Demostración práctica en el aula: utilización de comandos avanzados de búsqueda para analizar y detectar vulnerabilidades en sitios web

García, Marcelo Adrián[1,3], Mendoza, Carlos[2,4]

[1]Instituto de Administración - Facultad de Ciencias Económicas – U. N. Tucumán
[2]Instituto de Ingeniería y Agronomía - U. N. Arturo Jauretche

[3]mgarcia@face.unt.edu.ar, [4]cmendoza2015@gmail.com

Resumen. En el presente trabajo se realiza una recopilación bibliográfica a fin de reconocer los principales operadores avanzados de búsqueda utilizados para la detección de vulnerabilidades en sitios, plataformas y dispositivos publicados en la web, a través de la herramienta "Google Dorks".

Posteriormente, en el marco de la Asignatura "Seguridad y Control en Sistemas Informáticos" de la Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán, se efectuará una demostración práctica a los alumnos a través de una disertación a distancia. Se les mostrará a los asistentes la técnica objeto de estudio, mediante la búsqueda de vulnerabilidades en páginas web indexadas en Google, con el objetivo de complementar los contenidos abordados en la unidad de estudio "Ciberseguridad en el Comercio Electrónico".

Finalmente, se presentarán conclusiones, recomendaciones y buenas prácticas que los futuros graduados en Ciencias Económicas deberán tener en cuenta para asegurar las características de confidencialidad, integridad y disponibilidad de la información en plataformas y sitios publicadas en la web.

Palabras Claves: Ciberseguridad, Vulnerabilidades, Sitios Web

1 PLANTEO DEL PROBLEMA Y JUSTIFICACIÓN DE SU ABORDAJE

"Google Dorking" o "Google Hacking" es un mecanismo que utiliza la búsqueda avanzada del motor de búsqueda de Google para encontrar vulnerabilidades de seguridad en la configuración de los sitios publicados en la web, pudiéndose encontrar además "agujeros de seguridad" en su código (Wikipedia, s.f.).

A través de la utilización de esta herramienta es posible filtrar información y conseguir resultados normalmente no accesibles como, por ejemplo, nombre de usuario, contraseñas, páginas de inicio de sesión de administrador, datos sensibles, detalles de cuentas bancarias, entre otros. La técnica se basa en la utilización de operadores avanzados en el motor de búsqueda de Google, para encontrar cadenas específicas de texto en los

resultados. De este modo, se localizarán las páginas web que posean filtros, como ser título, texto principal, URL u otros.

Cabe destacar que cualquier persona puede rastrear el contenido de un sitio web a través de operadores de búsqueda avanzada, pues el motor de búsqueda, salvo que se lo bloquee de manera explícita, lo indexa automáticamente. Google es considerado como el más popular y poderoso buscador que actualmente existe, utilizando un robot que indexa gran cantidad de contenidos en la web y posibilita acceder a numerosas páginas (Tryhackme, 2021). Como consecuencia de ello, información sensible podría ser revelada sin que el propietario de esta se percate (Raggi, 2021).

Dada esta problemática, se plantea efectuar una demostración práctica en el aula a los alumnos de la asignatura "Seguridad y Control en Sistemas Informáticos" (UNT), con el objetivo de complementar los contenidos abordados en la unidad de estudio "Ciberseguridad en el Comercio Electrónico".

Se considera que esta actividad interdisciplinaria agrega considerable valor a la formación académica de los futuros egresados en Ciencias Económicas, quienes deberán estar preparados para gestionar organizaciones en contextos altamente informatizados e interconectados.

2 FORMULACIÓN DE OBJETIVOS

2.1 Objetivo General

Efectuar una demostración práctica en el aula sobre la utilización de comandos avanzados de búsqueda para el análisis y detección de vulnerabilidades en sitios web, a los alumnos de la asignatura "Seguridad y Control en Sistemas Informáticos" (UNT), con el objetivo de complementar los contenidos abordados en la unidad de estudio "Ciberseguridad en el Comercio Electrónico"

2.2 Objetivos específicos

- Realizar una recopilación bibliográfica a fin de reconocer los principales operadores avanzados de búsqueda utilizados para la detección y análisis de vulnerabilidades en sitios, plataformas y dispositivos publicados en la web, a través de la herramienta "Google Dorks".
- Organizar una actividad interdisciplinaria entre la Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán y el Instituto de Ingeniería y Agronomía de Universidad Nacional Arturo Jauretche, con el afán de agregar valor a la formación académica de los alumnos, en lo que respecto a la disciplina de Sistemas y Tecnologías de la Información en general y de Ciberseguridad en particular.
- Presentar recomendaciones y buenas prácticas que deben tener en cuenta los profesionales en Ciencias Económicas para asegurar las características de confidencialidad, integridad y disponibilidad de la información en plataformas y sitios web.

