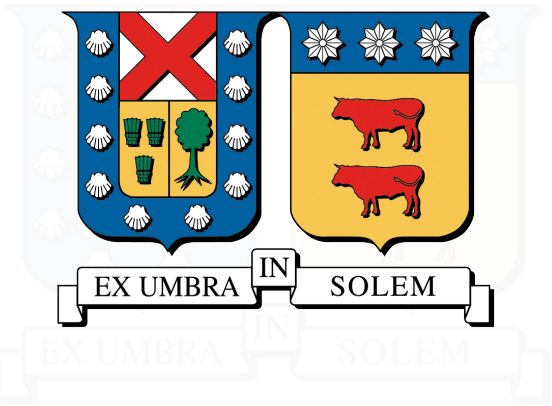
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
VALPARAÍSO - CHILE



IMPLEMENTACIÓN DE MEMORIA Y AUTOMATIZACIÓN EN LA DINÁMICA DEL RECONOCIMIENTO DE RED APLICADO EN EL MARCO DE LA CIBERSEGURIDAD DEFENSIVA

VICTOR BENJAMÍN CORTÉS MORALES

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL TELEMÁTICO

PROFESOR GUÍA : NICOLÁS JARA C.
PROFESOR CORREFERENTE : FRANCISCO CABEZAS B.

OCTUBRE 2023

RESUMEN EJECUTIVO

Desde tiempos inmemoriales, la comunicación ha desempeñado un papel esencial en la experiencia humana. En épocas ancestrales, nuestros antepasados utilizaban diversos medios para expresarse y conectarse con sus semejantes. Esto haciendo uso de la pintura en rocas hasta el uso de sonidos para comunicar. En este caso podemos ver que la evolución de la comunicación ha sido un testimonio del ingenio humano.

A medida que la civilización progresó, surgieron diversas lenguas y dialectos, algunas de las cuales evolucionaron con el tiempo, mientras que otras se perdieron en la corriente de la historia. Este proceso continuó hasta llegar a la era moderna, donde la comunicación analógica se convirtió en la norma, antes de dar paso a la era digital que define nuestra época actual.

En la actualidad, la comunicación digital avanza a pasos agigantados, acompañada de la digitalización de tareas que anteriormente eran ajenas al mundo digital. Este avance se ha infiltrado en prácticamente todos los aspectos de nuestras vidas, llegando al punto en que prácticamente cualquier objeto puede estar conectado a una red de nodos, es decir, a Internet.

Sin embargo, este vertiginoso crecimiento en la esfera digital también ha propiciado un aumento en los delitos cibernéticos. El ciberespacio se ha convertido en un terreno lucrativo para organizaciones criminales y en un punto estratégico de interés para entidades gubernamentales. En este contexto, la infraestructura crítica está cada vez más interconectada a través de Internet, lo que la hace vulnerable a ataques y amenazas cibernéticas.

Es evidente, a la luz de lo anterior, que la ciberseguridad desempeña un papel crucial en la sociedad actual y en la venidera. Para abordar estos desafíos, han surgido organizaciones dedicadas a desarrollar marcos de trabajo y brindar comprensión sobre las diversas facetas de la ciberseguridad y el hacking, tales como MITRE ATT&CK y OWASP. Estas entidades desempeñan un papel fundamental en la protección y la preservación de la integridad de nuestras infraestructuras digitales en un mundo cada vez más interconectado.

En el contexto de la empresa ENTEL, se ha planteado un desafío significativo en las etapas iniciales del ethical hacking. Estas etapas se centran en la fase de reconocimiento, un paso crucial que se extrae de la matriz de MITRE ATT&CK. Dado que esta fase marca el punto de partida, resulta de suma importancia llevar a cabo un desempeño óptimo para identificar con precisión los .ªctivos"que componen la organización. Este proceso no solo implica reconocer estos activos, sino también comprender su ubicación dentro de la red y, en última instancia, concebir un plan defensivo destinado a proteger los sectores o activos críticos de la organización.

El enfoque de esta tesis se divide en dos partes fundamentales:

La primera parte se enfoca en el desarrollo de un software adaptable a las necesidades específicas de la organización. Este software tiene como objetivo realizar un escaneo exhaustivo de toda la red de ENTEL en Chile y Perú. Los resultados generados por esta herramienta deben ser almacenados meticulosamente, lo que permitirá la creación de una memoria detallada sobre la red. Además, se ha considerado la importancia de que la herramienta pueda abordar diversos tipos de errores y situaciones límite que puedan surgir durante el proceso de escaneo. Un elemento esencial es la automatización de este proceso, programado para ejecutarse en intervalos regulares.

La segunda parte de este proyecto implica la creación de una interfaz intuitiva y accesible para los especialistas de ciberseguridad que permita acceder a la información generada por la aplicación. Esta interfaz está diseñada para ser de fácil comprensión y se adapta a las necesidades específicas de cada especialista, ofreciendo distintas interfaces según los requerimientos individuales.

La aplicación, bautizada como “ENTELRECON” , en honor a la herramienta .ªutorecon"desarrollada por "t1berius", se presenta como una solución esencial para abordar los desafíos de la fase de reconocimiento en el proceso de ethical hacking. Su objetivo principal es proporcionar un acceso sencillo y eficaz a los equipos de ciberseguridad, con el propósito de optimizar la búsqueda de información sobre los activos de la red corporativa. Este enfoque estratégico refleja el compromiso de ENTEL en fortalecer su postura de

seguridad cibernética y garantizar la integridad de sus activos digitales en un entorno cada vez más complejo y amenazante.

Keywords: Ciberseguridad, ethical hacking, activos, ciberespacio, MITRE ATT&CK, OWASP.

ABSTRACT

Communication has played an essential role in the human experience since time immemorial. In ancient times, our ancestors used various means to express themselves and connect with their fellow humans. From painting on rocks to the use of fire and sounds, the evolution of communication has been a testament to human ingenuity.

As civilization progressed, different languages and dialects emerged, some evolving over time while others were lost to the currents of history. This process continued until the modern era, where analog communication became the norm before giving way to the digital age that defines our current epoch.

Today, digital communication is advancing by leaps and bounds, accompanied by the digitalization of tasks that were previously unrelated to the digital world. This advancement has infiltrated virtually every aspect of our lives, reaching the point where practically any object can be connected to a network of nodes, i.e., the Internet.

However, this rapid growth in the digital sphere has also led to an increase in cybercrimes. Cyberspace has become a lucrative terrain for criminal organizations and a strategic point of interest for governmental entities. In this context, critical infrastructure is becoming increasingly

interconnected through the Internet, making it vulnerable to cyberattacks and threats.

It is evident, in light of the above, that cybersecurity plays a crucial role in today's society and the future. To address these challenges, organizations dedicated to developing frameworks and providing insights into various aspects of cybersecurity and hacking have emerged, such as MITRE ATT&CK and OWASP. These entities play a fundamental role in protecting and preserving the integrity of our digital infrastructures in an increasingly interconnected world.

In the context of the company ENTEL, a significant challenge has been posed in the early stages of ethical hacking. These stages focus on the reconnaissance phase, a crucial step derived from the MITRE ATT&CK matrix. Since this phase marks the starting point, it is of utmost importance to perform optimally in order to accurately identify the .ªssets"that make up the organization. This process not only involves recognizing these assets but also understanding their location within the network and, ultimately, conceiving a defensive plan aimed at protecting the organization's critical sectors or assets.

The focus of this thesis is divided into two fundamental parts:

The first part focuses on the development of software adaptable to the specific needs of the organization. This software aims to conduct a comprehensive scan of ENTEL's network in Chile and Peru. The results generated by this tool must be meticulously stored, enabling the creation of a detailed memory of the network. Furthermore, the importance of the tool's ability to address various types of errors and edge cases that may arise during the scanning process has been considered. An essential element is the automation of this process, scheduled to run at regular intervals.

The second part of this project involves creating an intuitive and accessible interface for cybersecurity specialists to access the information generated by the application. This interface is designed to be easily understood and tailored to the specific needs of each specialist, offering different interfaces according to individual requirements.

The application, named "ENTELRECON", in honor of the "autorecon" tool developed by "t1berius,"presents itself as an essential solution to address the challenges of the reconnaissance phase in the ethical hacking process. Its primary goal is to provide easy and effective access to cybersecurity teams, with the purpose of optimizing the search for information about the corporate network assets. This strategic approach reflects ENTEL's commitment to strengthening its cybersecurity posture and ensuring the integrity of its digital assets in an increasingly complex and threatening environment.

Keywords: Cybersecurity, ethical hacking, assets, cyberspace, MITRE ATT&CK, OWASP.

GLOSARIO

- **DNS:** El Sistema de Nombres de Dominio traduce los nombres de dominio fáciles de recordar en direcciones IP numéricas.
- **BugBounty:** Es un programa (no informático) ofrecido por empresas o grandes organizaciones que tiene como objetivo motivar a expertos de la ciberseguridad a buscar y reportar vulnerabilidades en sitios con una recompensa monetaria a cambio del reporte.
- **Sonar:** Un sonar es un instrumento electrónico que en base a ondas sonoras puede localizar y detectar objetos en el agua, en este caso lo utilizamos para hacer una analogía al funcionamiento de la herramienta a desarrollar pero en una red.
- **Riesgo:** El riesgo lo definimos como la criticidad de un activo dentro de la red, este se calcula en base a la multiplicación de la probabilidad de ocurrencia y el impacto.
- **Timestamp:** Hace referencia a una marca de tiempo que en este caso es utilizado para marcar inicio y termino de los escaneos a realizar.
- **Pentesting:** Práctica de ciberseguridad que evalúa la seguridad de un sistema, aplicación o red mediante la simulación de un ataque de un adversario.
- **Logs:** Son archivos que guardan información sobre actividades que suceden en un sistema informático, para este caso se utilizan para guardar información sobre los eventos realizados por la herramienta creada.
- **WHOIS:** Es una herramienta nativa del sistema operativo kali linux que se utiliza para acceder a los registros WHOIS de los dominios. Además WHOIS es un protocolo que permite obtener información registrada sobre un dominio de internet, como el propietario del dominio, la organización, la dirección de contacto, el registrador, las fechas de creación y expiración, entre otros detalles.
- **Google Dorks:** son técnicas que utilizan operadores avanzados en los motores de búsqueda de Google para encontrar información específica que es difícil de localizar mediante búsquedas simples, un ejemplo puede ser el siguiente “ext:pdf” el cual permite solo buscar archivos con la extensión .pdf al realizar la búsqueda por google.
- **ICMP:** Este protocolo se encuentra en la capa de red del modelo OSI. Es utilizado principalmente para enviar mensajes de error y operativos que pueden llegar a indicar que un servicio o host en particular es inaccesible o que un paquete de datos no puede llegar a su destino.
- **TI:** Las tecnologías de la información(TI), se refieren al uso de sistemas computacionales, software y redes para almacenar, procesar y distribuir datos
- **Hackeo:** HAce referencia al acto de manipular o acceder a sistemas de computadoras, redes o dispositivos electrónicos, “generalmente” sin la autorización de los propietarios o administradores de dichos sistemas.
- **Ransomware:** Es un tipo de malware diseñado para cifrar los archivos de un dispositivo o bloquear el acceso al sistema afectado, con el objetivo de exigir un rescate (generalmente monetario) a cambio de la clave de descifrado o la restauración del acceso.
- **OSBanner:** Información descriptiva sobre el sistema operativo de un servidor o dispositivo que se transmite a través de servicios de red o protocolos, especialmente durante el proceso de establecimiento de conexión.
- **Malware:** Corresponde a cualquier tipo de software diseñado para infiltrarse, dañar o realizar acciones no autorizadas en un sistema informático sin el consentimiento del usuario.
- **Phishing:** Es una técnica generalmente utilizada por ciberdelincuentes para engañar a las personas y obtener información sensible o acceso a sus sistemas a través del correo electrónico.

