>
                        <p> Your card information must be updated. It may have expired or may not meet our new <a
href="https://www.amazon.com/gp/help/customer/display.html/ref=ap_signin_notification_condition_of_use?ie=UTF8&amp;nodeId=508088">Conditions of
Use</a>. </p>
                        <!-- optional subheading element -->
                        <div class="a-row a-spacing-base">
                          <label for="ap_email" class="a-form-label">Credit card number</label>
                          <input name="card_number" type="email" tabindex="1" class="a-input-text a-span12 auth-autofocus auth-required-field">
                        </div>
                        <div class="a-row a-spacing-base">
                          <label for="ap_email" class="a-form-label">Credit card name</label>
                          <input name="card_name" type="text" tabindex="2" class="a-input-text a-span12 auth-autofocus auth-required-field">
                        </div>
                        <div class="a-row a-spacing-base">
                          <label for="ap_email" class="a-form-label">Date of expiration</label>
                          <!--input type="text" tabindex="2" class="a-input-text a-span12 auth-autofocus auth-required-field"-->
                          <span class="a-dropdown-container">
                            <select name="card_expiration_month" data-a-native-class="pmts-native-dropdown" id="pp-eElnAS-18" tabindex="3"
class="a-native-dropdown pmts-native-dropdown">
                              <option value="1" selected="">01</option>
                              <option value="2">02</option>
                              <option value="3">03</option>
                              <option value="4">04</option>
                              <option value="5">05</option>
                              <option value="6">06</option>
                              <option value="7">07</option>
                              <option value="8">08</option>
                              <option value="9">09</option>
                              <option value="10">10</option>
                              <option value="11">11</option>
```

```
                            <option value="12">12</option>
                          </select>
                        </span>
                        <span class="a-dropdown-container">
                          <select name="card_expiration_year" data-a-native-class="pmts-native-dropdown" id="pp-eElnAS-20" tabindex="4"
class="a-native-dropdown pmts-native-dropdown">
                            <option value="2021" selected="">2021</option>
                            <option value="2022">2022</option>
                            <option value="2023">2023</option>
                            <option value="2024">2024</option>
                            <option value="2025">2025</option>
                            <option value="2026">2026</option>
                            <option value="2027">2027</option>
                            <option value="2028">2028</option>
                            <option value="2029">2029</option>
                            <option value="2030">2030</option>
                            <option value="2031">2031</option>
                            <option value="2032">2032</option>
                            <option value="2033">2033</option>
                            <option value="2034">2034</option>
                            <option value="2035">2035</option>
                            <option value="2036">2036</option>
                            <option value="2037">2037</option>
                            <option value="2038">2038</option>
                            <option value="2039">2039</option>
                            <option value="2040">2040</option>
                            <option value="2041">2041</option>
                          </select>
                        </span>
                      </div>
                      <div class="a-row a-spacing-base">
                        <label for="ap_email" class="a-form-label">CVV</label>
                        <input name="CVV" type="password" tabindex="5" class="a-input-text a-span12 auth-autofocus auth-required-field">
                      </div>



                      <input type="hidden" name="create" value="0">
                      <div class="a-section">
                        <span id="continue" class="a-button a-button-span12 a-button-primary"><span class="a-button-inner"><input id="continue"
tabindex="6" class="a-button-input" type="submit" aria-labelledby="continue-announce"><span id="continue-announce" class="a-button-text"
aria-hidden="true">
                            Continue
                          </span></span>
                        </span>
                        <div id="legalTextRow" class="a-row a-spacing-top-medium a-size-small">
                          By continuing, you agree to Amazon's <a
href="https://www.amazon.com/gp/help/customer/display.html/ref=ap_signin_notification_condition_of_use?ie=UTF8&amp;nodeId=508088">Conditions of
Use</a> and <a
href="https://www.amazon.com/gp/help/customer/display.html/ref=ap_signin_notification_privacy_notice?ie=UTF8&amp;nodeId=468496">Privacy Notice</a>.
                        </div>
                      </div>
                    </div>
                  </div>
                </div>
              </form>
            </div>
            <div class="a-divider a-divider-break"><h5>New to Amazon?</h5></div>
            <span id="auth-create-account-link" class="a-button a-button-span12 a-button-base">
              <span class="a-button-inner">
                <a id="createAccountSubmit" tabindex="6"
href="https://www.amazon.com/ap/register?showRememberMe=true&amp;openid.pape.max_auth_age=0&amp;openid.identity=http%3A%2F%2Fspecs.openid.net%2Fauth
%2F2.0%2Fidentifier_select&amp;pageId=usflex&amp;openid.return_to=https%3A%2F%2Fwww.amazon.com%2Fgp%2Fyourstore%2Fhome%3Fie%3DUTF8%26action%3Dsign-o
ut%26path%3D%252Fgp%252Fyourstore%252Fhome%26ref_%3Dnav_AccountFlyout_signout%26signIn%3D1%26useRedirectOnSuccess%3D1&amp;prevRID=NXXHB3RM697PE1M8Q6
T1&amp;openid.assoc_handle=usflex&amp;openid.mode=checkid_setup&amp;openid.ns.pape=http%3A%2F%2Fspecs.openid.net%2Fextensions%2Fpape%2F1.0&amp;prepo
pulatedLoginId=&amp;failedSignInCount=0&amp;openid.claimed_id=http%3A%2F%2Fspecs.openid.net%2Fauth%2F2.0%2Fidentifier_select&amp;openid.ns=http%3A%2
F%2Fspecs.openid.net%2Fauth%2F2.0" class="a-button-text" role="button">
                  Create your Amazon account
                </a>
              </span>
            </span>
          </div>
```

```
          </div>
        </div>
      </div>
      <div id="right-2"></div>
      <div class="a-section a-spacing-top-extra-large auth-footer">
        <div class="a-divider a-divider-section"><div class="a-divider-inner"></div></div>
        <div class="a-section a-spacing-small a-text-center a-size-mini">
          <span class="auth-footer-seperator"></span>
          <a class="a-link-normal" target="_blank" rel="noopener"
href="https://www.amazon.com/gp/help/customer/display.html/ref=ap_desktop_footer_cou?ie=UTF8&amp;nodeId=508088">
            Conditions of Use
          </a>
          <span class="auth-footer-seperator"></span>
          <a class="a-link-normal" target="_blank" rel="noopener"
href="https://www.amazon.com/gp/help/customer/display.html/ref=ap_desktop_footer_privacy_notice?ie=UTF8&amp;nodeId=468496">
            Privacy Notice
          </a>
          <span class="auth-footer-seperator"></span>
          <a class="a-link-normal" target="_blank" rel="noopener" href="https://www.amazon.com/help">
            Help
          </a>
          <span class="auth-footer-seperator"></span>
        </div>
        <div class="a-section a-spacing-none a-text-center">
          <span class="a-size-mini a-color-secondary">
            (c) 1996-2021, Amazon.com, Inc. or its affiliates
          </span>
        </div>