3 ENCUADRES TEÓRICO

Una gran cantidad de personas utiliza servicios de internet para almacenar información en distintos tipos de formato (documentos, fotos, etc.). Sin embargo, muchos de ellos no son conscientes respecto a que estos datos pueden quedar públicamente expuestos y ser utilizados con objetivos maliciosos.

En este sentido, se hace referencia al concepto de "inteligencia de fuentes abiertas" (OSINT por su sigla en inglés "Open Source INTelligence"), que es el conjunto de técnicas y herramientas para recopilar información pública, correlacionar datos y procesarlos. Es decir, "aplicar análisis e inteligencia a la gran cantidad de información públicamente accesible en Internet con el objetivo de extraer conclusiones útiles para una investigación, monitoreo, campaña de marketing, etc." (Pastorino, 2019). Pero para que ello sea posible, es importante definir previamente los selectores y las palabras claves que se utilizarán para obtener la información que se desea.

Operadores utilizados en búsquedas avanzadas en "Google Dorks".

Es necesario indicar en la "barra de búsqueda" los comandos que se desea encontrar de acuerdo con los criterios requeridos. Estos operadores avanzados pueden ser utilizados individualmente o combinados.

Las principales instrucciones avanzadas de búsqueda se detallan a continuación.:

- Intitle: indica el título buscado
- Inurl: URL del sitio web en donde se realiza la búsqueda
- "": busca una frase exacta
- And y not: operadores lógicos "y" o "no"
- + y -: incluye o excluye términos de una búsqueda
- * (asterisco): comodín, cualquier palabra (solo una)
- (punto): comodín, cualquier palabra (una o muchas)
- Site: busca resultados dentro de la web indicada
- Filetype: busca archivos de un formato determinado (doc, xlsx, txt)
- Link: busca sitios web que tienen un enlace a un determinado sitio web
- Inanchor: busca en sitios que poseen en el texto de enlace la expresión buscada
- Cache: arroja el resultado en el caché de Google de un sitio web
- Related: busca sitios web relacionados a una determinada

Posteriormente se describen los principales comandos de búsqueda que pueden utilizarse para acceder a información sensible o detectar vulnerabilidades en sitios web

1. Archivos con usuarios y contraseñas: permiten a quien realiza la búsqueda, ingresar en un sitio web.

- intitle:"index of" "Index of /" password.txt: permite visualizar en el índice del sitio web los servidores que contenga un archivo llamado password.txt.
- filetype:inc intext: mysql_connect password -please -could -port: muestra usuarios y contraseñas de bases de datos MySQL.

2. Páginas con formularios de acceso: permite realizar ataques de diccionario o fuerza bruta.

- Inurl:":10000" webmin: posibilita administrar remotamente un servidor Linux por el puerto 10000. Al ingresar se solicita un usuario y contraseña.
- Inurl:/admin/login.asp: se pueden obtener datos de ingreso a sitios de WordPress, que están contenidos en la carpeta /wp-admin o en /administrator, y configuradas por defecto.

3. Archivos con nombres de usuario o mensajes de error: posibilita atacar mediante una técnica de diccionario o fuerza bruta, debido a que se muestran los usuarios (a los cuales solo resta obtener su contraseña).

- "Access denied for user" "using password" "general error"
- inurl: phpbb "sql error": se muestran foros phpBB que arrojan errores.

4. Detección de la versión del servidor: si el servidor web o alguno de sus programas instalados no cuentan con la actualización de su última versión, seguramente presentan vulnerabilidades. A partir de ello, puede buscarse un exploit que permita acceder al servidor, explotando la vulnerabilidad detectada.

- Intitle: index.of "Apache/*" "Server at": arroja servidores Apache. Debe especificarse la versión con vulnerabilidades en donde se indica el asterisco.
- Intitle: index.of "Microsoft-IIS/* server at": muestra servidores Microsoft-IIS. Debe especificarse la versión con vulnerabilidades en donde se indica el asterisco.

5. Dispositivos hardware online: a través de estos comandos es posible husmear de forma remota los dispositivos conectados a la web de un usuario en particular.

- Camera linksys inurl: main.cgi: la búsqueda arroja enlaces a cámaras disponibles a las que se puede ingresar sin contraseñas.
- Inurl:"ViewerFrame?Mode=": otra variante que permite visualizar cámaras.
- Intitle:"network print server" filetype:shtm: otra variante que posee el operador interior.

6. Bases de datos con información sensible:

- "robots.txt" "disallow:" filetype:txt: permite visualizar el sector del sitio web que el buscador no muestra, es decir, la información más confidencial de la web.
- Intitle:index.of "parent directory": listado de directorios de un servidor, lo que permitiría su exploración.