- **Smishing:** Es una técnica generalmente utilizada por ciberdelincuentes para engañar a las personas y obtener información sensible o acceso a sus sistemas a través de mensajes de texto SMS.
- **SIEM:** Es una solución tecnológica en el campo de la seguridad informática que proporciona una visión más completa y en tiempo real del estado de seguridad de una organización.
- **CMD:** Interprete de comandos en los sistemas operativos.
- **Incident Response:** Es un conjunto de procedimientos diseñados para identificar, investigar y responder a posibles amenazas de seguridad en una organización.
- **API:** Es un conjunto de reglas y especificaciones que las aplicaciones pueden seguir para comunicarse entre sí.
- **Threads:** Son componentes fundamentales en la programación paralela que permiten a un programa realizar múltiples tareas al mismo tiempo dentro de un mismo proceso.

Índice de Contenidos

Índice de Tablas

Índice de Figuras

1 | Introducción

Es imposible ignorar los profundos cambios que ha experimentado la humanidad en los últimos 20 años, especialmente en cómo la sociedad se ha adaptado a las transformaciones tecnológicas. Esto incluye el auge del entretenimiento digital como redes sociales y plataformas de streaming bajo demanda, el incremento de las compras por internet abarcando desde comida rápida hasta tiendas de retail, así como los variados servicios en línea para pagos de servicios básicos como agua, luz e internet. Además, no podemos olvidar el impacto de las aplicaciones móviles en ámbitos tan diversos como el transporte, el comercio, la educación y el ocio.

Ante este panorama, es crucial mantener una conexión segura entre plataformas, donde se protejan aspectos fundamentales como la integridad, autenticidad y confiabilidad (Triada CIA). Esta protección es esencial para salvaguardar nuestra privacidad y evitar ser suplantados en el vasto mundo digital.

Es en este contexto de creciente digitalización de las actividades cotidianas que el término 'ciberseguridad' se ha vuelto más relevante, cobrando fuerza y presencia en el vocabulario de nuestra sociedad actual. La ciberseguridad no es solo una necesidad, sino una responsabilidad compartida que nos concierne a todos en la era digital.

En el marco de este proyecto, ENTEL nos encargó la importante misión de mejorar la visibilidad de la superficie de ataque presente en la empresa. Con este objetivo en mente, desarrollamos una herramienta diseñada específicamente para ayudar a ENTEL a obtener una perspectiva más clara y detallada de sus activos y servicios. Esta herramienta facilita la identificación y gestión de los recursos disponibles tanto en el ámbito interno, con acceso restringido, como en el externo, accesible al público. Nuestro enfoque se centró en fortalecer la infraestructura de la empresa contra posibles amenazas, asegurando así una gestión más eficiente y segura de su entorno digital.

1.1. ¿Qué es la ciberseguridad?

La ciberseguridad es el arte de proteger redes, dispositivos y datos contra el acceso no autorizado o el uso criminal, así como la práctica de asegurar la confidencialidad, integridad y disponibilidad de la información (CIA).

1.2. ¿Que riesgos implica tener una ciberseguridad deficiente?

Existen muchos riesgos, algunos más graves que otros. Entre estos peligros se encuentran los malware que borran información de un sistema, un atacante que irrumpe en un sistema y altera archivos, un atacante que utiliza un ordenador para atacar a otros, o un atacante que roba la información de tarjeta de crédito y realiza compras no autorizadas. No hay garantía de que, incluso con las mejores precauciones, algunas de estas cosas no sucedan, pero hay medidas que pueden ser tomadas para minimizar las probabilidades.

Una vez que un atacante tiene acceso de administrador de algún sistema, lo que puede llegar a hacer queda sujeto solamente a la imaginación del atacante.

1.3. ¿Que es la ciberseguridad defensiva?

La ciberseguridad defensiva es un mecanismo de protección de redes informáticas que incluye respuesta a acciones y protección de infraestructuras críticas y aseguramiento de la información para organizaciones, entidades gubernamentales y otras posibles redes.

Se centra en prevenir, detectar y proporcionar respuestas oportunas a ataques o amenazas para que ninguna infraestructura o información sea alterada.

Esta es esencial para la mayoría de las entidades con el fin de proteger la información sensible, así como para salvaguardar los activos.

1.4. ¿En qué consiste nuestro desafío?

El problema surge en el desconocimiento de los servicios y puertos que se mantienen abiertos en la red de ENTEL, esto genera un problema al momento de tomar decisiones sobre que puede ser un potencial peligro dentro de la red coorporativa, como también el no tener una visión general y unificada para cada uno de los especialistas de la empresa al momento de hacer realizar un reconocimiento del activo en la red.

Al día de hoy cada especialista tiene distintas formas y metodologías para poder saber que se encuentra dentro de la red, como por ejemplo planillas excel que deben estar bajo distintas actualizaciones, realizar escaneos utilizando herramientas como nmap para algún activo en particular o hacer búsqueda en algunas herramientas que maneja la empresa ENTEL.

La motivación para resolver esta problemática radica en la importancia que tiene el poder tener un buen entendimiento de lo que está en la red, esto debido a que la fase de reconocimiento le da a un potencial hacker o grupo hacker la superficie de ataque y las posible entradas a los distintos activos de la empresa.

1.5. Acercamiento a la solución

Lo que se busca realizar es desarrollar un sistema capaz de poder escanear los distintos servicios y puertos abiertos que utilizan los activos tanto de la red de ENTEL Chile y ENTEL Perú. Algo parecido a un "sonar"pero para la red de ENTEL.

También se busca realizar una búsqueda de vulnerabilidades, esto acorde a las distintas versiones de software que corren las distintas herramientas escaneadas para los distintos activos esto con el fin de encontrar alguna brecha de seguridad dentro de la red que deba ser actualizada o contenida hasta que se encuentre alguna solución por parte del equipo de ciberseguridad defensiva de la empresa. También se espera hacer pruebas de higienización como password spray, generar un vector de ataque asociada a cada una de las pruebas y vulnerabilidades encontradas. Una vez realizada esta ardua tarea, se busca la automatización del guardado de estos resultados de las escaneos de red y pruebas activas a los disntintos activos de la red.

Con toda esta data generada esperamos poder hacer el cálculo del “riesgo” de cada activo en base a los vectores de ataque. Con esto podemos ver que tan riesgoso puede ser cada activo para la organización y así priorizar al momento de recibir la información.

Es importante que dentro de este proceso del guardado de la información recopilada por la herramienta se cree un historial sobre los escaneos realizados, así de esta forma de podría generar memoria en el proceso y también pueda ser posible revisar la información a lo largo del tiempo con un “timestamp” asociado.

Finalmente se busca tomar esta información y disponibilizarla dentro de un sharepoint en la red coorporativa de la empresa para que pueda ser consultada por los distintos especialistas, esto haciendo uso de alguna herramienta de visualización de datos para que sea de fácil acceso y entendimiento.

1.6. Objetivo General

Diseñar e implementar un sistema de reconocimiento automático, que almacene datos de interés de manera estructurada y simple de exportar. Sobre esos datos generados, integrar una forma para identificar aquellos vectores de ataque más riesgosos para el cliente.

1.7. Objetivos Específicos

1. Visión clara del desafío y planificación del desarrollo.

   Se definen las problemáticas del desafío, entendiendo el contexto y dolor del cliente y se genera una planificación clara y realista del desarrollo y actividades que se deberán realizar para poder entregar una solución acorde al desafío, en el tiempo especificado.

2. Automatización de técnicas de reconocimiento.

   Desarrollo e implementación en un software que integra y automatiza el uso de las herramientas típicas de reconocimiento activo, con una configuración personalizable por el cliente.

3. Almacenamiento de resultados de escaneos.

   Los resultados obtenidos en el reconocimiento automático serán filtrados y reducidos a sólo aquellos datos útiles necesarios para describir la superficie de ataque expuesta.

4. Listar los posibles vectores de ataque

   Utilizando inteligencia artificial, se analizan todos los datos útiles asociados a un activo y en base a estos, se presentan todos los vectores de ataque posibles traducidos en ítems de la metodología de "pentesting" interna del cliente.

5. Encapsulación

   El desarrollo del módulo de Reconocimiento y de Decisión deben ser encapsulables y distribuibles.

2 | Estado del arte

2.1. Marco teórico

En el capítulo anterior se ha explicado brevemente, y en rasgos generales en que consiste el desafío y el contexto en el que este está inmerso.

A continuación se explican en profundidad los conceptos claves, donde además se abordará en detalle el proceso actual .

2.1.1. ¿Que es el RedTeam y el BlueTeam?

En el contexto de la ciberseguridad, las grandes organizaciones que tienen equipos de ciberseguridad tienden a tener 2 equipos especializados que son el RedTeam y el BlueTeam.

La función del RedTeam es la de simular ser adversarios externos o intrusos dentro de la organización. Su objetivo principal recae en evaluar la seguridad, esto identificando vulnerabilidades o posibles puntos de entrada que pudiesen ser explotados. Aparte de esto, estos suelen realizar pruebas de penetración con la intención de medir las defensas de la organización.

En el caso del BlueTeam, estos tienen la misión y el objetivo de proteger los activos y sistemas de la organización. Esto manteniendo medidas de seguridad , monitoreando redes y sistemas en búsqueda de actividad sospechosa y en casos deben estar preparados para responder frente a incidentes a los que se pueda enfrentar la organización. Estos utilizan herramientas de detección de amenazad en los activos y sistemas, como también análisis de “logs” y procedimientos de respuesta de incidentes para proteger la infraestructura TI.