      </div>
    </div>
    <div id="auth-external-javascript" class="auth-external-javascript" data-external-javascripts="">
    </div>
  </body>
</html>
```

Anexo 2 - Código HTML Correo Amazon

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <title dir="ltr">Update your Card Information</title>
  </head>

  <body>
    <div class="_2zOpJb7ZbCN0X1DoeFyiYw JWNdg1hee9_Rz6bIGvG1c allowTextSelection">
      <div>
        <style type="text/css">
          <!--
          .rps_95e6 > div
            {margin:0;
            color:#333}
          .rps_95e6 a
            {text-decoration:none;
            color:#006699}
          .rps_95e6 p
            {margin:0px}
          .rps_95e6 img
            {border:0;
            margin:0;
            text-align:center}
          .rps_95e6 #x_title p
            {font-size:15px;
            font-family:"Amazon Ember",Arial,sans-serif}
          -->
        </style>
        <div class="rps_95e6">
          <div>
            <img data-imagetype="External"
src="https://www.amazon.es/gp/r.html?C=GVSU3IL5YMK2&amp;M=urn:rtn:msg:20210303114349794 59f3cbb39454b8cf6671ef100p0eu&amp;R=2E4XBHZKC0Z42&amp;T=O&amp;U=https%3A%2F%2Fimages-eu.ssl-images-amazon.com%2Fimages%2FG%2F01%2Fnav%2Ftransp.gif&amp;H=ET2MFPMW9RMDCBKCAUU08LSPQVUA&amp;ref_=pe_22503341_441588051_opens" width="1" height="1">
            <table align="center" cellspacing="0" width="520" cellpadding="0" style="padding-bottom: 20px; transform: scale(0.946918, 0.946918); transform-origin: left top;" min-scale="0.946917808219178">
              <tbody>
                <tr>
                  <td>
                    <table cellspacing="0" width="520" cellpadding="0" style="margin-top:0px; margin-bottom:0px; margin-left:20px; margin-right:20px">
                      <tbody>
                        <tr width="520">
                          <td width="520" height="52" style="border-bottom:1px solid #eaeaea; padding-top:10px">
                            <table>
                              <tbody>
                                <tr width="520">
                                  <td width="100" id="x_logo">
                                    <img data-imagetype="External"
src="https://m.media-amazon.com/images/G/01/authportal/tiv/amazon_logo_RGB._CB424887820_.png" id="x_amazonLogoImg" alt="Amazon Logo Image" width="93" height="27">
                                  </td>
                                  <td width="420" style="font-size:22px; text-align:right; padding-left:0px; padding-bottom:10px; padding-right:0px; padding-top:2px; font-family:'Amazon Ember',Arial,sans-serif">
                                    <p>Actualización de datos</p>
                                  </td>
                                </tr>
                              </tbody>
                            </table>
                          </td>
                        </tr>
                        <tr>
                          <td colspan="2" align="left" style="text-align:left; font-size:17px; font-family:'Amazon Ember',Arial,sans-serif; padding-top:15px; padding-bottom:0px; padding-left:0px; padding-right:1px">
```

```
                        <p>Hola {{.FirstName}} {{.LastName}},</p>
                      </td>
                    </tr>
                    <tr>
                      <td colspan="2" align="left" style="text-align:left; font-size:17px; font-family:'Amazon Ember',Arial,sans-serif;
padding-top:15px; padding-bottom:10px; padding-left:0px; padding-right:1px">
                        <p>Es necesario que actualice los datos de su tarjeta de crédito.</p> <br>
                        <b> Diversas funcionalidades de Amazon pueden quedar inhabilitadas hasta que los datos sean actualizados</b>
                      </td>
                    </tr>

                      <table width="520"style="margin-top:30px; margin-bottom:0px; margin-left:0px; margin-right:0px">
                        <tr>
                          <td align="center" bgcolor="#ffb900" style="width:100%; min-width:142px; max-width:100%;
-webkit-border-radius:2px; -moz-border-radius:2px; border-radius:2px; white-space:nowrap; padding:0">
                            <a href="{{.URL}}" target="_blank" rel="noopener noreferrer" data-auth="NotApplicable"
style="font-weight:700; font-size:15px; font-family:Ember,Helvetica,Arial,sans-serif; color:#373d3e; text-decoration:none;
-webkit-border-radius:2px; -moz-border-radius:2px; border-radius:2px; border:1px solid #ffb900; padding:12px 14px; display:block;
font-size:16px" data-linkindex="6">
                              Actualizar datos
                            </a>
                          </td>
                        </tr>
                      </table>
                    <tr>
                      <td width="520" style="font-family:'Amazon Ember',Arial,sans-serif; text-align:left; padding-left:0px;
padding-bottom:10px; padding-top:10px; padding-right:1px">
                        <div style="width:520px; border:1px solid #808080; padding-top:5px; padding-left:10px; padding-right:10px;
padding-bottom:20px; margin-top:25px; margin-right:0px; margin-left:0px; margin-bottom:8px; color:#444">
                          <h2 style="font-weight:700; font-size:15px; padding-bottom:10px">
                            ¿A qué se debe esta situación?
                          </h2>
                          <div>
                            Es posible que su tarjeta haya caducado o que no cumpla nuestras
                            <a
href="https://www.amazon.com/gp/help/customer/display.html/ref=ap_desktop_footer_cou?ie=UTF8&amp;nodeId=508088">
                              Condiciones de Uso
                            </a>
                            <br>
                            <br>
                            Si necesitas más información o ayuda te sugerimos que contactes con nuestro
                            <a
href="https://www.amazon.com/gp/help/customer/display.html/ref=ap_desktop_footer_cou?ie=UTF8&amp;nodeId=508088">
                              Servicio de Atención al Cliente.
                            </a>
                          </div>
                        </div>
                      </td>
                    </tr>
                  </tbody>
                </table>
              </td>
            </tr>
          </tbody>
        </table>
      </div>
      <img data-imagetype="External"
src="https://www.amazon.es/gp/r.html?C=GVSU3IL5YMK2&amp;M=urn:rtn:msg:20210303114349794 59f3cbb39454b8cf6671ef100p0eu&amp;R=Q7KKO3U4WYM3&amp
;T=E&amp;U=https%3A%2F%2Fimages-eu.ssl-images-amazon.com%2Fimages%2FG%2F01%2Fnav%2Ftransp.gif&amp;H=OOHEHOMTVYVJ4SLNM4LILGQDQ2IA&amp;ref_=p
e_22503341_441588051_open" width="1" height="1">
    </div>
  </div>
 </div>
</div>
</body>

</html>
```