7. Información de apoyo al acceso:

- “Microsoft (R) Windows * (TM) Version * DrWtsn32 Copyright (C)” ext:log: permite identificar el antivirus que está instalado en el servidor web y si el mismo cuenta con seguridad perimetral, como por ejemplo firewall.
- Inurl:”8080” -intext:8080: servidores que ejecutan servicios en el puerto 8080.
- Intitle: index.of ws_ftp.log: logs de acceso por FTP que incluyen las rutas locales de los archivos que se suben a la web.

8. Seguridad en Sitios Webs

Es recomendable hacer una verificación de los sitios web propios (o administrados), para determinar y evaluar los riesgos de seguridad existentes, mediante pruebas manuales o automatizadas, a fin de prevenir ataques de terceros mal intencionados. Las primeras evaluaciones se realizan a través de Google Dork y las segundas por aplicaciones que permiten evaluar automáticamente las vulnerabilidades de una página web y determinar qué información confidencial está expuesta a la comunidad digital.

Algunos de los softwares que realizan estas acciones son los siguientes:

- Foca
- SiteDigger
- Wikto
- SearchDiggiti
- Athena
- Gooscan

Cabe destacar que también existen “programas de recompensa”, como los denominados “bugbounty”. Estos consisten en un ofrecimiento realizado por sitios web, organizaciones y/o desarrolladores de software mediante el cual las personas pueden recibir reconocimiento y compensación por informar errores, esencialmente aquellos relacionados con vulnerabilidades y exploits de seguridad.

Estas actividades permiten descubrir y solucionar vulnerabilidades antes que el público en general los sepa, evadiendo incidentes con posterioridad (Wikipedia, s.f.).

4 PROPUESTA DIDÁCTICA O DE CONTENIDO PARA RESOLUCIÓN DE PROBLEMAS

En el marco de la asignatura “Seguridad y Control en Sistemas Informáticos” de la Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán, se efectuará una aplicación práctica, en una disertación mediante videoconferencia, con un docente del Instituto de Ingeniería y Agronomía de la Universidad Nacional Arturo Jauretche. Se les mostrará a los asistentes la técnica objeto de estudio, mediante la búsqueda de vulnerabilidades en páginas web indexadas en Google, con el objetivo de complementar los contenidos abordados en la unidad de estudio “Ciberseguridad en el Comercio Electrónico”.

Esta actividad interdisciplinaria entre la Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán y el Instituto de Ingeniería y Agronomía de Universidad Nacional Arturo Jauretche, se presenta como un valor agregado a formación académica a los alumnos en lo que respecta a la disciplina de Sistemas y Tecnologías de la Información en general y a Seguridad de la Información en particular.

5 ANÁLISIS DE RESULTADOS

En este apartado se mostrarán las principales capturas de pantalla de la demostración práctica realizada, utilizando comandos avanzados de búsqueda para la detección de vulnerabilidades en sitios web indexados en Google. Estas fueron compartidas con los alumnos en una videoconferencia organizada a tal fin.

Se destaca que, para la realización de esta actividad se contrató un servicio de proxy (ProtonVPN), con el objeto de enmascarar la IP pública desde donde se realizaron las búsquedas.

Asimismo, se hace referencia a que solo se efectuaron capturas de pantallas a modo ilustrativo y con fines académicos, no explotando las vulnerabilidades detectadas ni accediendo los archivos encontrados.

A continuación, se muestran los principales hallazgos obtenidos en el trabajo llevado a cabo:

- Password o user name de cuenta de correo electrónico

En esta búsqueda se localizan archivos de Microsoft Word, que contienen cuentas de Gmail con usuario y/o clave de acceso.

- Comando utilizado: allintext:"*.@gmail.com" OR "password" OR "username" filetype:docx

Este código busca archivos de Word que contengan las palabras password, user name y @gmail.com. Se encontraron 181.000 resultados de páginas con archivos que contenían datos de acceso.

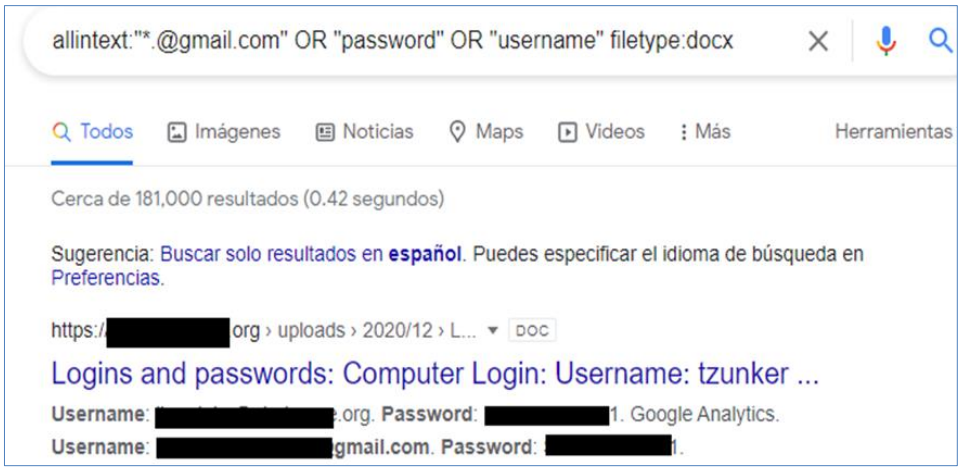

Fig. 1. Listado de archivos de Word con datos de usuario y contraseña. *Fuente*: elaboración propia