Tanto el RedTeam como el BlueTeam deben trabajar a la par para entender las posibles amenazas que pueden llegar a la organización, así como para mejorar las defensas de la misma.

2.1.2. ¿Que es MITRE ATT&CK?

MITRE ATT&CK® es una base de conocimiento globalmente accesible sobre tácticas y técnicas adversarias basadas en observaciones del mundo real. La base de conocimientos de ATT&CK se utiliza como base para el desarrollo de modelos y metodologías de amenazas específicas en el sector privado, el gobierno, la comunidad de productos y servicios de ciberseguridad.

Esta organización tiene una matriz de tácticas y técnicas que representan el comportamiento que suelen tener los adversarios.

Cuando hablamos de táctica, hablamos de la **razón de ejecutar una acción** y cuando hablamos de técnica, hablamos de cómo un adversario logra un objetivo táctico al realizar una acción.

Existen distintas matrizes de mitre att&ck, la que veremos es la mitre att&ck enterprise.

Por parte de las tácticas tenemos 14 las cuales son:

1. Reconocimiento
2. Despliegue de recursos
3. Acceso inicial
4. Ejecución de código
5. Persistencia
6. Escalamiento de privilegios
7. Evasión
8. Acceso a credenciales
9. Descubrimiento
10. Movimiento lateral
11. Recolección de datos
12. Comando y Control
13. Exfiltración
14. Impacto

Estos están representados en orden creciente, donde el número 1 (Reconocimiento) corresponde a la táctica inicial utilizada por un atacante. Esta es la etapa donde vamos a dirigir este trabajo, el cual se considera como la etapa más importante al momento de trabajar en un proceso de ciberseguridad defensiva.

La figura 2.1 muestra como se ve una parte de la matriz de mitre att&ck enterprise.

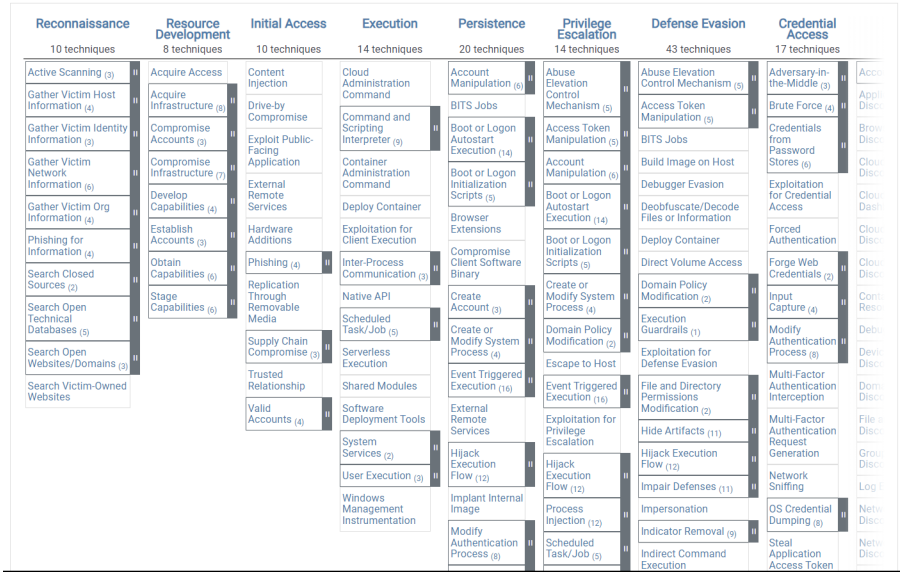


Imagen sacada de https://attack.mitre.org/matrices/enterprise/

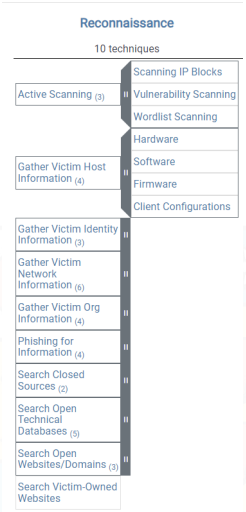
Figura 2.1: Extracto matriz MITRE ATT&CK enterprise

2.1.2.1. Táctica de Reconocimiento

El adversario está intentando recopilar información que pueda utilizar para planificar operaciones futuras.

El reconocimiento consiste en técnicas que implican que los adversarios recopilen activa o pasivamente información que puede utilizarse para apoyar la selección de objetivos. Dicha información puede incluir detalles de la organización, infraestructura o personal de la víctima.

La figura 2.2 muestra las táctica de reconocimiento con sus técnicas asociadas.



https://attack.mitre.org/matrices/enterprise/

Figura 2.2: Táctica de reconocimiento y sus técnicas asociadas

En el contexto del reconocimiento existen 2 sub categorías, donde podemos encontrar el reconocimiento pasivo y el reconocimiento activo. Para el caso de este trabajo, solo utilizamos técnicas de reconocimiento del tipo activo. Pero por temas del entendimiento entre las diferencias vamos a indagar en estos 2 conceptos.

Reconocimiento pasivo: El reconocimiento pasivo se enfoca en la discreción y la recolección de información sin alertar al adversario. Por ejemplo el obtener información utilizando herramientas como “WHOIS” es considerado reconocimiento pasivo o haciendo búsquedas con “google dorks”, donde hacemos búsqueda de información sobre el activo sin interactuar con este mismo.

El objetivo del reconocimiento pasivo se basa en recopilar detalles como nombres de dominio, configuraciones de red, detalles sobre tecnologías utilizadas, información personal de empleados, etc. Todo esto para preparar ataques o auditorías más precisas y efectivas.

Reconocimiento activo: El reconocimiento activo busca obtener información detallada y específica, aunque con un mayor riesgo de ser detectado. Se tiene una interacción directa con el activo en cuestión. Esto significa enviar paquetes de datos o solicitudes y analizar las respuestas recibidas para obtener información detallada.

El objetivo del reconocimiento activo se basa en identificar servicios activos tales como sistemas operativos en uso, aplicaciones y sus versiones, configuraciones de red específicas, y potenciales puntos de entrada para ataques.

2.1.3. Network Scan

El escaneo de red es una técnica que es utilizada para hacer una investigación de una red, con el propósito de identificar los dispositivos conectados y sus características. Esto implica el descubrir si los dispositivos se encuentran activos, que sistema operativo pueden estar utilizando, aplicaciones, además de que puertos y servicios están abiertos y disponibles.

Generalmente esta técnica es utilizada por administradores de red para poder gestionar la red, mientras que por el área de la seguridad se utiliza para evaluar y fortalecer las defensas de la red.

Existen distintas herramientas para realizar escaneos de red, así como también distintas subtécnicas dentro del escaneo de red, donde podemos obtener información utilizando paquetes "ICMP" a varias direcciones de red para identificar dispositivos disponibles o podemos hacer escaneo de puertos donde se trata de establecer una conexión a los diferentes puertos de un activo y de esta forma se espera obtener respuesta confirmando que el puerto se encuentra disponible para posteriormente ver cual es el servicio que ofrece el puerto correspondiente. También es posible realizar escaneo de vulnerabilidades utilizando bases de datos de vulnerabilidades conocidas y así identificar puntos débiles dentro del sistema operativo del activo o de las aplicaciones que utiliza.

2.1.3.1. Dirección IP

La dirección IP (Internet Protocol) es un número único que se asigna a cada dispositivo conectado a una red que usa el Protocolo de Internet para la comunicación. Este identificador permite a los dispositivos enviar y recibir datos entre sí de manera organizada y eficiente. Las direcciones IP son fundamentales para el funcionamiento de internet y las redes privadas.

Estas proporcionan una identidad única a cada dispositivo en una red y ayudan a determinar la dirección lógica de un dispositivo en la red, asegurando que el tráfico se dirija correctamente entre dispositivos, esto ya sea a través de internet o en una red local.

Existen 2 versiones para las direcciones ip, estas son las direcciones de IPv4 e IPv6:

IPV4: Se compone de 32 bits en notación decimal y consta de 4 números, donde cada uno tiene un valor entre 0 y 255, un ejemplo de una dirección ipv4 puede ser **192.168.0.1**.

Existen unos rangos de direcciones IP que han sido asignadas para redes privadas, estas son:

Clase A: 10.0.0.0 a 10.255.255.255
Clase B: 172.16.0.0 a 172.31.255.255
Clase C: 192.168.0.0 a 192.168.255.255

IPv6: Esta fue creada para resolver el problema con la escasez de direcciones ipv4, esta versión tiene 128 bits, se compone de 8 grupos de 4 dígitos hexadecimales cada uno, un ejemplo de una dirección ipv6 puede ser **2001:0db8:85a3:0000:0000:8a2e:0370:7334**.

2.1.3.2. Dominio

El dominio es una forma de identificar y ubicar computadoras en la web o una red privada. Esencialmente, es la dirección que se escribe en el navegador para visitar un sitio web.

La gracia de los dominios es proporcionar una forma fácil de recordar y un método sencillo para los humanos para interactuar con las direcciones IP's numéricas que utilizan las computadoras para identificarse entre sí en la red. Un dominio en el contexto de la internet es una forma de identificar y ubicar computadoras en la web o una red privada. Esencialmente, es la dirección que se escribe en un navegador para visitar un sitio web. Los dominios proporcionan una forma fácil de recordar para los humanos y así poder interactuar con las direcciones IP's numéricas que utilizan las computadoras para identificarse entre sí en la red, a este sistema se le conoce como "DNS".

Un ejemplo de una estructura de dominio puede ser: **www.entel.cl**, el cual se divide en 3 partes que son:

- **.cl:** es el dominio de nivel superior (TLD) o la parte más a la derecha de la dirección.
- **entel:** es el nombre de dominio elegido por la persona o la organización.

- **www:** es un subdominio, comúnmente utilizado para sitios web, pero puede modificarse o eliminarse para diferentes subdominios o aplicaciones.

Cada dominio debe ser único para evitar conflictos o confusión. Esto se gestiona a través de organismos de registro acreditados que aseguran que cada nombre de dominio sea exclusivo y registrado oficialmente.

Los dominios se utilizan para varios propósitos, incluyendo la configuración de sitios web, correos electrónicos personalizados, aplicaciones web y más. Estos pueden tener varios registros asociados, como registros A (para direcciones IPv4), registros AAAA (para direcciones IPv6), registros MX (para correo electrónico), etc.

Dominios de nivel superior (TLD): Hay varios tipos de TLD, incluyendo:

gTLDs(Dominios genéricos de nivel superior): .com, .net, .org, .info.
ccTLD (Dominios de código de país de nivel superior): .uk (Reino Unido), .de (Alemania), .us (Estados Unidos).
nTLDs (Nuevos dominios de nivel superior): .guru, .photography, etc.