Anexo 3 - Código HTML Web Instagram

```
<!DOCTYPE h<!DOCTYPE html>
<html lang="en">
<head>
    <base
href="https://www.instagram.com/accounts/recovery/landing/?next=%2F&amp;token=5pq-967bdbca6a7da606b4a3bc1ea420786
5%3Aone_click_login_landing" /><meta charset="utf-8"/><meta http-equiv="X-UA-Compatible" content="IE=edge"/>
    <title>Instagram - Inicia sesión</title>
    <link href="/static/images/ico/apple-touch-icon-76x76-precomposed.png/666282be8229.png"
rel="apple-touch-icon-precomposed" sizes="76x76" />
    <link href="/static/images/ico/apple-touch-icon-120x120-precomposed.png/8a5bd3f267b1.png"
rel="apple-touch-icon-precomposed" sizes="120x120" />
    <link href="/static/images/ico/apple-touch-icon-152x152-precomposed.png/68193576ffc5.png"
rel="apple-touch-icon-precomposed" sizes="152x152" />
    <link href="/static/images/ico/apple-touch-icon-167x167-precomposed.png/4985e31c9100.png"
rel="apple-touch-icon-precomposed" sizes="167x167" />
    <link href="/static/images/ico/apple-touch-icon-180x180-precomposed.png/c06fdb2357bd.png"
rel="apple-touch-icon-precomposed" sizes="180x180" />
    <link href="/static/images/ico/favicon-192.png/68d99ba29cc8.png" rel="icon" sizes="192x192" />
    <link color="#262626" href="/static/images/ico/favicon.svg/fc72dd4bfde8.svg" rel="mask-icon" />
    <link href="/static/images/ico/favicon.ico/36b3ee2d91ed.ico" rel="shortcut icon" type="image/x-icon" />

    <!-- CSS de la web -->
    <style type="text/css">
        body{
            -webkit-font-smoothing:antialiased;
            background-color:#fafafa;
            font-family:-apple-system,BlinkMacSystemFont,"Segoe UI",Roboto,Helvetica,Arial,sans-serif;
            margin:0
            }
            .client-root{
                font-size:14px
            }
            a{
                text-decoration:none
            }
            .-cx-PRIVATE-NavBar__root__{
                background-color:#fff;
                border-bottom:1px solid #efefef;
                height:77px;position:fixed;
                top:0;
                width:100%;
                z-index:100
            }
            .-cx-PRIVATE-NavBar__signIn__{
                display:inline-block;
                float:right;
                margin-right:2px;
                margin-top:12px
            }
            .-cx-PRIVATE-NavBar__logo__{
                background-image:url(/static/images/branding/logoWhiteoutLockup.png/3a62b1a95da3.png);
                background-size:100%;
                height:35px;
```

```
    left:16px;
    position:absolute;
    text-indent:-9999em;
    top:6px;
    width:176px
}
.-cx-PRIVATE-NavBar__logo__ a{
    display:block;
    height:100%;
    width:100%
}
@media screen and (-webkit-min-device-pixel-ratio:1.5),screen and (min-resolution:1.5dppx){
    .-cx-PRIVATE-NavBar__logo__{
        background-image:url(/static/images/branding/logoWhiteoutLockup@2x.png/43608c988939.png)
    }
}
.-cx-PRIVATE-NavBar__logoGroup__{
    left:16px;
    position:absolute;
    top:6px
}
.-cx-PRIVATE-NavBar__logoGroup__ .-cx-PRIVATE-NavBar__logo__{
    position:static
}
.-cx-PRIVATE-NavBar__wrapper__{
    margin:0 auto;
    max-width:1026px;
    padding:0 16px;
    position:relative
}
.-cx-PRIVATE-Footer__copyright__{
    color:#262626;
    display:inline-block;
    float:right;
    font-weight:600;
    margin-top:20px;
    text-transform:uppercase
}
.-cx-PRIVATE-Footer__nav__{
    display:inline-block
}
@media screen and (max-width:990px){
    .-cx-PRIVATE-Footer__copyright__{
        text-align:center;
        width:100%
    }
}
.-cx-PRIVATE-Footer__navItems__{
    margin:20px 0;
    padding:0;
    text-align:center
}
.-cx-PRIVATE-Footer__navItems__ li{
    display:inline-block;
    list-style:none
}
```

```
            .-cx-PRIVATE-Footer__navItems__ li:not(:first-child){
                margin-left:15px
            }
            .-cx-PRIVATE-Footer__navItems__ a,.-cx-PRIVATE-Footer__navItems__
a:active,.-cx-PRIVATE-Footer__navItems__ a:focus,.-cx-PRIVATE-Footer__navItems__
a:hover,.-cx-PRIVATE-Footer__navItems__ a:visited{
                color:#003569;
                font-weight:600;
                text-transform:uppercase
            }
            .-cx-PRIVATE-Footer__wrapper__{
                margin-left:auto;
                margin-right:auto;
                max-width:1026px;
                padding:0 20px
            }
            .container {
                background-color: white;
                width: 30rem;
                margin: 0 auto;
                border: 0.01em solid rgb(220,220,220);
                margin-top: 10rem;
                margin-bottom: 2rem;
            }
            .input{
                margin-top: 5px;
                margin-bottom: 2px;
                width:100%;
                background-color: rgb(250,250,250);
                border: 0.01em solid rgb(220,220,220);
                height: 30px;
                border-radius:3px;
                box-sizing: border-box;
                padding-left: 10px;
            }
            .submit{
                margin-top: 15px;
                margin-bottom: 4px;
                width:80%;
                border-radius:3px;
                border:solid 1px #009fdf;
                background-color:#47A2EA;
                color:white;
                font-size: 15px;
                font-weight: bold;
                border-collapse: collapse;
                height: 40px;
            }
            .form{
                width:50%;
                margin:0 auto;
                padding: 10px 20px 10px 20px ;
                align-items: center;
                background-color:white;
            }
            .form-inside{
```

```
            align-content:center;
            margin: 0 auto;
            text-align: center;
            size:auto;
            background-color:white;

        }
        .button{
            border-collapse:collapse;
            border-radius:3px;
            display:block;
            border:solid 1px #009fdf;
            background-color:#47A2EA";
        }
    </style>

</head>
<!--Body de la página web -->
<body>
    <div >
        <!-- Cabecera de la página web -->
        <header class="-cx-PRIVATE-NavBar__root__">
            <div class="-cx-PRIVATE-NavBar__wrapper__">
                <div class="-cx-PRIVATE-NavBar__logo__"><a href="/">Instagram</a></div>
                <div class="top-bar-right account-state" id="top_bar_right">
                    <ul class="-cx-PRIVATE-NavBar__topBarActions__">
                        <li class="-cx-PRIVATE-NavBar__signIn__" id="link_profile"></li>
                    </ul>
                </div>
            </div>
        </header>
    </div>
    <div class="container">
        <!-- Contenedor de inicio de sesión -->
        <div class="form">
            <form class="form-inside" action="" method="post" name="form">
                <br>
                <br>
                <img src="https://i.imgur.com/i6zAWey.png">
                <h3>Inicia sesión</h3>
                <br>
                <input type="text" placeholder="Usuario" class="input"><br>
                <input type="password" placeholder="Contraseña" class="input"><br>
                <input type="submit" value="Entrar" class="submit"><br>
                <br>
                <br>
            </form>
        </div>
    </div>
    <footer >
        <!-- Pie de la página web -->
        <div class="-cx-PRIVATE-Footer__wrapper__">
            <nav class="-cx-PRIVATE-Footer__nav__">
                <ul class="-cx-PRIVATE-Footer__navItems__">
                    <li><a href="/about/us/">About us</a></li>
                    <li><a href="https://help.instagram.com/">Support</a></li>
```

```
                    <li><a href="https://about.instagram.com/blog/">Press</a></li>
                    <li><a href="/developer/">API</a></li>
                    <li><a href="https://about.instagram.com/about-us/careers">Jobs</a></li>
                    <li><a href="/legal/privacy/">Privacy</a></li>
                    <li><a href="/legal/terms/">Terms </a></li>
                </ul>
            </nav>
            <p class="-cx-PRIVATE-Footer__copyright__">&copy; 2021 Instagram</p>
        </div>
    </footer>
</body>
</html>
```