Modificando los operadores utilizados anteriormente, ampliamos la búsqueda para archivos "xlsx" con cualquier cuenta de correo electrónico. Se buscó archivos de Excel que contengan las palabras password, user name y el símbolo @, que hace referencia a casillas de correo electrónico globales.

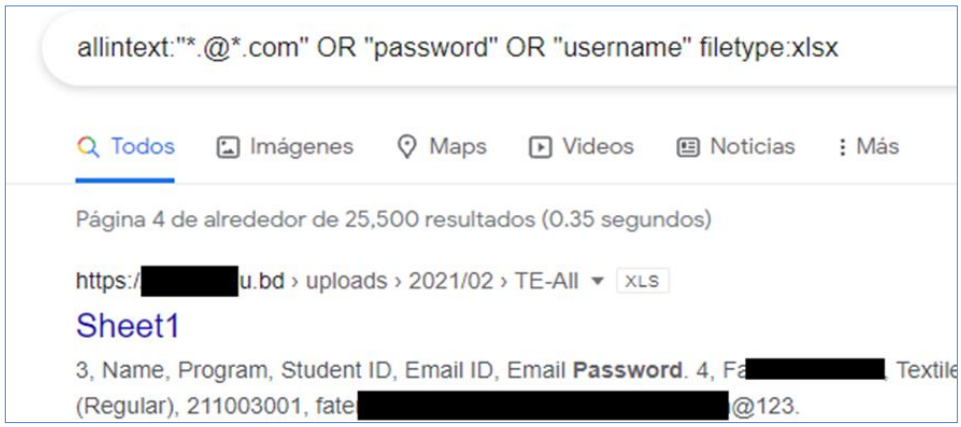

Fig. 2. Listado de archivos de Excel con datos de usuario y contraseña. *Fuente*: Elaboración propia

- Exposición de datos sensibles en "Laravel"

Laravel es un framework de PHP para desarrollos de sitios Web, que almacena un archivo de extensión ".env" que guarda las variables principales y almacena la base de datos de usuarios y contraseñas, accesos al servidor, entre otros datos.

– Comando utilizado: filetype:env intext:APP_NAME

Permite obtener el archivo de configuración de “Laravel” que contiene nombre de usuario, claves, nombre de usuario y toda la información relacionada con la base de datos. Se obtuvo el siguiente resultado:

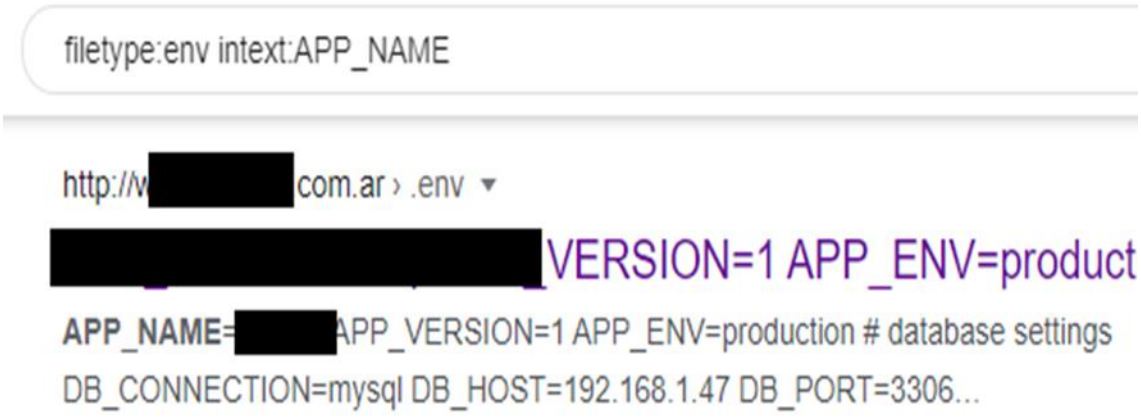

Fig. 3. Comando aplicado a Framework Laravel. *Fuente*: Elaboración Propia

Para mitigar esta vulnerabilidad se recomienda definir la carpeta “Public” como la de acceso a la web y otorgar permisos de solo lectura (400) al archivo “.env”, con la finalidad de que la información no quede publicada en la web (Cuartas, 2020).