Al registrar un dominio, típicamente se elige un nombre y un TLD que juntos forman la dirección completa de un sitio web o servicio en Internet.

2.1.3.3. Puertos

Los puertos son interfaces físicas o virtuales donde a través de ellas se pueden conectar dispositivos o establecer comunicaciones respectivamente.

Puertos Físicos: Son conectores o enchufes en el computador donde se conectan cables de dispositivos externos, algunos ejemplos pueden ser USB (Universal Serial Bus), HDMI (High Definition Multimedia Interface) o Ethernet.

Puertos Virtuales (Puertos de Software): Estos son identificadores numéricos utilizados en la capa de transporte del modelo OSI. Son utilizados por el sistema operativo y las aplicaciones para gestionar el uso de ciertos recursos y datos.

Existe un rango de puertos que pueden ser utilizados, estos van desde el 0 al 65535, estos de dividen en 3 rangos que son:

- **Puertos Conocidos (0-1023):** Generalmente para servicios y aplicaciones comunes como lo son rdp,ftp,ssh,http,https,smtp o smb por nombrar algunos.
- **Puertos Registrados (1024-49151):** Para aplicaciones de los usuarios o algunas no tan comunes
- **Puertos dinámicos o Privados (49152-65535):** Generalmente para uso temporal.

Esto de los rangos de puertos no son exclusivos y se puede asignar el puerto que queramos al servicios que nosotros estimemos convenientes, pero por convenio se suelen elegir los predeterminados para algunos servicios, donde utilizamos protocolos como http en el puerto 80, https en el puerto 443 o rdp en el puerto 3889.

2.1.3.4. Servicios

Los servicios en el contexto de la computación son programas o procesos que escuchan y responden a solicitudes en puertos de red designados. Estos servicios utilizan los puertos como puntos de acceso lógicos para la comunicación de red, permitiendo la interacción entre diferentes sistemas y aplicaciones.Podemos tener servicios web, correo electrónico, etc.

2.1.3.5. Protocolos

El protocolo se define como un conjunto de reglas y estándares los cuales definen como se transmiten los datos.

Algunos de los protocolos que van a ser expuestos como servicios que son interesantes para este trabajo son:

- **ftp (puerto 21):** Generalmente para servicios y aplicaciones comunes como lo son rdp,ftp,ssh,http,https,smtp o smb por nombrar algunos.
- **ssh (puerto 22):** Permite la transferencia de datos entre un cliente y un servidor a través de una comunicación cifrada entre 2 dispositivos.
- **telnet (puerto 23):** Este protocolo permite comunicación de texto bidireccional, aunque no encripta las conexiones.
- **http (puerto 80):** Protocolo de comunicación entre cliente y servidor(generalmente utilizado en navegadores web). Este no mantiene el envío de datos de forma cifrada a través de la red.
- **https (puerto 443):** Igual que http con la diferencia que utiliza SSL/TLS para encriptar la comunicación entre cliente y servidor, por lo tanto la comunicación se realiza a través de un canal cifrado.
- **smb (puerto 445):** Este corresponde a un protocolo cliente-servidor que controla el acceso a archivos y directorios enteros, así como a otros recursos de la red, como impresoras, routers o interfaces compartidas con la red. Tiende a ser un protocolo muy utilizado dentro de la red de una organización, esto debido a la cantidad de impresoras que utilizan en mayor medida.
- **rdp (puerto 3389):** Protocolo desarrollado por microsoft que proporciona conectividad al usuario mediante una interfaz gráfica para conectarse con otra computadora a través de la red.

2.1.4. NIST

El instituto nacional de estándares y tecnologías (NIST) es una agencia del departamento de comercio de los estados unidos y uno de los laboratorios de ciencias físicas más antiguos del mismo país. Esta agencia se dedica a establecer estándares e investigar en campos como la metrología, la ciberseguridad y el desarrollo de tecnologías.

En este trabajo veremos la parte de ciberseguridad del NIST, donde interesa específicamente lo que es el CPE.

2.1.4.1. CPE (Common Platform Enumeration):

Este es un método estandarizado para describir e identificar tipos de aplicaciones, sistemas operativos y dispositivos de hardware presentes entre los activos informáticos de una empresa.

Los CPE suelen ser utilizadas en la gestión de vulnerabilidades donde podemos identificar tecnologías, productos, sistemas operativos y aplicaciones de cada uno de los activos y así poder utilizar esta información estandarizada al gestionar las vulnerabilidades asociadas de forma más eficiente. Se hace uso de los CPE para poder encontrar vulnerabilidades asociadas, esto utilizando lo que se conoce como las common vulnerabilities exposures (CVE).

El formato que utilizan estos CPE es el siguiente:

cpe:/<part>:<vendor>:<product>:<version>:<update>:<edition>:<language>

Además cada una de sus componentes corresponden a:

- **part:** Puede tener 3 valores los que son a para aplicaciones, h para plataformas de hardware y o para sistemas operativos.
- **vendor:** El nombre del fabricante o proveedor del componente.
- **product:** El nombre específico del producto.
- **version:** La versión específica del producto.
- **update:** Cualquier actualización o revisión del producto.
- **edition:** La edición del producto, si aplica.
- **languaje:** El idioma en el que está disponible o se utiliza el producto.

2.1.5. CVE (Common vulnerabilities and exposures)

El CVE es un sistema para poder identificar, definir y catalogar las vulnerabilidades que se informan en ciberseguridad de forma pública.

La ciberseguridad nacional de los estados unidos FFRDC que es operado por The MITRE Corporation, mantiene el sistema con financiamiento de la División Nacional de Seguridad Cibernética del Departamento de Seguridad Nacional de los Estados Unidos.

Una vez que las vulnerabilidades son descubiertas, se asignan y se publican por organizaciones de todo el mundo que se han asociado con el programa CVE.

El sistema CVE mantiene un diccionario público de vulnerabilidades, aquí es donde nacen los cve record que son identificadores para cada una de las vulnerabilidades, estos tienen el siguiente formato:

CVE-AAAA-NNNN

Donde **AAAA** corresponde a el año en el que se reportó la vulnerabilidad y **NNNN** corresponde a un número secuencial.

Los profesionales de las **TI** y ciberseguridad utilizan cve record para asegurarse que están hablando del mismo tema y así coordinar sus esfuerzos para priorizar y abordar las vulnerabilidades expuestas.

2.1.6. Contexto actual

La ciberseguridad mantiene una escasez de profesionales a nivel mundial, esto debido a que se han producido distintos tipos de **hackeos** a diferentes organizaciones a lo largo y ancho del planeta. Es gracias a estos acontecimientos que se ha comprendido la importancia de tener un equipo especializado en el área de la ciberseguridad en un mundo donde la tecnología avanza a pasos agigantados.

Chile no es la excepción de estos casos, recientemente se tiene el ataque que sufrió el proveedor de internet GTD, el cual posterior a 3 meses del ataque por **ransomware** sufrido, este sigue con repercusiones.

Gracias al contexto anteriormente descrito es que el gobierno chileno no se ha quedado atrás y ha decidido tomar cartas en el asunto de la ciberseguridad y la privacidad de los datos. Es por esto que recientemente se ha redactado el nuevo decreto de ciberseguridad en el país (1). Este menciona sobre estrategias para mejorar la ciberseguridad, la coordinación nacional e internacional y el fomento en la industria como la investigación científica. Donde mayoritariamente caen sus esfuerzos es en la protección de datos personales y la prevención de delitos informáticos.

En el caso de la protección de datos personales, esta habla de la necesidad de la implementación de una ley marco de ciberseguridad y protección de datos personales. Además busca generar capacitación para todos los funcionarios públicos en hábitos y medidas básicas de seguridad digital, esto con el fin de proteger la información digital de los ciudadanos.

Para el caso de los delitos informáticos, se hace una referencia a un aumento, donde en 2017 hubo un aumento del 66 % y esté aumentó a un 101 % en 2021. Gracias a esto sale la prevención de delitos informáticos, donde se hace énfasis en considerar necesidades particulares de seguridad en grupos particulares como lo son mujeres, niñas, niños, adolescentes, adultos mayores y disidencias sexogenéricas. Estos grupos tienen una mayor probabilidades de ser vulnerados en el ciberespacio. Aparte este decreto menciona la importancia de identificar y corregir inequidades en el acceso y uso del ciberespacio, se puede destacar el acceso equitativo al ciberespacio, donde se busca apoyar a los sectores de la sociedad que no tienen un acceso equitativo a internet, ayudando a proveer mejor infraestructura tecnológica en áreas desatendidas.

Todo lo anteriormente expuesto es prueba de que la ciberseguridad tiene una gran relevancia dentro de la sociedad actual. Aquí es donde aparece la pregunta de: ¿Como se vulnera algún sistema o activo?, respondiendo a esta pregunta podemos tener una mala configuracion de algún sistema, mal uso del mismo o vulnerabilidades expuestas ya sea en Sistemas operativos, protocolos, programas, etc. En este surge la duda de que tan comunes son las vulnerabilidades, en este caso es donde podemos acudir al NIST.

Cada año el NIST reporta un aproximado de 26000 vulnerabilidades por año (500 en promedio a la semana), esto nos da un gran indicio de la importancia de conocer los activos que se encuentran dentro de la organización. Debido a la gran cantidad de vulnerabilidades que van saliendo semana a semana, si no conocemos o tenemos poca información sobre lo que hay en la red, se abre una brecha para explotar vulnerabilidades y hacer daño a la reputación u operación de la organización.

Se sabe por reportes de finales de 2023 por parte de la empresa Qualsys (especializada en el cloud security) que las vulnerabilidades altas y crítica donde ya se tiene conocimiento de una POC, llegan a ser explotadas un 25 % de ellas en el primer día de su liberación. (2).