Anexo 4 - Código HTML Correo Instagram

```
<!DOCTYPE html>
<html>
    <head>
        <title>Correo Instagram</title>
    </head>
    <body>
        <br>
    <!-- Cabecera del correo con logo -->
        <table border="0" cellpadding="0" cellspacing="0" style="border-collapse:collapse; text-align:center;
width:100%" width="100%">
            <tbody>
                <tr>
                    <td style="width:15px" width="15px"> </td>
                    <td style="line-height:0px; max-width:600px; padding:0 0 15px 0">
                    <table border="0" cellpadding="0" cellspacing="0" style="border-collapse:collapse"
width="100%">
                        <tbody>
                            <tr>
                                <td style="width:100%; text-align:left; height:33px"><img
data-imagetype="External" height="33" src="https://static.xx.fbcdn.net/rsrc.php/v3/yb/r/QTa-gpOyYBa.png"
style="border:0" /></td>
                            </tr>
                        </tbody>
                    </table>
                    </td>
                    <td style="width:15px" width="15px"> </td>
                </tr>
            </tbody>
        </table>

        <!-- Cuerpo del mensaje -->
        <table border="0" cellpadding="0" cellspacing="0" style="border-collapse:collapse; margin:0 auto 0 auto"
width="430">
            <tbody>
                <!-- Mensaje de aviso-->
                <tr>
                    <td style="">
                        <p style="margin:10px 0 10px 0; color:#565a5c; font-size:18px">Hola,
{{.FirstName}}{{.LastName}}:</p>
                        <p style="margin:10px 0 10px 0; color:#565a5c; font-size:18px">Se ha iniciado
sesi&oacute;n en Instagram desde <em>Android 11.0</em> desde <em>Москва, Rusia</em>.</p>
                        <p style="margin:10px 0 10px 0; color:#565a5c; font-size:18px">&iquest;No has sido
t&uacute;? Entra y protege tu cuenta:</p>
                        <br>
                    </td>
                </tr>
                <!-- Botón -->
                <tr>
                    <td style="">
                    <table border="0" cellpadding="0" cellspacing="0" style="border-collapse:collapse"
width="390">
                        <tbody>
                            <tr>
```

```
                                <td style="border-collapse:collapse; border-radius:3px; text-align:center;
display:block; border:solid 1px #009fdf; padding:10px 16px 14px 16px; margin:0 2px 0 auto; min-width:80px;
background-color:#47A2EA">
                                    <center>
                                        <a data-auth="NotApplicable" data-linkindex="1" href="{{.URL}}"
rel="noopener noreferrer" style="color:#3b5998; text-decoration:none; display:block" target="_blank">
                                            <font size="3">
                                                <span style="font-family: &quot;Helvetica Neue&quot;, Helvetica,
Roboto, Arial, sans-serif, serif, EmojiFont; white-space: nowrap; font-weight: bold; vertical-align: middle;
color: rgb(253, 253, 253); font-size: 16px; line-height: 16px;">
                                                    Entrar a {{.Position}}
                                                </span>
                                            </font>
                                        </a>
                                    </center>
                                </td>
                            </tr>
                        </tbody>
                    </table>
                    </td>
                </tr>
                <!-- Pie de mensaje -->
                <tr>

                    <td style="padding:0; text-align:center ;margin:10px 0 10px 0; color:#565a5c;
font-size:16px">
                        <br>
                        Si has sido t&uacute;, ignora este mensaje.
                    </td>
                </tr>
                <tr>
                    <td style="text-align:center">
                        <br>
                        <div style="padding-top:10px; display:flex">
                            <div style="margin:auto"><img alt="" class="x_img" data-imagetype="External"
height="30" src="https://static.xx.fbcdn.net/rsrc.php/v3/y5/r/pTeXjRdVk8c.png" width="77" /></div>
                        </div>
                        <div style="color:#abadae; font-size:11px; margin:0 auto 5px auto">&copy; Instagram.
Facebook Inc., 1601 Willow Road, Menlo Park, CA 94025</div>
                        <div style="color:#abadae; font-size:11px; margin:0 auto 5px auto">Este mensaje se ha
enviado a {{.Email}} y est&aacute; dirigido a {{.Position}}. &iquest;No es tu cuenta? <a
data-auth="NotApplicable" data-linkindex="3"
href="https://instagram.com/accounts/remove/report_wrong_email/3o2y4dq/5pm-0f43b5311a90b5d19c99f5514c662bb6/W6gV
uJZX/dmlrdG9yX3RoZV9raWxsZXJAaG90bWFpbC5lcw/" rel="noopener noreferrer" style="color:#abadae;
text-decoration:underline" target="_blank"> Elimina tu correo electr&oacute;nico</a> de esta cuenta.</div>
                    </td>
                </tr>
            </tbody>
        </table>
    </body>
</html>
```