- Exposición de archivos shadow y passwd en sitios Webs

En esta búsqueda encontramos el directorio /root/ de diferentes páginas web. Se logra llegar hasta los archivos shadow, passwd y a otros que contienen el hash de las credenciales de los usuarios.

Comando utilizado: inurl:/sym/root/ intitle:index.of

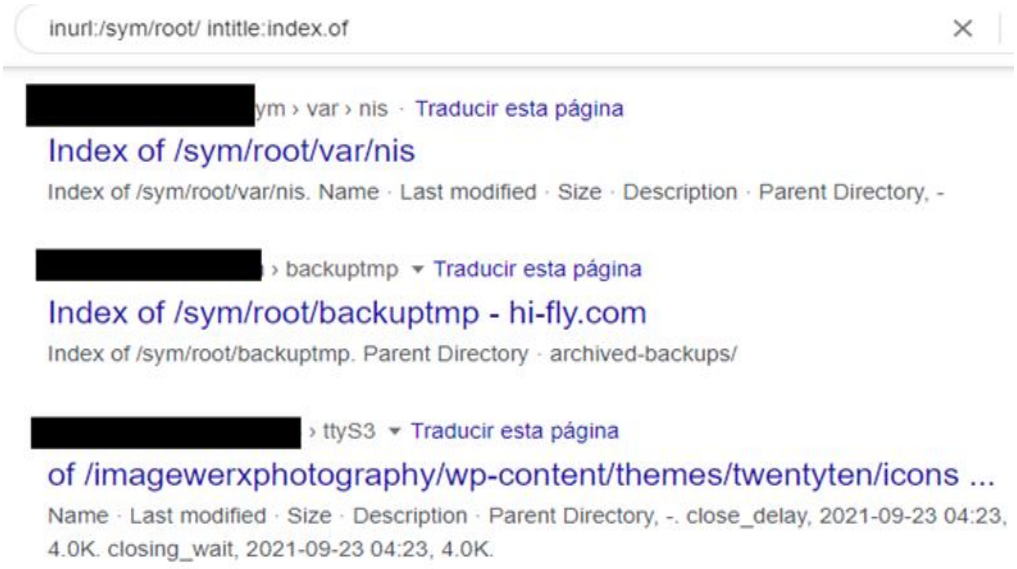

Fig. 4. Acceso a los archivos passwd y shadow. *Fuente*: Elaboración Propia

En la mayoría de los resultados obtenidos, se muestra el acceso a los archivos /etc/passwd, /etc/shadow, lo que permitiría ingresar al sitio web y de este modo vulnerarse la confidencialidad de los datos expuestos.

- Curriculum Vitae con datos personales

El siguiente dork localiza Curriculum Vitae (CV) alojados en sitios webs “edu.ar”. Los mismos contienen información personal sensible de los docentes de la institución, con los cuales podría efectuarse alguna técnica de ingeniería social.

Comando utilizado: site:*.edu.ar “teléfono * * *” “dirección *” “e-mail” intitle:”curriculum vitae”

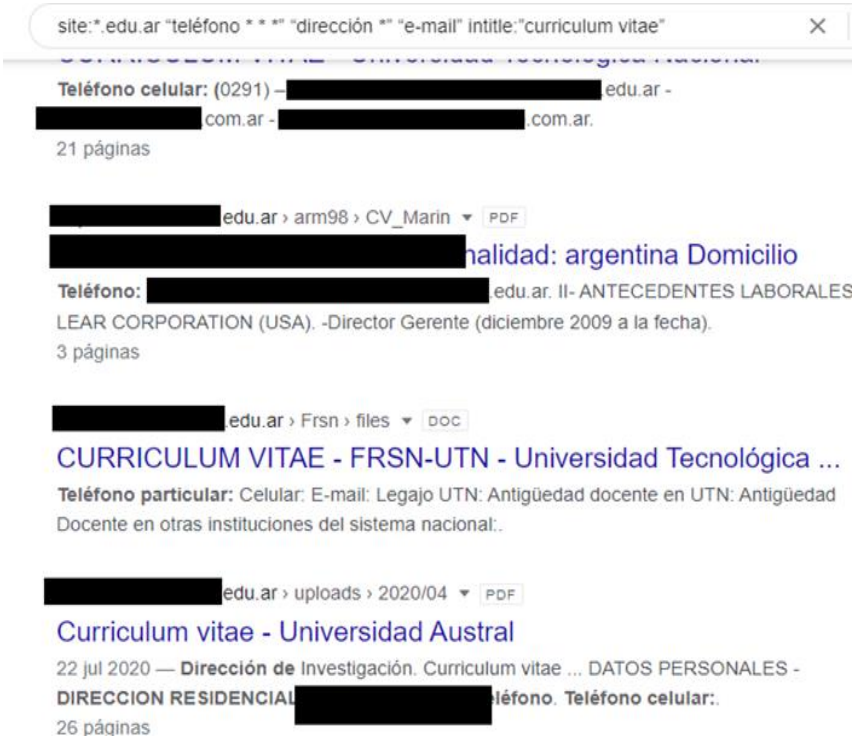

Fig. 5. CV vinculados a sitios web ".edu.ar". *Fuente*: Elaboración Propia

Posteriormente, se efectuó una búsqueda de los CV vinculados a dominios ".uba.ar". Se obtienen las siguientes respuestas:

Fig. 6. CV vinculados a UBA. *Fuente*: Elaboración Propia

- Credenciales expuestas de una Base de Datos

Este comando nos permite encontrar el archivo “db.conf” de SQL Server, en dónde se encuentra información relacionada con la base de datos, usuario y contraseña.

Comando utilizado: intitle:"index of" "db.conf"

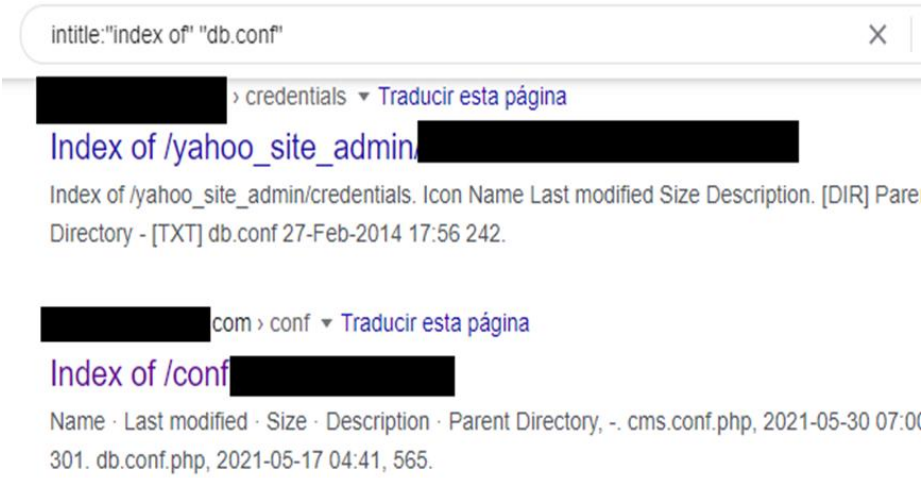

Fig. 7. Archivos de configuración de la BD de SQL Server. *Fuente*: Elaboración Propia

- Sistemas de Refrigeración expuestos

Con la siguiente instrucción podemos obtener acceso a sistemas de refrigeración no protegidos. Esto es considerado una vulnerabilidad grave, debido a que podrían modificarse parámetros de refrigeración en industrias y generar perjuicios económicos en este tipo de organizaciones.

Comando utilizado: inurl:/cgi-bin/cgi.cgi?Cont=

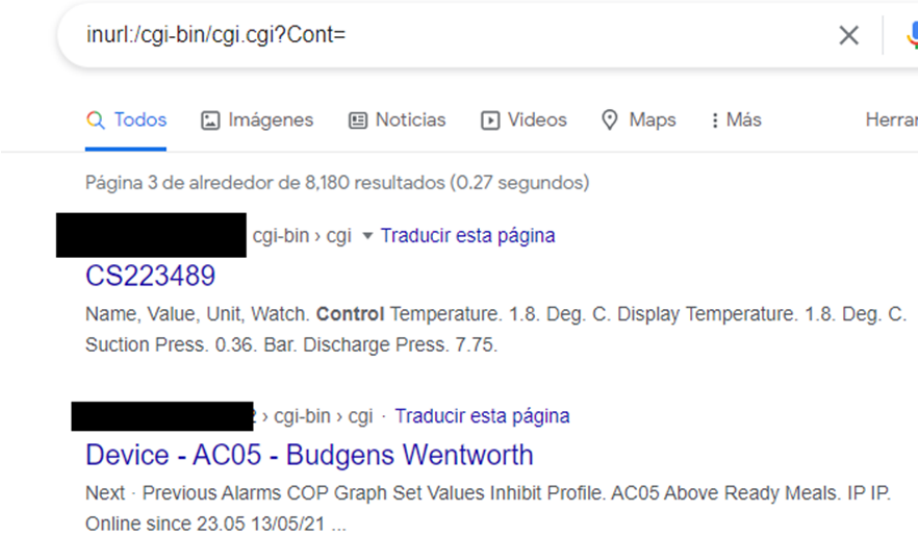

Fig. 8. Acceso a sistemas de refrigeración expuestos en la web. *Fuente*: Elaboración Propia