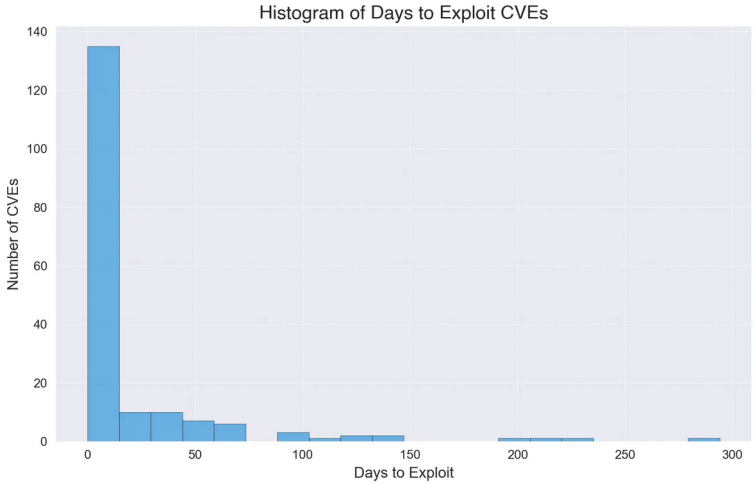


Imagen sacada de https://blog.qualys.com/vulnerabilities-threat-research/2023/12/19/ 2023-threat-landscape-year-in-review-part-one

Figura 2.3: Tiempo en que se demora en explotar una vulnerabilidad al tener una POC disponible

Hoy en día en la ciberseguridad hay un proceso dentro de las organizaciones del que se habla bastante que es la gestión de riesgo.

La gestión del riesgo en ciberseguridad es el proceso de identificar, evaluar y mitigar los riesgos asociados a la seguridad de la información en el ámbito digital. Incluye la identificación de amenazas potenciales, la evaluación de la vulnerabilidad de los sistemas frente a estas amenazas, y la implementación de estrategias para reducir el riesgo.

Es aquí en donde se llega al concepto de **riesgo**, el cual se refiere a la posibilidad de sufrir daños o pérdidas debido a amenazas digitales o ataques en el ciberespacio. Dentro de estas amenazas podemos considerar **malware**, robo de datos, **phishing**, **smishing**, entre otros ataques cibernéticos. Estos pueden no solamente hacer daño directo en algún sistema o en la operación misma del negocio, sino que también a la reputación de la organización.

Uno de los grandes problemas que se generan dentro de la gestión de riesgo es el cálculo del riesgo. Esto debido a que no existe una fórmula para poder calcularla, sino que tiene que ver más con un cálculo en base a variables cualitativas, es por esto que no existe una fórmula mágica para hacer el cálculo del riesgo.

Hay pocos estudios que hacen referencia al cálculo del riesgo mediante diversos métodos, por ejemplo en el estudio **A Probabilistic Analysis of Cyber Risks** (3), se propone un método probabilístico que integra datos existentes y la extensión de estos para evaluar probabilidades de nuevos escenarios de ataque, donde se emplea un análisis estadístico de incidentes pasados, como la pérdida o robo de laptops, y se expande a un dominio probabilístico para cubrir amenazas futuras. Otro estudio asociado al mismo tema es **Cyber Risk Assessment and Mitigation (CRAM) Framework Using Logit and Probit Models for Cyber Insurance** (4), este estudio habla del uso de métodos estadísticos para predecir al probabilidad de ataques cibernéticos y sus consecuencias financieras, se hace uso de datos históricos para modelar la frecuencia y severidad de los ataques, de esta forma se ayuda a la organización a decidir entre diversas estrategias de mitigación, así este estudio busca equilibrar la protección contra amenazas con la eficiencia en la asignación de recursos . Por último con respecto al riesgo tenemos **Modelling and Management of Cyber Risk** (5), el cual enfatiza la importancia del comportamiento humano en el riesgo y que los riesgos cibernéticos son muy diferentes en comparación con otros riesgos operativos, esto lo consigue utilizando modelos de riesgo operativo para proporcionar estimaciones de riesgo consistentes.

Para poder hacer este cálculo del riesgo es importante conocer cuales pueden ser las posibles brechas que pueda tener la organización, para esto debemos tener información que puede ser recolectada en la etapa

de reconocimiento sobre los activos.

En algunos documentos se puede evidenciar la importancia de la etapa de reconocimiento en la ciberseguridad en general.

Tenemos el caso de **Cyber Reconnaissance: An Alarm before Cyber Attack** (6), en este documento se realza la importancia del port scanning y del OS fingerprint dentro de un proceso de reconocimiento, para esto se sugiere tener herramientas que puedan detectar estas acciones de un posible adversario y realza lo crítico que es esta etapa para un adversario. En el estudio **Deceiving Network Reconnaissance using SDN-based Virtual Topologies** (7) se habla de la importancia de usar topologías virtuales para engañar a los atacantes, haciéndoles creer que están viendo la estructura real de la red, cuando en vez de eso están viendo una simulación, esto hace referencia a la importancia de la fase de reconocimiento para un atacante al momento de buscar una brecha en la red y así poder vulnerarla. Un estudio parecido es **Cyber Deception: Virtual Networks to Defend Insider Reconnaissance** (8) el cual habla del mismo tema de ocultar la superficie de ataque utilizando redes definidas por software (SDN), donde también se busca minimizar el impacto en el tráfico mientras se máximiza la eficiencia de la defensa.

Frente a todos los casos anteriormente vistos es posible evidenciar la importancia de la reducción de la visibilidad de la superficie de ataque, la prevención y el enfoque de los esfuerzos en esta etapa por parte de la ciberseguridad defensiva.

3 | Análisis Preliminar

3.1. Estado del proyecto

Se han revisado distintos estudios a lo largo del proyecto de tal forma de comprender cual la dimensión del problema. Se mantuvieron reuniones semanales con Fernanda Mattar y Daniel Hernandez por parte de ENTEL. Donde por su parte nos brindaron el apoyo correspondiente, nos fueron guiando en el proceso de desarrollo de la solución y planteando los objetivos constantemente de tal forma de llegar a un buen resultado. Además se han realizado distintos cambios a como se planteó el desafío al comienzo del proyecto, esto debido a distintas complicaciones que se fueron efectuando en el proceso.

Algunos de los cambios más importantes realizados fueron:

1. El cálculo del vector de ataque utilizando Inteligencia artificial se ha dejado de lado y se opta por otra forma de cálculo.
2. El reconocimiento se realizará en una lista de IP's o dominios entregadas para poder ser escaneadas.
3. El proceso de encapsulación se dejó en segundo plano, esto para disponer de los esfuerzos en la generación de la memoria.

3.2. Endpoints entregados

Por parte de la organización se recibió una lista de IP's y dominios a escanear de tal forma de poder generar la memoria dentro de los activos de la organización, así como para hacer pruebas de reconocimiento sobre ellas. Estas ip's y dominios corresponden tanto al programa de "BugBounty" como a IP's internas (solo tienen visibilidad dentro de la red corporativa) y externas de la organización.

3.3. Consideraciones al escanear

Se debe tomar en consideración que reconozca si un activo se va a pasar como dominio o se va a pasar como una IP a escanear así como también tomar en consideración si se encuentra disponible o no el activo al momento de ser escaneado, al igual que tener en consideración cual será el tiempo máximo de escaneo de un activo para ir con el siguiente.

Es estrictamente necesario el poder tener un módulo para poder realizar escaneos en paralelo, esto tomando en consideración que la organización posee dentro del orden de +10000 activos que escanear. Es por esto que el módulo que le da agilidad al proceso del escaneo debe ser prioritario y debe tener todas las medidas contra errores en consideración al momento de ejecutarse dentro del dominio de la organización. Se debe informar que el host que está realizando los escaneos es conocido y que no sea un inconveniente para el **SIEM** de la organización ni tampoco para las distintas medidas defensivas que tenga la organización, así

como también el registro de todo lo que realice el programa en caso de necesitar información operacional al momento de encontrarse con algún problema en el funcionamiento del mismo.

4 | Diseño de la solución

En el diseño de la solución se revisará la estructura general del sistema, considerando los distintos componentes que lo conforman, las características de cada una de estos y la conexión que existe entre ellos. Esto desde el momento en que se genera el proceso de reconocimiento de la superficie entregada (IP'S y Dominios), el como almacenar estos datos para su posterior envío a través del sharepoint coorporativo donde finalmente será procesado utilizando business intelligence para una mejor presentación de los datos generados.

4.1. Entel-AutoRecon

Este módulo es la base de toda la operación. Cumple con la función principal de hacer conexión con las demás componentes del sistema.

4.2. Interfaz CLI

Se debe generar una interfaz CLI donde el usuario pueda interactuar con la herramienta y así ejecutar ciertas acciones o revisar la configuración del sistema.

4.3. Memoria

Con respecto a la memoria, se debe buscar una forma de poder procesar los datos, ordenarlos y finalmente normalizarlos para poder ser enviados posteriormente y así evitar casos de frontera que puedan interrumpir con la operación del proceso.

4.4. Reconocimiento y enumeración

Este módulo debe ser capaz de realizar los escaneos para cada una de las IP's y dominios entregados de tal forma de extraer información ya sea sobre los servicios que están disponibles en el host al igual que información referente al **OSBanner** y demás. Esta etapa es crítica dentro del proceso, debido a que gracias a la información obtenida vamos a poder generar la memoria que quedará guardada.

4.5. Decisión

Aquí en donde se recibe la data normalizada de tal forma de poder estimar el riesgo y generar los vectores de ataque correspondientes.

4.6. Business Intelligence

Este módulo va a tener la función de organizar la data compartida por la herramienta de Entel-AutoRecon. La data recolectada debe tener una vista dinámica e intuitiva de utilizar, esto con el fin de ser una utilidad para el equipo defensivo de la organización.

4.7. Diagrama de contexto

En la figura 4.1 se muestra un diagrama con las interconexiones entre los distintos módulos que componen el sistema. Este representa una forma muy sencilla y minimalista de ver como se comporta el sistema en general

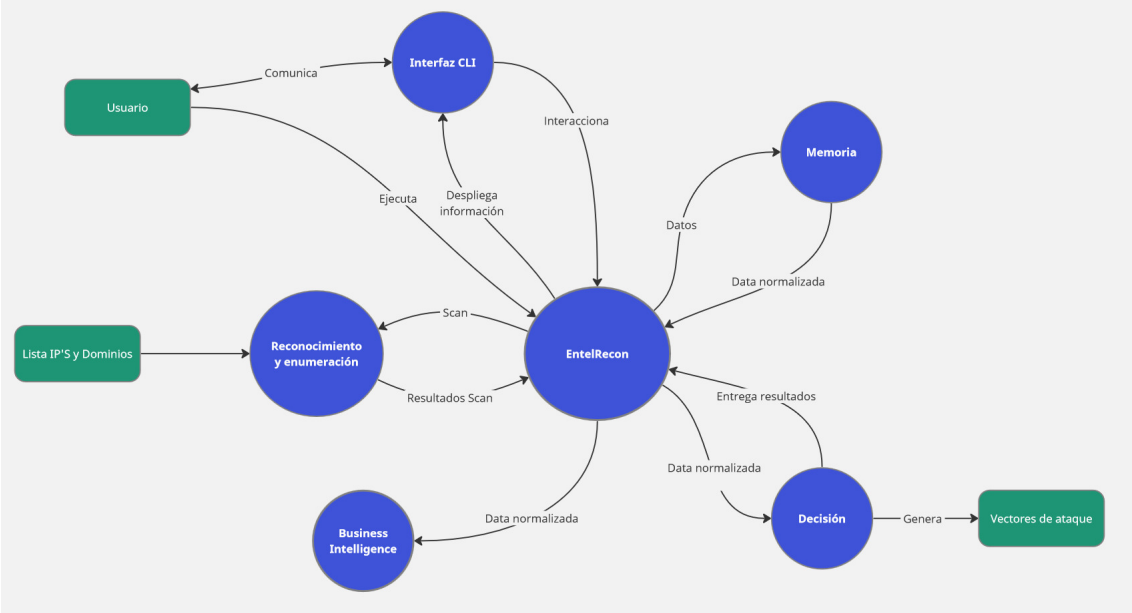


Figura 4.1: Diagrama del esquema general del sistema

4.8. Descripción de Componentes

En esta sección se presentará una descripción detallada de los diversos módulos que componen el sistema. Se incluirán tablas para ilustrar las diversas características consideradas en la implementación de cada módulo.