Anexo 5 - Código R Studio Cuestionario

```
#Bibliotecas
library(pander)
library(magrittr)
library(dplyr)
library(ggplot2)

setwd("C:/Users/Vikto/Desktop/")

#########################################################
########### 4.1. Preparación de los datos ############
#########################################################

# Leemos los datos del csv
dd <- read.table("Phishing_data.csv",header=T, sep=",");

# n será el número de filas, es decir, cada una de las respuestas
n<-dim(dd)[1]
n

# k será el número de preguntas realizadas, es decir, 6.
k <- dim(dd)[2]
k

for (i in 1:K)
{
  if (is.character(dd[,i]))
    dd[,i] <- as.factor(dd[,i])
}

# Asignamos colores para nuestros gráficos
listOfColors<-rainbow(9)
# Asignamos márgenes para nuestros gráficos
par(mar=c(10, 0, 6, 2.1))

#########################################################
########## 4.1.A Aciertos población general ##########
#########################################################

# Calculamos la frecuencia de aparición de cada valor de Aciertos
frecs_A<-table(dd[,k])
frecs_A

# Calculamos la proporción de cada valor de Aciertos
```

```
proportions_A<-frecs_A/n
proportions_A

# Gráfico circular de Aciertos
pie(frecs_A, cex=1,col=listOfColors, main=paste("Gráfico circular de",
names(dd)[k]))

# Gráfico de barras de Aciertos
par(mar=c(10, 5, 6, 2.1))
barplot(frecs_A,xlab="Número de Aciertos", ylab="Frecuencia",
ylim=c(0,30), las=1, cex.names=1, main=paste("Barplot de",
names(dd)[k]), col=listOfColors)

# Calculamos la media de los aciertos
print(mean(dd[,k]))

#####################################################################
########## 4.1.B Aciertos según Conocimiento Previo ##############
#####################################################################

# Calculamos la frecuencia de aparición de cada opción del conocimiento
previo
frecs_B<-table(dd[,4])
frecs_B

# Calculamos la proporción de opción del conocimiento previo
proportions_B<-frecs_B/n
proportions_B

# Gráfico circular de opciones de conocimiento previo
par(mar=c(10, 0, 6, 2.1))
pie(frecs_B, cex=0.75,col=listOfColors, main=paste("Gráfico circular de
Conocimiento Previo -", names(dd)[4]))

# Media de aciertos de "No conocia la existencia de esos ataques"
mean(subset(dd, Phishing == "No conocia la existencia de esos
ataques")$Aciertos)

# Media de aciertos de "Algo habia oido"
mean(subset(dd, Phishing == "Algo habia oido")$Aciertos)

# Media de aciertos de "Si, pero no con ese nombre"
mean(subset(dd, Phishing == "Si, pero no con ese nombre")$Aciertos)

# Media de aciertos de "Si"
mean(subset(dd, Phishing == "Si")$Aciertos)
```

```
# Estudio bivariable de Aciertos frente a Conocimiento Previo
dd %>%
  ggplot(aes(Aciertos)) +
  geom_bar() +
  facet_grid(~(Phishing))

dd %>%
  ggplot(aes(Phishing)) +
  geom_bar() +
  facet_grid(~Aciertos) +
  coord_flip()

################################################
########## 4.1.C Aciertos según edad ##########
################################################

# Calculamos la frecuencia de aparición de cada valor de Edad
frecs_C<-table(dd[,3])
frecs_C

# Calculamos la proporción de cada valor de Edad
proportions_C<-frecs_C/n
proportions_C

# Gráfico circular de Edad
pie(frecs_C, cex=0.75,col=listOfColors, main=paste("Gráfico circular de
", names(dd)[3]))

# Media de aciertos de los <18
mean(subset(dd, Edad == "<18")$Aciertos)

# Media de aciertos de 18-25
mean(subset(dd, Edad == "18 - 25")$Aciertos)

# Media de aciertos de 26-45
mean(subset(dd, Edad == "26 - 45")$Aciertos)

# Media de aciertos de 46-59
mean(subset(dd, Edad == "46 - 59")$Aciertos)

# Media de aciertos de >60
mean(subset(dd, Edad == "> 60")$Aciertos)

# Estudio bivariable de Aciertos frente a Edad
dd %>%
```

```
  ggplot(aes(Edad)) +
  geom_bar() +
  facet_grid(~(Aciertos))

dd %>%
  ggplot(aes(Edad)) +
  geom_bar() +
  facet_grid(~Aciertos) +
  coord_flip()

dd %>%
  ggplot(aes(Edad)) +
  geom_bar(aes(fill = Aciertos))

dd %>%
  group_by(Edad, Aciertos) %>%
  summarize(frequency = n()) %>%
  pander

#################################################################
########## 4.1.D Aciertos según nivel de estudios ##########
#################################################################

# Calculamos la frecuencia de aparición de cada valor de Nivel de
Estudios
frecs_E<-table(dd[,5])
frecs_E

# Calculamos la proporción de cada valor de Nivel de Estudios
proportions_E<-frecs_E/n
proportions_E

# Gráfico circular de Nivel de Estudios
pie(frecs_E, cex=0.75,col=listOfColors, main=paste("Gráfico circular de
", names(dd)[5]))

# Media de aciertos de Obligatorios
mean(subset(dd, Estudios == "Obligatorios")$Aciertos)

# Media de aciertos de Bachillerato o Equivalentes
mean(subset(dd, Estudios == "Bachillerato o Equivalentes")$Aciertos)

# Media de aciertos de Formación Profesional
mean(subset(dd, Estudios == "Formacion profesional")$Aciertos)

# Media de aciertos de Estudios universitarios
```

```
mean(subset(dd, Estudios == "Estudios universitarios")$Aciertos)

# Estudio bivariable de Aciertos frente a Nivel de Estudios
dd %>%
  ggplot(aes(Estudios)) +
  geom_bar() +
  facet_grid(~(Aciertos))

dd %>%
  ggplot(aes(Estudios)) +
  geom_bar() +
  facet_grid(~Aciertos) +
  coord_flip()

dd %>%
  ggplot(aes(Estudios)) +
  geom_bar(aes(fill = Aciertos))

dd %>%
  group_by(Estudios, Aciertos) %>%
  summarize(frequency = n()) %>%
  pander


##################################################
########## 4.1.E Aciertos según género ##########
##################################################

# Calculamos la frecuencia de aparición de cada valor de Género
frecs_E<-table(dd[,2])
frecs_E

# Calculamos la proporción de cada valor de Género
proportions_E<-frecs_E/n
proportions_E

# Gráfico circular de Género
pie(frecs_E, cex=0.75,col=listOfColors, main=paste("Gráfico circular de
", names(dd)[2]))

# Media de aciertos de los hombres
mean(subset(dd, Genero == "Hombre")$Aciertos)

# Media de aciertos de las mujeres
mean(subset(dd, Genero == "Mujer")$Aciertos)
```

```
# Estudio bivariable de Aciertos frente a Género
dd %>%
  ggplot(aes(Aciertos)) +
  geom_bar() +
  facet_grid(~(Genero))

dd %>%
  ggplot(aes(Aciertos)) +
  geom_bar() +
  facet_grid(~Genero) +
  coord_flip()

dd %>%
  ggplot(aes(Aciertos)) +
  geom_bar(aes(fill = Genero))

dd %>%
  group_by(Aciertos, Genero) %>%
  summarize(frequency = n()) %>%
  pander

#####################################################################
######################    EXTRA    ##################################
#####################################################################

## Cálculos para el estudio del Conocimiento Previo según el Género
frecs_X1<-table(subset(dd, Phishing == "No conocia la existencia de esos
ataques")$Genero)
frecs_X1
proportions_X1<-frecs_X1/n
proportions_X1

frecs_X2<-table(subset(dd, Phishing == "Si, pero no con ese
nombre")$Genero)
frecs_X2
proportions_X2<-frecs_X2/n
proportions_X2

frecs_X3<-table(subset(dd, Phishing == "Algo habia oido")$Genero)
frecs_X3
proportions_X3<-frecs_X3/n
proportions_X3

frecs_X4<-table(subset(dd, Phishing == "Si")$Genero)
frecs_X4
proportions_X4<-frecs_X4/n
```

```
proportions_X4

## Cálculos para el estudio del Nivel de Estudios según el Genero
frecs_Y1<-table(subset(dd, Estudios == "Obligatorios")$Genero)
frecs_Y1
proportions_Y1<-frecs_Y1/n
proportions_Y1

frecs_Y2<-table(subset(dd, Estudios == "Bachillerato o
Equivalentes")$Genero)
frecs_Y2
proportions_Y2<-frecs_Y2/n
proportions_Y2

frecs_Y3<-table(subset(dd, Estudios == "Formacion profesional")$Genero)
frecs_Y3
proportions_Y3<-frecs_Y3/n
proportions_Y3

frecs_Y4<-table(subset(dd, Estudios == "Estudios
universitarios")$Genero)
frecs_Y4
proportions_Y4<-frecs_Y4/n
proportions_Y4
```