6 RECOMENDACIONES DE SEGURIDAD Y REPORTE DE VULNERABILIDADES

6.1 Recomendaciones de Seguridad

Es posible que las vulnerabilidades de un sitio web queden expuestas en Internet. Por ello, es importante tener en cuenta las siguientes recomendaciones de seguridad para evitar un impacto económico y/o reputacional al ocasionarse una brecha de datos (Kopca, 2021):

- Mantener el sistema operativo, servicios, plugins, aplicaciones y frameworks (por ejemplo: Laravel y WordPress) actualizadas. deben instalarse los parches de seguridad que sean necesarios.
- Utilizar tecnologías de seguridad, como por ejemplo antivirus o Web application firewall (WAF), para evitar el acceso no autorizado a las plataformas.
- Auditar la exposición de la información en motores de búsqueda, utilizando técnicas avanzadas como las descriptas en este trabajo (Google Dork). Debe tenerse en cuenta que es posible la identificación del usuario que efectúa este tipo de búsquedas a través de la IP de su computadora personal, por lo que siempre la información obtenida debe utilizarse sin perjuicio de terceras partes y con fines legales.
- No almacenar información confidencial en repositorios públicos.
- Efectuar técnicas de evaluación de vulnerabilidades web y de penetración.
- Debe configurarse un archivo con el nombre “robots.txt”, para evitar que los buscadores accedan a datos sensibles del sitio web.
- Si se utiliza el framework “Laravel”, se recomienda definir la carpeta “Public” como la de acceso a la web y dar permisos de solo lectura (400) al archivo “.env”, a fin que la información no quede publicada en la web. Además, ocultar las variables: app_key, db_database, db_username, db_password, tanto en la llave _env como en _server.
- No vincular al sitio web publicado archivos con palabras claves como ser: usuario, contraseña, password, user name, entre otros.
- En caso de tener que publicar archivos en sitios web, tomar la precaución de que los mismos no contengan datos personales sensibles.
- Respecto a sistemas OT y dispositivos expuestos en la web, deben cambiarse las credenciales de acceso por defecto. Además, combinándose esta información con otros buscadores más avanzados (ejemplo: Shodan), pueden explotarse las vulnerabilidades expuestas y generar impactos más significativos.
- Tener en cuenta las recomendaciones para el desarrollo web seguro que presenta el OWASP Top 10.

6.2 Reporte de vulnerabilidades a las compañías

Una de las maneras de reportar las vulnerabilidades detectadas en sitios web, es contactar por mail al administrador de seguridad de la compañía afectada o al desarrollador del sitio, para informarle sobre la vulnerabilidad encontrada.

Otro modo más rentable económicamente es utilizar plataformas de BugBounty como “HackerOne” o “Bugcrowd”, donde las empresas que participan pagan una recompensa (las búsquedas a reportar son con un target específico). También existen foros para realizar este tipo de actividad.

Se hace referencia a que el 15 de noviembre del 2021, uno de los autores de este trabajo envió un mail a una empresa, reportando vulnerabilidades y enviando capturas de pantallas de lo encontrado. Sin embargo, hasta el día de la fecha no se obtuvo respuesta y los datos en el sitio web continúan expuestos.

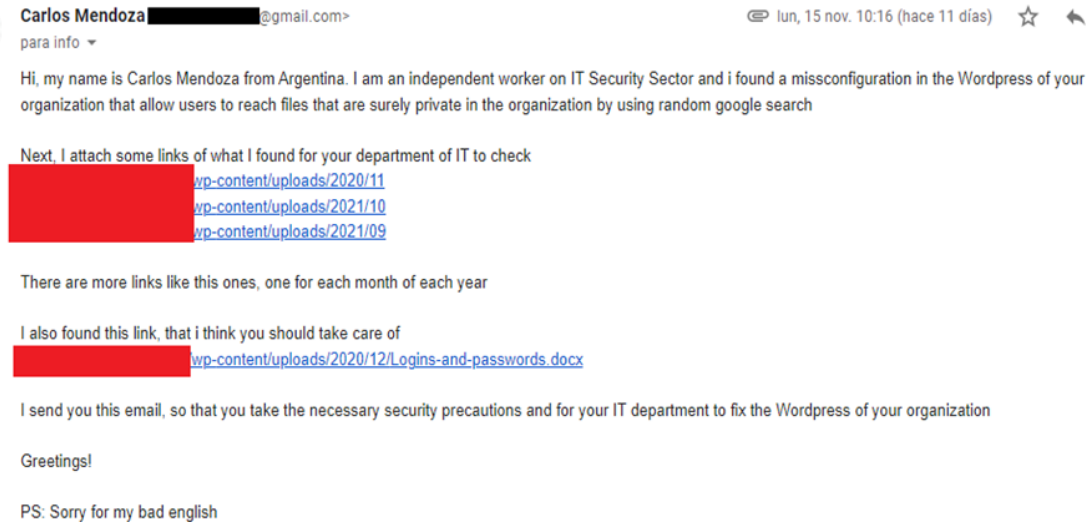

Fig. 9. Mail reportando vulnerabilidades en sitio web

7 DISCUSIÓN, CONCLUSIÓN, RECOMENDACIONES PARA FUTUROS ESTUDIOS

La recopilación bibliográfica realizada nos permitió reconocer los principales operadores avanzados de búsquedas utilizados para la detección de vulnerabilidades en sitios web, a través de la herramienta “Google Dorks”.