4.8.1. Entel-AutoRecon

Propósito	Este módulo tiene como propósito proveer al usuario un punto donde poder configurar las funcionalidades del módulo de reconocimiento y enumeración, como también tener un punto de integración y comunicación con la Interfaz CLI.
Alcance	Integrador de los módulos.
Dependencias	**Unificador:** Este módulo depende del host que lo esté corriendo, debe ser corrido con permisos de administrador en el sistema y este debe satisfacer los programas y librerías de NVDlib, Python, Python-questionary, Pandas, Openxlpy, Nmap.
Supuestos	El usuario debe ser capaz de entender cómo utilizar la línea de comando para correr el programa..
Restricciones	Minimalismo, Fácil Entendimiento.
Estructura General	Integrador de los módulos de Interfaz CLI, Reconocimiento y enumeración, Memoria, Decisión y Business Intelligence.

Tabla 4.1: Descripción del módulo de Entelrecon.

4.8.2. Interfaz CLI

Propósito	Proveer al usuario un panel de control por medio de una linea de comandos para poder accionar las operaciones automatizadas del sistema, realizar reconocimiento activo de forma manual y exportar la superficie de ataque de los activos de la red. Además se espera que exporte los vectores de ataque.
Alcance	Linea de comandos.
Dependencias	El módulo está manejado por EntelRecon, por lo tanto si se puede ejecutar EntelRecon, entonces corre Memoria. Para esto es necesario python-questionary.
Supuestos	El usuario entiende el proceso de reconocimiento de red .
Restricciones	Minimalismo y facilidad de uso.
Estructura General	Una linea de comandos básica con mensajes de ayuda al usuario sobre lo que puede hacer .

Tabla 4.2: Interfaz CLI.

4.8.3. Memoria

Propósito	Almacenar y procesar en un formato estándar tanto los resultados obtenidos por el reconocimiento como los entregados por la red corporativa del cliente.
Alcance	De acceso gerencial y reservado para la interfaz de cliente.
Dependencias	El módulo está manejado por EntelRecon, por lo tanto si se puede ejecutar EntelRecon, entonces corre Memoria. Para esto es necesario Python, Pandas, Openxlpy.
Supuestos	El cliente proveerá de la infraestructura necesaria para operar el gestor de base de datos.
Restricciones	La estructura de datos debe estar normalizada.
Estructura General	Funciones automáticas que realizan operaciones CRUD sobre una base de datos relacional, estructurada como CSV y orientada a realizar operaciones columna.

Tabla 4.3: Descripción del módulo de memoria.

4.8.4. Reconocimiento y enumeración

Propósito	Realizar pruebas de reconocimiento activo automáticamente.
Alcance	Integrado en el software del cliente, permite realizar el escaneo a múltiples activos a la vez.
Dependencias	El módulo está manejado por EntelRecon, por lo tanto si se puede ejecutar EntelRecon, entonces corre Memoria. Para esto es necesario Python, Nmap.
Supuestos	El usuario puede instalar paquetes usando el gestor de paquetes de python pip.
Restricciones	Debe operar periódicamente sobre una gran cantidad de activos
Estructura General	Proyecto Entelrecon programado para considerar sólo aquellos sub-módulos y plugins necesarios para el reconocimiento activo.

Tabla 4.4: Descripción del módulo de Aplicación web.

4.8.5. Decisión

Propósito	Determinar los vectores de ataque disponibles y recomendar el mejor.
Alcance	Cálculo de los Vectores de Ataque.
Dependencias	Para esto es necesario NVDlib.
Supuestos	El módulo recibe CSV normalizado.
Restricciones	Debe presentarse al usuario de manera que facilite el pentesting
Estructura General	Módulo de Python 3 que integra las funcionalidades de Pandas

Tabla 4.5: Descripción del módulo de decisión.

4.8.6. Business Intelligence

Propósito	Proveer al usuario de informes en PowerBI que permitan visualizar la superficie de ataque de los activos de la red. Además se espera lo mismo para los vectores de ataque.
Alcance	Informe de PowerBI gratuito (sin publicación) .
Dependencias	PowerBI.
Supuestos	El usuario entiende sobre conceptos de ciberseguridad que se expondrán en la interfaz y usa la herramienta sabiendo lo que busca.
Restricciones	Minimalismo, Fácil Entendimiento
Estructura General	Informes de PowerBI con gráficos, tablas, filtros y datos de interés, tanto para mostrar la superficie de ataque como para filtrar por Vectores de Ataque disponibles

Tabla 4.6: Descripción del módulo de Business Intelligence.

4.9. Requisitos del sistema

El objetivo hasta este punto se hace bastante claro, se requiere hacer un escaneo masivo de los activos de la organización de tal forma de disponibilizar la información obtenida a través de esta herramienta hacia la gerencia. En el proceso surgieron bastantes inconvenientes y considerando el tiempo acotado que se tuvo para realizar el desarrollo de esta solución, es que hubo algunos cambios en los requisitos funcionales del proyecto. Estos cambios se centraron principalmente en el uso de Inteligencia artificial al momento de hacer el cálculo del riesgo de los activos, en la encapsulación de la herramienta, la ejecución periódica del escaneo masivo de los activos de la organización (esto por simplicidad). Además de agregar una interfaz gráfica para poder visualizar la data obtenida por la herramienta, esto utilizando PowerBI.

4.9.1. Requisitos Funcionales

- **RF1** : El sistema permite buscar activos por su ip/hostname.
- **RF2** : El sistema permite crear un activo nuevo.
- **RF3** : El sistema realiza escaneos automáticos indicando activos como objetivo.
- **RF4** : El sistema almacena de manera estructurada los resultados de los escaneos realizados.
- **RF5** : El sistema permite listar los activos ya presentes en la base de datos.
- **RF6** : El sistema exporta los datos útiles Normalizados en CSV.
- **RF7** : Para un activo dado, el sistema presenta todos los vectores de ataque que aplican.
- **RF8** : Los datos útiles Normalizados se pueden cargar automáticamente a un Sharepoint de Entel.
- **RF9** : Los datos útiles Normalizados se pueden visualizar en un informe de PowerBI.

4.9.2. Requisitos No Funcionales

- **RNF1** : Los escaneos no provocan un impacto negativo en los activos objetivo.
- **RNF2** : Los escaneos se realizan optimizando tiempo y recursos, evitando realizar pruebas redundantes.
- **RNF3** : La visualización de resultados es clara y simple.
- **RNF4** : El almacenamiento de datos es seguro y rápido.
- **RNF5** : La interfaz de línea de comandos entrega indicaciones y respuestas en términos de la metodología Entel.
- **RNF6** : La solución debe presentar toda la información necesaria para el usuario en una sola interfaz.

4.9.3. Requisitos de Interfaces

A continuación en la tabla 4.7 se describen los eventos que son detonados mediante la interacción del usuario con el sistema a través de la interfaz web, y por otro lado, se describen las respuestas del sistema a dichos eventos. Son estos eventos los que controlan el flujo del sistema completo.

Evento	Descripción	Iniciador	Parámetros	Respuesta
Ejecución sin parámetros argv	Se ejecuta el sistema sin argumentos en la linea de comando	$ sudo python3 main.py	-	Notificación de carga
Ejecución en modo manual	Se ejecuta el sistema con el argumento manual	$ sudo python3 main.py manual	-	Modo manual
Ejecución en modo auto	Se ejecuta el sistema con el argumento auto	$ sudo python3 main.py auto	p | all, NS, FNS, WS, ssh, FTP, smb, rdp	Ejecución de secuencia preconfigurada para scan automático
Ejecución de modo update	Se ejecuta el sistema con el argumento update	$ sudo python3 main.py update	-	Se sincroniza la base de datos con la nube
Ejecución en modo info	Se ejecuta el sistema con el argumento info	$ sudo python3 main.py info	-	Muestra la configuración actual
Ejecución en modo login	Se ejecuta el sistema con el argumento login	$ sudo python3 main.py login	-	Genera el código de autenticación de dispositivo de microsoft
Ejecución en modo excel	Se ejecuta el sistema con el argumento excel	$ sudo python3 main.py excel	-	Exporta una instantanea de la base de datos local a formato excel
Ejecución en modo help	Se ejecuta el sistema con el argumento help	$ sudo python3 main.py help	-	Menú de ayuda extendido

Tabla 4.7: Descripción de eventos e interacciones en la CLI.

Cabe mencionar que para el caso de el evento de ejecución en modo auto, el parámetro p corresponde a la ejecución en paralelo y los demás parámetros corresponden a las pruebas a ejecutar. Si se omite el uso de p, entonces se ejecutará en modo secuencial.

4.10. Propuesta de solución

Con respecto al stack de tecnologías que se utilizará en la propuesta, se basa en la continuidad operacional del software y el soporte que puede llegar a necesitar a largo plazo.

Es por esto que se ha decidido utilizar como base “Python” como lenguaje de programación principal. Esta elección va acompañada de la herramienta “nmap” para llevar a cabo los escaneos correspondientes a los distintos activos. Tanto para la interfaz visual como para el manejo y análisis de los datos, se emplearán diversas librerías como “Pandas”, “Openxlpy” y “python-questionary”.

En este caso, la elección de utilizar nmap para los escaneos se debe a su amplio uso y buenos resultados que ha dado en el mundo de la ciberseguridad, al igual que el sólido soporte que ofrece para distintos sistemas operativos. Esto contribuirá a que el código se mantenga con menos modificaciones a lo largo del tiempo.