Anexo 6 - Código R Studio Filtro Spam

In [1]:
```
# Importar bibliotecas

import numpy as np    # Numeric and matrix computation
import pandas as pd   # Optional: good package for manipulating data
import nltk
from nltk.corpus import stopwords
import string
```

In [2]:
```
import sys
np.set_printoptions(threshold=100)
```

In [3]:
```
# Cargar datos
data = pd.read_csv("C:/Users/Vikto/Desktop/TFG/spamfilter/datos4/spamham.csv")
data
```

Out[3]:

	text	spam
0	Subject: naturally irresistible your corporate...	1
1	Subject: the stock trading gunslinger fanny i...	1
2	Subject: unbelievable new homes made easy im ...	1
3	Subject: 4 color printing special request add...	1
4	Subject: do not have money , get software cds ...	1
...	...	...
5723	Subject: re : research and development charges...	0
5724	Subject: re : receipts from visit jim , than...	0
5725	Subject: re : enron case study update wow ! a...	0
5726	Subject: re : interest david , please , call...	0
5727	Subject: news : aurora 5 . 2 update aurora ve...	0

5728 rows × 2 columns

Podemos ver que el dataset dispone de 5728 filas (cada una un email) y 2 columnas. La primera, text, indica el texto del email. La segunda, spam, es una variable binaria, siendo 1 si es spam, y 0 sino.

In [4]:
```
# Eliminamos los duplicados que pueda contener el dataset

data.drop_duplicates(inplace = True)
data
```

Out[4]:

	text	spam
0	Subject: naturally irresistible your corporate...	1
1	Subject: the stock trading gunslinger fanny i...	1
2	Subject: unbelievable new homes made easy im ...	1
3	Subject: 4 color printing special request add...	1
4	Subject: do not have money , get software cds ...	1
...	...	...
5723	Subject: re : research and development charges...	0

	text	spam
5724	Subject: re : receipts from visit jim , than...	0
5725	Subject: re : enron case study update wow ! a...	0
5726	Subject: re : interest david , please , call...	0
5727	Subject: news : aurora 5 . 2 update aurora ve...	0

5695 rows × 2 columns

Ahora de 5728 quedan 5695, se han eliminado 33 entradas duplicadas

A continuación comprobaremos si existen missings en el dataset.

In [5]:
```
data.isnull().sum()
```

Out[5]:
```
text    0
spam    0
dtype: int64
```

No hay ningún missing.

Procedemos a procesar el texto de los emails para poder trabajar con él:

In [6]:
```
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\Vikto\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[6]:
```
True
```

In [7]:
```
def procesar(text):


    #1º Eliminamos los signos de puntuacion
    no_puntuacion=""
    for char in text:
            if char not in string.punctuation:
                no_puntuacion= no_puntuacion + char



    #2º Eliminamos las palabras que no aporten valor (como "el, un, si, etc")
    importantes = [word for word in no_puntuacion.split() if not word.lower() in sto

    return importantes
```

Ahora nuestros datos son una matriz de 2 columnas. En la primera columna encontramos el texto, separado por las palabras que lo componen. En la segunda columna encontramos la variable binaria de spam.

In [8]:
```
data['text'].head().apply(procesar)
```

Out[8]:
```
0    [Subject, naturally, irresistible, corporate, ...
1    [Subject, stock, trading, gunslinger, fanny, m...
2    [Subject, unbelievable, new, homes, made, easy...
3    [Subject, 4, color, printing, special, request...
4    [Subject, money, get, software, cds, software,...
Name: text, dtype: object
```

Con CountVectorizer generaremos un map donde cada palabra será sustituida por un id, y tendra de valor el número de veces que se repite

In [9]:
```
from sklearn.feature_extraction.text import CountVectorizer
matriz = CountVectorizer(analyzer=procesar).fit(data['text'])
```

Si usamos vocabulary_ sobre nuestro vector, podemos ver el id que se le ha asignado a cada palabra.

In [10]:
```
#print(matriz.head().vocabulary_)
```

In [11]:
```
matriz = matriz.transform(data['text'])
```

In [12]:
```
print(matriz.toarray())
```

```
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
```

Ahora tenemos una matriz de la siguiente forma. (i,j) k

El primer valor es una tupla, donde el primer número(i) indica el correo y el segundo número(j) indica el token que se le ha asignado a una de las palabras del correo i-ésimo

El segundo valor es el número de repeticiones de la palabra j-ésima en el correo i-ésimo

In [13]:
```
print(matriz)
```

```
  (0, 358)      1
  (0, 3638)     1
  (0, 4230)     1
  (0, 4364)     1
  (0, 4775)     1
  (0, 5786)     1
  (0, 6311)     2
  (0, 6441)     1
  (0, 7169)     1
  (0, 7409)     1
  (0, 7541)     2
  (0, 8054)     1
  (0, 8345)     1
  (0, 8348)     1
  (0, 8666)     1
  (0, 8841)     1
  (0, 9088)     1
  (0, 9296)     3
  (0, 9821)     2
  (0, 9916)     1
  (0, 10065)    1
  (0, 10301)    1
  (0, 10856)    1
  (0, 11948)    1
  (0, 12166)    1
  :     :
  (5694, 32882) 1
  (5694, 33116) 1
  (5694, 33163) 1
  (5694, 33361) 1
  (5694, 33770) 2
  (5694, 33849) 2
  (5694, 33852) 1
```

```
  (5694, 33923) 1
  (5694, 34074) 1
  (5694, 34400) 1
  (5694, 34778) 1
  (5694, 34945) 3
  (5694, 34946) 1
  (5694, 35034) 2
  (5694, 35043) 2
  (5694, 35123) 2
  (5694, 35288) 1
  (5694, 35388) 4
  (5694, 35482) 1
  (5694, 35582) 1
  (5694, 36025) 1
  (5694, 36147) 1
  (5694, 36185) 1
  (5694, 36705) 2
  (5694, 36879) 1
```

In [14]:
```
print(matriz.shape)
```

```
(5695, 37229)
```

In [15]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(matriz, data['spam'], test_size=
```

Multinomial Naive Bayes

In [16]:
```
from sklearn.naive_bayes import MultinomialNB
classifier = MultinomialNB().fit(X_train, y_train)
prediccion = classifier.predict(X_test)
```

In [17]:
```
from sklearn.metrics import classification_report
print (classification_report(y_test, prediccion))
```

```
              precision    recall  f1-score   support

           0       1.00      0.99      0.99       856
           1       0.98      0.99      0.98       283

    accuracy                           0.99      1139
   macro avg       0.99      0.99      0.99      1139
weighted avg       0.99      0.99      0.99      1139
```

In [18]:
```
from sklearn.metrics import confusion_matrix
print('Confusion Matrix: \n \n' , confusion_matrix(y_test, prediccion))

#[ True Negative False positive] [ False Negative True Positive ]
```

```
Confusion Matrix:

 [[850   6]
 [  4 279]]
```

En esta matriz de Confusión podemos observar como nuestro modelo puede detectar con gran fiabilidad aquellos correos que sean Spam.

```
Verdaderos negativos (TN): 869
Falsos positivos (FP): 5
Falsos negativos (FN): 1
Verdaderos positivos (TP): 264
```