Se evidencia que las técnicas y herramientas para la recopilación y análisis de información son cada vez más populares, lo que las constituye en un aspecto clave de la seguridad de la información. Un acceso con fines maliciosos, aprovechando las vulnerabilidades de configuración de un sitio web, podrían causar graves consecuencias económicas y reputacionales a una organización o causar daños a personas en el mundo real.

El espionaje electrónico y el uso malintencionado de la información son problemas a los que se enfrentan tanto las organizaciones como los usuarios. Es por esto que la seguridad en Internet se ha convertido en una prioridad, ya que las técnicas que utilizan los ciberdelincuentes para acceder a información confidencial, se volvieron altamente

sofisticadas. Muchas de ellas explotan vulnerabilidades y/o malas configuraciones de los sitios web publicados.

Se considera que los profesionales en Ciencias Económicas deben tener en cuenta las recomendaciones de seguridad planteadas en este trabajo, a fin de alertar a los administradores de los sitios web de la organización. Esto contribuirá a garantizar la confidencialidad e integridad de la información almacenada en la web. Por el contrario, su inobservancia podría ocasionar un perjuicio hacia los involucrados, por exponer datos sensibles que pueden ser utilizados con fines maliciosos.

Se hace hincapié en que la técnica desarrollada en el presente trabajo también puede ser utilizada para buscar vulnerabilidades en sitios web propios y aprender de ello, a fin de alcanzar niveles aceptables de seguridad y evitar que terceros no autorizados accedan a información restringida. Lo descripto anteriormente, puede combinarse con aplicaciones y herramientas de búsqueda de vulnerabilidades para profundizar el análisis a tal fin (por ejemplo: Zap, Nmap, Foca, Nessus, Metaesploit, pwnXSS, etc).

Finalmente se enfatiza en que la demostración práctica en el aula a los alumnos de la asignatura “Seguridad y Control en Sistemas Informáticos” (UNT), con el objetivo de complementar los contenidos abordados en la unidad de estudio “Ciberseguridad en el Comercio Electrónico”, agrega considerable valor a su formación académica. Esto se fundamenta en que los mismos deberán estar capacitados para gestionar organizaciones en contextos altamente informatizados e interconectados, y en donde la presencia en el entorno digital de las organizaciones reviste una trascendental importancia para su posicionamiento estratégico

8 REFERENCIAS BIBLIOGRÁFICAS

1. Añais, L. (abril de 2021). Google Dorks. Users - Especial Nº136.
2. Cuartas, J. (16 de abril de 2020). ¡Cuidado con tu archivo .env! No olvides hacer esto. Obtenido de Laraveltip: https://www.laraveltip.com/cuidado-con-tu-avides-hacer-esto/
3. Kopca, T. (16 de abril de 2021). Google Dorks Como Encontrar Datos Interesantes y Buscar Como Un Hacker. Obtenido de Ma-no.org: https://www.ma-no.org/es/seguridad/google-dorks-como-encontrar-datos-interesantes-y-buscar-como-un-hacker
4. OWASP (s.f). Top 10 de OWASP – 2021. Obtenido de owasp.org: https://owasp.org/Top10/A00_2021_Introduction/
5. Pastorino, C. (07 de octubre de 2019). Técnicas y herramientas OSINT para la investigación en Internet. Obtenido de We live security. https://www.welivesecurity.com/la-es/2019/10/07/tecnicas-herramientas-osint-investigacion-internet/
6. Raggi, N. (29 de julio de 2021). Google hacking: averigua cuanta información sobre ti o tu empresa aparece en los resultados. Obtenido de We live security: https://www.welivesecurity.com/la-es/2021/07/29/google-hacking-averigua-que-informacion-sobre-ti-o-empresa-aparece-resultados/
7. Tryhackme. (16 de enero de 2021). Google Dorking. Obtenido de Tryhackme: https://tryhackme.com/room/googledorking
8. Wikipedia. (S.F.). Google Hacking. Obtenido de Wikipedia: https://es.wikipedia.org/wiki/Google_Hacking

9. Wikipedia. (S.f.). Programa de recompensas por errores. Obtenido de Wikipedia: https://en-m-wikipedia-org.translate.goog/wiki/Bug_bounty_program?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=nui,sc