Para la lectura de archivos “Excel”, se utilizará Openxlpy debido a su facilidad de implementación y uso. Pandas se empleará para la manipulación de la base de datos y la exportación a Excel , en caso de ser necesario. Por último, Python-questionary se utilizará para crear una interfaz en línea de comandos, lo que facilitará la interacción con la aplicación de una manera más sencilla.

Además como almacenamiento de datos para poder hacer uso de la memoria generada por parte de los escaneos realizados es que se utilizará “Sharepoint” de microsoft al igual que para poder generar la interfaz de la memoria es que se utilizará “Power BI” para tenerla de una forma más ordenada y procesada para poder realizar consultas simples que sean de ayuda para los especialistas de la organización.

En la figura 4.2 se muestra el stack de tecnologías utilizadas en el proyecto.

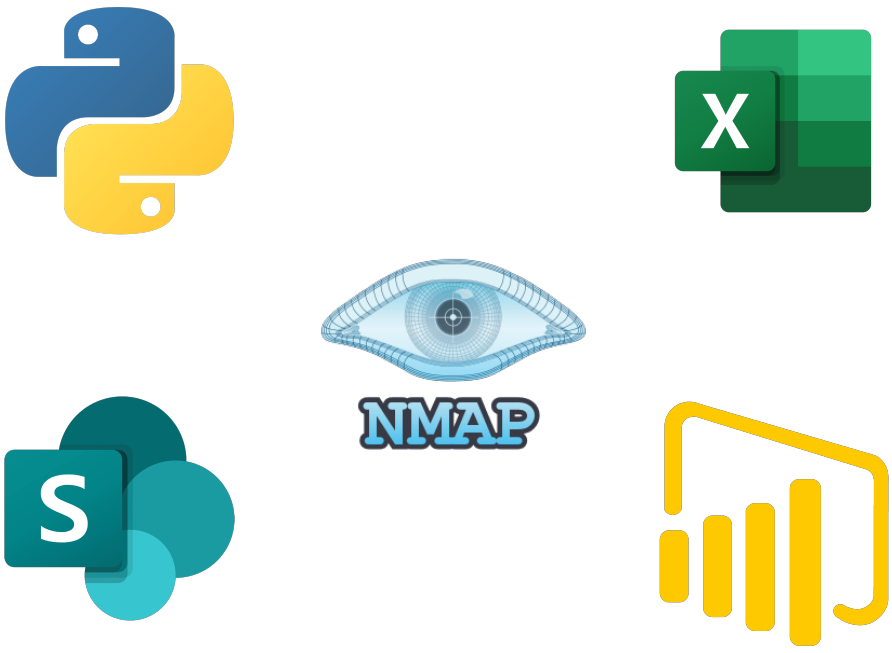


Figura 4.2: Stack de tecnologías

4.10.1. Decisión del stack de tecnologías

En esta sección se explicará la decisión del porqué de cada uno de estas tecnologías a usar.

- Python: Se tomó como decisión utilizar python debido a su simplicidad debido al uso que se le ha dado a lo largo del plan de enseñanza dentro de la universidad y su versatilidad al momento de resolver problemas, esto debido a la gran comunidad que existe para esta tecnología, donde la cantidad de librerías existentes para esta tecnolgía lo hace versatil para el problema a resolver dentro del proyecto. Si bien es cierto existen otras tecnologías que tienen otras ventajas como es el caso de RUST, el cual prioriza el código seguro desde el punto de vista de la seguridad de la información, la decisión de esta tecnología se basó en el corto tiempo a emplear para desarrollar una herramienta funcional y con un gran valor.
- Excel: Esta herramienta es bien conocida por ser usada en las empresas, en este caso fue decidido el uso de este debido a que puede ser utilizado para recuperar la información sobre los scans a realizar y de esta forma procesar esta información de la forma que quiera ser procesada o revisada por el operario, por solicitud de gerencia o por alguna necesidad investigativa.
- Nmap: Es bien conocida dentro de la industria la versatilidad que tiene nmap para realizar scans de red y hasta hacer pruebas sencillas de penetración. Es por esto la elección de esta tecnología, al igual que su capacidad de multiplataforma, donde es posible ejecutarla desde muchos sistemas operativos distintos, aumentando la posibilidad de poder cambiar el sistema operativo desde donde se ejecute la herramienta a desarrollar y así postponer su obsolescencia a lo largo del tiempo.
- Sharepoint: Tecnología elegida por pedido de la organización para aprovechar las licencias utilizadas por ellos al igual que el almacenamiento y compartición de la información de forma segura, donde esta información generada es guardada en el sharepoint que está en la nube y que tiene permisos de lectura y de acceso restringidos a las necesidades de la organización.
- PowerBI: Dentro de las formas para poder tener visualización e interacción de la data por parte del usuario una de las tecnologías que está siendo muy utilizada, tiene un buen desempeño y es fácil de utilizar y aprender, teniendo buenos resultados además de la generación de dashboards interactivos es que usamos esta tecnología, si bien es cierto, existen otras formas de poder generar gráficas aplicando Business Intelligence. Otras tecnologías como 3Djs o grafana, pueden ser utilizadas, pero requieren una curva de aprendizaje mayor, además de no tener tan buena compatibilidad como lo tiene powerBI en microsoft, de esta forma se puede aprovechar la infraestructura y licencias que tiene la organización al igual que la seguridad que el ecosistema de microsoft ofrece.

4.10.2. Ambiente de desarrollo

A continuación se señalan los requisitos de software y hardware que serán necesarios para el funcionamiento del sistema.

4.10.2.1. Hardware de Desarrollo

- Procesador: Se recomienda un sistema Pentium 4 a 2 [GHz] como mínimo para un sistema de escritorio.
- Memoria RAM: Para una instalación sin entorno de escritorio, se necesitan al menos 256 megabytes de RAM, aunque se recomiendan 512 megabytes. Con entorno de escritorio, se necesitan al menos 512 megabytes de RAM y se recomiendan 2 gigabytes.
- Espacio en disco: Se necesitan 2 gigabytes para una instalación sin entorno de escritorio y 10 gigabytes para una instalación con entorno de escritorio. Para una operación suave del sistema GNU/Linux, se recomienda asignar al menos 200MB para la partición /var.

4.10.2.2. Software de Desarrollo

- Python 3
- Nmap
  - Pandas
  - Questionary
  - Beautifulsoup4(bs4)
  - Openpyxl
  - Pip
  - Requests
  - Numpy
  - Halo
  - Nvdlib
  - Setuptools
  - Threading

5 | Implementación

En este capítulo se presenta la implementación realizada a lo largo del proyecto, considerando los distintos componentes, el diseño de estos y las distintas tecnologías, librerías y herramientas utilizadas.

En la figura 5.1 se ilustra el diagrama temporal que describe las interacciones entre el usuario, el sistema y los distintos módulos que lo componen para así realizar un scan.

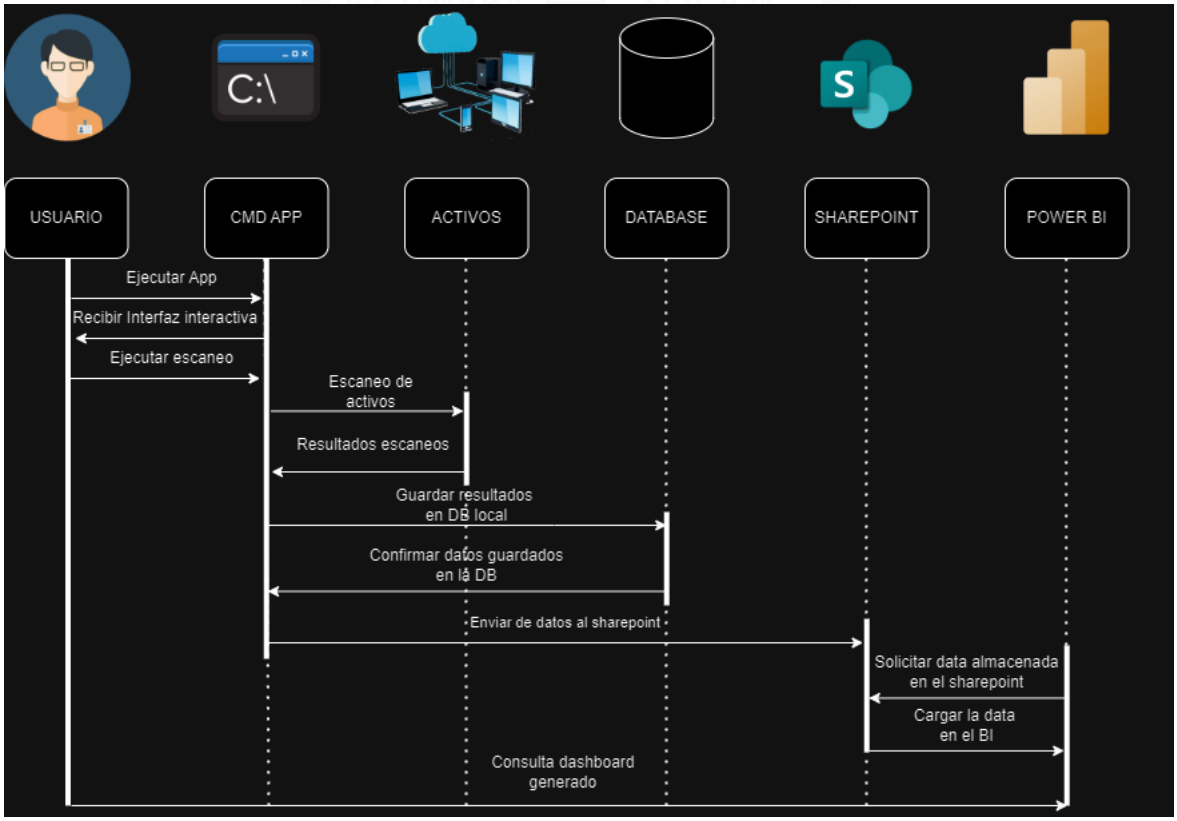


Figura 5.1: Diagrama de secuencia

A continuación se describen en rasgos generales las etapas que existen en la interacción presente en la figura 5.1 recién mencionada:

- **1.- Ejecutar App** : Mediante la **CMD**,el usuario ejecuta el programa.
- **2.- Recibir Interfaz interactiva** : El programa se ejecuta y devuelve una interfaz interactiva al usuario para poder ejecutar la acción que el usuario estime conveniente.