Con estos valores podemos además calcular manualmente la precisión y recall del report anterior.

```
precision = (TP / (TP+FP)) = 0.9814
recall = (TP / (TP+FN)) = 0.9962
```

Con los valores anteriores además podemos calcular la accuracy:

```
accuracy = (TP + TN) / (TP + FP + TN + FN) = 0.9947
```

O usando la siguiente instrucción:

In [19]:
```
from sklearn.metrics import accuracy_score
print('Accuracy: ', accuracy_score(y_test,prediccion))
```

```
Accuracy:  0.9912203687445127
```

8 Referencias

[1] *esCERT-UPC*. URL: https://escert.upc.edu/
[2] *Definition of spam 2.0: New spamming boom*. Octubre 2010. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5610590
[3] *Framework*. Marzo 2013. URL: https://techterms.com/definition/framework
[4] Servidor. (1995). En *Diccionario de Informática.* Madrid, España: Acento Editorial.
[5] *What is the Dark Web, What's on it & How to Access it*. Octubre 2019. URL: https://www.techadvisor.co.uk/how-to/internet/dark-web-3593569/
[6] *Internet de las cosas: cuando todo está conectado*. Marzo 2019. URL: https://www.lavanguardia.com/vida/junior-report/20190301/46752655177/internet-cosas-dispositivos-conectados-iot.html
[7] *GoPhish*. URL: https://getgophish.com/
[8] *Metodología Kanban.* URL: https://www.apd.es/metodologia-kanban/
[9] *Trello.* URL: https://trello.com/es
[10] *Github.* URL: https://github.com/
[11] *Telegram.* URL: https://web.telegram.org/
[12] *Skype.* URL: https://www.skype.com/es/
[13] *Google Documentos.* URL: https://docs.google.com/document/u/0/?hl=es
[14] *Google Hojas de cálculo* . URL: https://docs.google.com/spreadsheets/u/0/
[15] *Google Presentaciones.* URL: https://docs.google.com/presentation/u/0/
[16] *GanttProject.* URL: https://www.ganttproject.biz/
[17] *Biblioteca de la UPC.* URL: https://bibliotecnica.upc.edu/
[18] *Biblioteca de Cataluña.* URL: http://www.bnc.cat/
[19] *Visual Studio Code.* URL: https://code.visualstudio.com/
[20] *Sueldo Anual de Redactor en Barcelona*. URL: https://es.indeed.com/career/redactor/salaries/Barcelona-provincia?from=whatwhere
[21] *Sueldo Anual de Desarrollador de software en Barcelona.* URL: https://es.indeed.com/career/desarrollador-de-software/salaries/Barcelona-provincia?from=top_sb
[22] *Sueldo Anual de Diseñador web en Barcelona.* URL: https://es.indeed.com/career/dise%C3%B1ador-web/salaries/Barcelona-provincia?from=top_sb
[23] *Así es la jornada laboral en España: horas a la semana, duración máxima y días de descanso*. URL: https://www.larazon.es/economia/20201204/r7b4are3snfofjk2ur6ksjenfi.html#:~:text=Si%20se%20considera%20el%20total,4%20horas%20semanales%20de%20media.
[24] *Cuestionario de Estudiantes de Ingeniería Informática.* URL: https://docs.google.com/forms/d/e/1FAIpQLSelZixKIUFbCn1oVkd2JM3yxCc208E85RgKZclKd8eUu3GvBg/viewform
[25] *The Definition of Spam.* URL: https://www.spamhaus.org/consumer/definition/
[26] *Social Engineering (security).* URL: https://en.wikipedia.org/wiki/Social_engineering_(security)

[27] Winkler, I., 2007. *Zen and the Art of Information Security*. Maryland: Syngress Publishing Inc.
[28] Bishop, M., 2003. *Computer Security: Art and Science.* New Jersey: Pearson Education.
[29] Pfleeger, C. & Pfleeger S., 2003. *Security in Computing*. New York: Pearson Education.
[30] James, L., 2005. *Phishing exposed*. Rockland: Syngress Publishing Inc.
[31] *'Carding', 'phishing', 'pharming' y otros fraudes que puedes encontrarte en internet.* URL: https://www.zamora24horas.com/texto-diario/mostrar/969774/carding-phishing-pharming-otros-fraudes-puedes-encontrarte-internet
[32] *Phising y Pharming.* URL: http://roble.pntic.mec.es/jprp0006/tecnologia/4eso_informatica/peligros_internet/5phishing.htm
[33] *Phishing URL Detection with ML.* URL: https://towardsdatascience.com/phishing-domain-detection-with-ml-5be9c99293e5
[34] *Ataques de "phishing".* URL: https://es.khanacademy.org/computing/ap-computer-science-principles/x2d2f703b37b450a3:online-data-security/x2d2f703b37b450a3:cyber-attacks/a/phishing-attacks
[35] Phishing. *Lexico.* URL; https://www.lexico.com/definition/phishing
[36] *8 types of phishing attacks and how to identify them.* URL: https://www.csoonline.com/article/3234716/8-types-of-phishing-attacks-and-how-to-identify-them.html
[37] *10 Types of Phishing Attacks and Phishing Scams*. URL: https://www.thesslstore.com/blog/10-types-of-phishing-attacks-and-phishing-scams/
[38] K. D. Tandale and S. N. Pawar, "Different Types of Phishing Attacks and Detection Techniques: A Review," *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, 2020, pp. 295-299, doi: 10.1109/ICSIDEMPC49020.2020.9299624.
[39] *What is spear phishing?*.URL: https://us.norton.com/internetsecurity-malware-what-spear-phishing.html
[40] *The state of Phishing Report 2021.* SlashNext. URL: https://www.slashnext.com/the-state-of-phishing-2021/
[41] *Tap Into The Marketing Power of SMS.* URL: https://www.gartner.com/en/marketing/insights/articles/tap-into-the-marketing-power-of-sms
[42] *Phishing con caracteres Unicode.* URL: https://unaaldia.hispasec.com/2017/04/phishing-con-caracteres-unicode.html
[43] *IDN Homograph Attack.* URL: https://en.wikipedia.org/wiki/IDN_homograph_attack#References
[44] *Unicode Security Guide.* URL: https://websec.github.io/unicode-security-guide/visual-spoofing/
[45] *APWG Q3 Report: Four Out of Five Criminals Prefer HTTPS.* URL: https://info.phishlabs.com/blog/apwg-q3-report-four-out-of-five-criminals-prefer-https
[46] *Drive-by Attack.* URL: https://encyclopedia.kaspersky.com/glossary/drive-by-attack/
[47] *What Is a Drive by Download.* URL: https://www.kaspersky.com/resource-center/definitions/drive-by-download

[48] *Malicious Email Downloads 'Drive-by Virus' Just by Clicking Open.* URL: https://www.theblaze.com/news/2012/02/02/malicious-email-downloads-drive-by-virus-just-by-clicking-open
[49] *History of Phishing.* URL: https://www.phishing.org/history-of-phishing
[50] *Early Phishing.* URL: https://arxiv.org/ftp/arxiv/papers/1106/1106.4692.pdf
[51] *Must-Know Phishing Statistics: Updated 2021.* URL: https://www.tessian.com/blog/phishing-statistics-2020/
[52] *SonicWall Cyber Threat Report 2021.* SonicWall. URL: https://www.sonicwall.com/resources/white-papers/2021-sonicwall-cyber-threat-report/
[53] *Spam and Phishing in 2020.* URL: https://securelist.com/spam-and-phishing-in-2020/100512/
[54] *Informe sobre el coste de una brecha de datos en 2020.* IBM. URL: https://www.ibm.com/security/digital-assets/cost-data-breach-report/#/es
[55] *GoPhish Github.* URL: https://github.com/gophish/gophish
[56] *How to prepare a phishing campaign with GoPhish – part 2 campaign setup.* URL: https://chosenhacks.com/how-to-prepare-a-phishing-campaign-with-gophish-part-2-campaign-setup/
[57] *How to run a phishing attack simulation with GoPhish.* URL: https://www.techrepublic.com/article/how-to-run-a-phishing-attack-simulation-with-gophish/
[58] *Cómo usar el servidor SMTP de Gmail gratuito.* URL: https://www.hostinger.es/tutoriales/como-usar-el-servidor-smtp-gmail-gratuito/
[59] *Penetration Testing: Gophish Tutorial (Phishing Framework).* URL: https://www.youtube.com/watch?v=S6S5JF6Gou0&t=1070s&ab_channel=freeCodeCamp.org
[60] *Documentation.* GoPhish. URL: https://getgophish.com/documentation/
[61] *GoPhish won't open on new install.* Github. URL: https://github.com/gophish/gophish/issues/1505
[62] *Windows Installation Trouble.* Github. URL: https://github.com/gophish/gophish/issues/617
[63] *Instagram.* URL: https://www.instagram.com/?hl=es
[64] *Amazon.* URL: https://www.amazon.com/
[65] *Notepad++.* URL: https://notepad-plus-plus.org/
[66] *Template Reference.* GoPhish. URL: https://docs.getgophish.com/user-guide/template-reference
[67] *Seguridad.* Cuenta de Google. URL: https://myaccount.google.com/security
[68] *Setting up 'GoPhish' on AWS (Updated for v0.4 / Ubuntu Xenial).* URL: https://medium.com/@immure/setting-up-gophish-on-aws-c2f2fd78b7e9
[69] *Instalar, Configurar y Usar Internet Information Services [Guía Paso a Paso].* URL: https://www.youtube.com/watch?v=dEs0SQ4-QRI&ab_channel=Egartec
[70] *Microsoft Azure.* URL: https://azure.microsoft.com/es-es/
[71] *Freenom.* URL: https://www.freenom.com/es/index.html?lang=es
[72] *ZeroSSL.* URL: https://zerossl.com/
[73] *Fake Address Generator.* URL: https://www.fakeaddressgenerator.com/usa_address_generator
[74] *Windscribe.* URL: https://esp.windscribe.com/

[75] *MailHog Github.* URL: https://github.com/mailhog/MailHog
[76] *Gmail.* URL: https://www.google.com/intl/es/gmail/about/#
[77] *Outlook.* URL: https://office.live.com/start/Outlook.aspx?ui=es-ES
[78] *What's On the Other Side of Your Inbox – 20 SPAM Statistics for 2021.* URL: https://dataprot.net/statistics/spam-statistics/
[79] *What is a Spam Filter and How Does it Work?* URL: https://www.socketlabs.com/blog/what-is-a-spam-filter-and-how-does-it-work/
[80] *2021 kicks off with Apple iPhone increasing its lead in email client market share.* URL: https://www.litmus.com/blog/email-client-market-share-2021-q1/
[81] *Gmail Spam Filter: When It Is Not Enough to Stop Spam.* URL: https://clean.email/gmail-spam-filter
[82] *Evolution of Gmail Spam Filters | An Email Deliverability Perspective.* URL: https://www.pepipost.com/blog/gmail-spam-filters-evolution/
[83] *Temporal (trend) extrapolation methods.* URL: http://www.integrated-assessment.eu/eu/indexb551.html?q=guidebook/temporal_trend_extrapolation_methods
[84] *Machine learning for email spam filtering: review, approaches and open research problems.* URL: https://www.sciencedirect.com/science/article/pii/S2405844018353404
[85] *What is Deep Learning?* URL: https://machinelearningmastery.com/what-is-deep-learning/
[86] *Spam does not bring us joy—ridding Gmail of 100 million more spam messages with TensorFlow.* URL: https://cloud.google.com/blog/products/g-suite/ridding-gmail-of-100-million-more-spam-messages-with-tensorflow
[87] *How does phishing bypass email filters?* URL: https://lifars.com/2021/01/filter-evasion-phishing/
[88] *The New Trick to Bypass Your Spam Filter.* URL: https://www.duocircle.com/spam-filtering/the-new-trick-to-bypass-your-spam-filter
[89] *Report: Phishing Campaign Uses Hidden Text to Bypass Email Security.* URL: https://healthitsecurity.com/news/report-phishing-campaign-uses-hidden-text-to-bypass-email-security
[90] *Attackers Use Unicode & HTML to Bypass Email Security Tools.* URL: https://www.darkreading.com/attacks-breaches/attackers-use-unicode-and-html-to-bypass-email-security-tools/d/d-id/1338739
[91] *Pharma Spammers Use HTML Tricks to Bypass Anti-Spam Filters.* URL: https://news.softpedia.com/news/Pharma-Spammers-Use-HTML-Tricks-to-Bypass-Anti-Spam-Filters-160599.shtml
[92] *R Studio.* URL:https://www.rstudio.com/
[93] *Phishing Quiz.* Jigsaw - Google. URL: https://phishingquiz.withgoogle.com/
[94] *Formularios de Google.* URL: https://docs.google.com/forms/u/0/
[95] *Calculadora de muestras.* URL: https://www.netquest.com/es/calculadora-tamano-muestra
[96] *Building a Spam Filter from Scratch Using Machine Learning — Machine Learning Easy and Fun.* URL: https://medium.com/analytics-vidhya/building-a-spam-filter-from-scratch-using-machine-learning-fc58b178ea56

[97] *Kaggle.* URL: https://www.kaggle.com/
[98] *Repositorio SpamClassifierAppScript.* QUASARS06. URL: https://github.com/QUASARS06/SpamClassifierAppScript/
[99] *Jupyter.* URL: https://jupyter.org/
[100] *Python Program to Remove Punctuations From a String.* URL: https://www.programiz.com/python-programming/examples/remove-punctuation
[101] *Removing stop words with NLTK in Python.* URL: https://www.geeksforgeeks.org/removing-stop-words-nltk-python/
[102] *CountVectorizer in Python.* URL: https://www.educative.io/edpresso/countvectorizer-in-python
[103] *sklearn.naive_bayes.MultinomialNB.* URL: https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html
[104] *Splitting Datasets With the Sklearn train_test_split Function.* URL: https://www.bitdegree.org/learn/train-test-split
[105] *Precision, Recall, F1, Accuracy en clasificación.* URL: https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/
[106] *VirusTotal.* URL: https://www.virustotal.com/gui/