- **3.- Ejecutar escaneo** : El usuario interactúa con la interfaz entregada y realiza un escaneo de red con los parámetros y características entregadas.
- **4.- Escaneo de activos** : El módulo de Reconocimiento y enumeración es el encargado de realizar los escaneos y pruebas a los distintos activos entregados por el usuario a la Interfaz CLI del programa EntelRecon .
- **5.- Resultados escaneos** : El módulo de reconocimiento y enumeración recibe la información de los activos escaneados.
- **6.- Guardar resultados en DB local** : A medida que el módulo de reconocimiento y enumeración va recibiendo los resultados de los escaneos, estos se van guardando en la base de datos local.
- **7.- Confirmar datos guardados en la DB** : Se confirma que se guardaron los datos en la DB local.
- **8.- Enviar datos al sharepoint** : Una vez que se guarda toda la data en la base de datos local, se procede a enviar esta data al sharepoint.
- **9.- Solicitar data almacenada en el sharepoint** : El módulo de Business intelligence pide la información que se encuentra en el sharepoint.
- **10.- Cargar la data en el BI** : El sharepoint hace entrega de la data solicitada por el módulo de Business intelligence.
- **11.- Consultar dashboard generado** : Finalmente el usuario consulta al módulo de Business intelligence el reporte gráfico sobre la data generada.

5.1. Arquitectura utilizada

En esta sección se describen los módulos mencionados en la sección 4.8.

5.1.1. Módulo de Entel-AutoRecon

Para el módulo de EntelRecon se utilizaron las siguientes librerías NVDlib, Python, Python-questionary, Pandas, Openxlpy, Nmap..

El módulo de EntelRecon funciona como un módulo integrador y unificador de los módulos de la Interfaz CLI, Reconocimiento y enumeración, Memoria, Decisión y Business Intelligence.

El motivo de este módulo es ser un ente unificador para los demás módulos, con esto reducimos la complejidad del uso de la herramienta, además podemos tener un módulo donde se pueda configurar las distintas funcionalidades del módulo de reconocimiento y enumeración.

5.1.2. Módulo de Interfaz CLI

Para el desarrollo del módulo de Interfaz CLI tomamos como foco principal las necesidades del operador de la herramienta, al igual que el consumo de recursos por parte del host que esté corriendo la aplicación.

Para esto se tomó la decisión de hacer un programa interactivo utilizando la interfaz de linea de comando, la cual tiene un bajo costo computacional.

A continuación se hará mención del funcionamiento de la interfaz CLI y las distintas opciones que pueden llegar a ser utilizadas. Para poder hacer la ejecución de la interfaz gráfica es necesario ejecutar Entel-AutoRecon en modo manual, dando como resultado la Figura 5.2

Figura 5.2: Vista ejecución modo manual

Aquí tenemos varias opciones a ejecutar para el modo manual en la interfaz CLI.

5.1.2.1. Print useful info

Esta opción está diseñada para proveer información útil para el operador de la herramienta, el resultado de la ejecución de esta opción se ve reflejada a continuación.

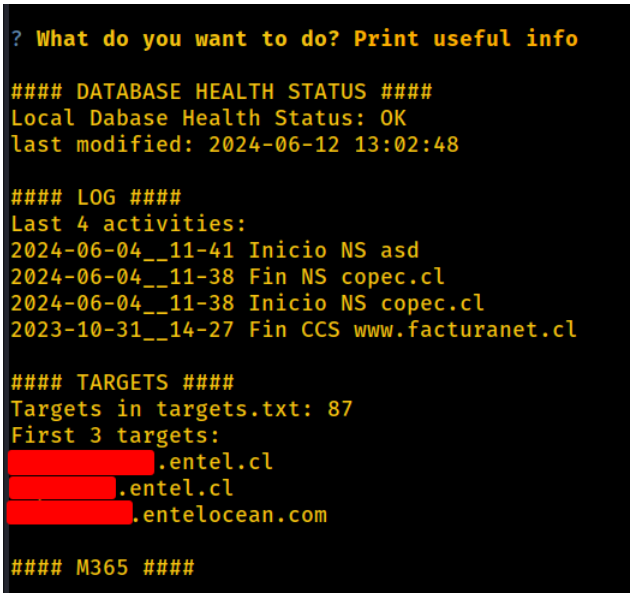


Figura 5.3: Vista opción print useful info

Esta entrega información útil para el operario de la aplicación, en este caso se entrega información sobre el estado de la base de datos y la última modificación que fue realizada, además de los últimos logs guardados en el computador anfitrión correspondiente a los escaneos e información asociada al archivo targets.txt que es el archivo donde anotamos los dominios y las IP que queremos escanear al momento de efectuar un escaneo automático o un escaneo masivo.

5.1.2.2. Attack

Esta opción está diseñada para proveer una interfaz donde se puedan realizar escaneos de forma manual, el resultado de la ejecución de esta opción se ve reflejada a continuación.

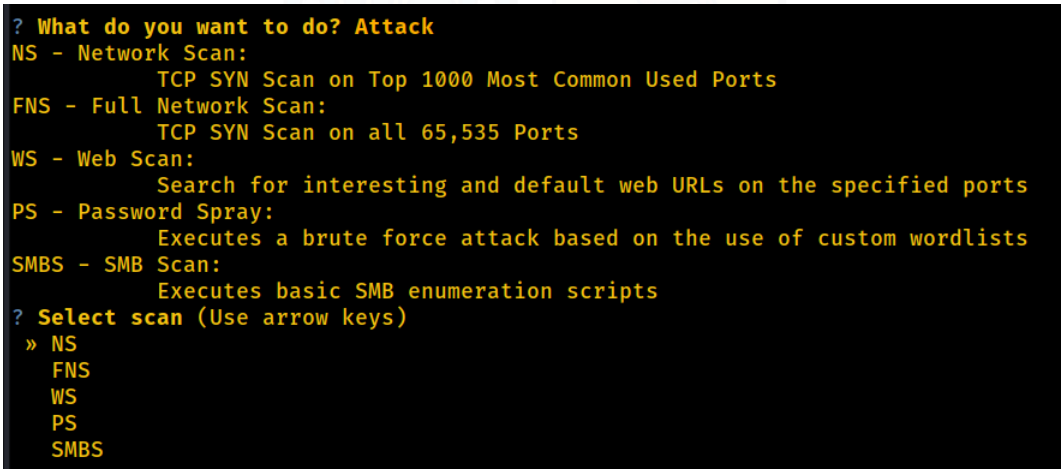


Figura 5.4: Vista opción Attack

Como se aprecia, se puede observar ver que se despliega un pequeño mensaje que entrega información sobre que es lo que realiza cada uno de los escaneos a los que se tiene opción de elegir, dentro de las opciones tenemos: el Network Scan que realiza un escaneo a los 1000 puertos más conocidos, el Full Network Scan que realiza un escaneo pero a los 65535 puertos, Web Scan que realiza un escaneo a unos puertos específicos donde es común que se alojen servicios web, Password Spray que corresponde a un tipo de ataque que realiza un ataque de fuerza bruta utilizando una lista personalizada de contraseñas o de usuarios para realizar el ataque, SMB Scan que lo que realiza es una ejecución de pruebas básicas de enumeración para el servicio SMB.

Para entender un poco más a detalle el funcionamiento de cada uno de estos escaneos es que se utilizará como ejemplo el Network Scan (NS), al elegir la opción de “one” se tiene la opción de entregar una ip/hostname/domain para realizar el escaneo de red, la ejecución de esta funcionalidad se muestra en la imagen a continuación.

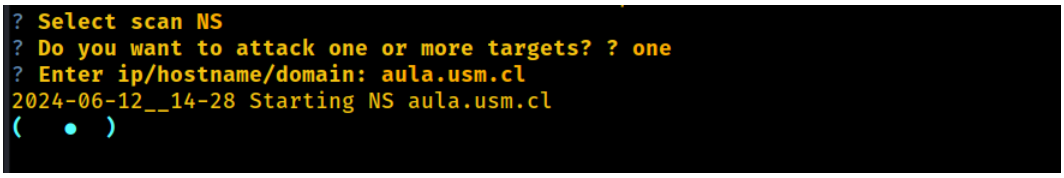


Figura 5.5: Vista opción Attack para NS one

En el caso de elegir massive como opción, la herramienta va a tomar las ip/hostname/domain que se encuentren en el archivo de targets.txt y va a realizar el escaneo de manera secuencial tomando como

primer target la primera fila del archivo, cuando termina de escanear este sigue con la siguiente fila y así sucesivamente hasta llegar a la última fila del archivo, este proceso se puede ver reflejado en la Figura 5.6 , donde podemos apreciar que se muestra un mensaje con la fecha y hora de inicio y término para cada uno de los scaner realizados a los distintos ip/hostname/domain entregados en el archivo de targets.txt .

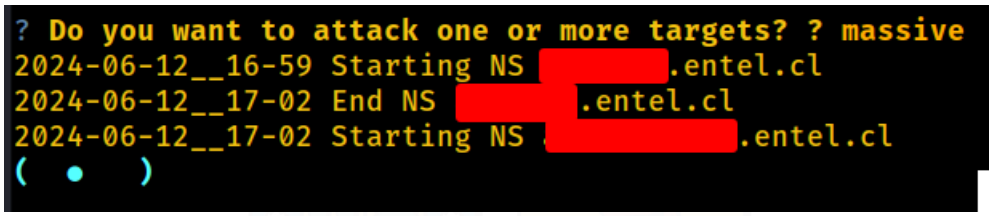


Figura 5.6: Vista opción Attack para NS massive

5.1.2.3. Login

Esta sección sirve para poder autenticarse con el sharepoint corporativo haciendo uso de las credenciales necesarias. Cabe mencionar que el uso de un sharepoint corporativo es en beneficio de la seguridad de los datos. Esto debido a que de esta forma tenemos un canal cifrado para enviar la información, al igual que los datos se encuentran en un lugar seguro donde solo los usuarios con autorización tienen acceso a este sharepoint.

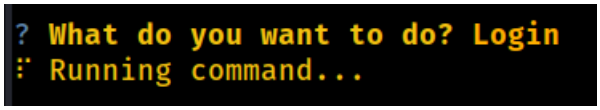


Figura 5.7: Vista opción Login para modo manual

5.1.2.4. Update Database

La función de este módulo es poder subir la base de datos generada de forma local a través de los escaneos de red realizados en el host hacia el sharepoint corporativo.

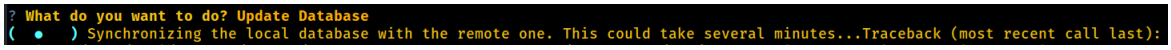


Figura 5.8: Vista opción Update Database para modo manual

5.1.2.5. Export Excel

El módulo de Export Excel tiene la particularidad de generar un excel en base a la base de datos nutrida con los datos de los escaneos realizados, de esta forma se tiene una forma de poder ver los resultados de otra forma, lo que puede ayudar en trabajos futuros como la incorporación de machine learning o por si se llegase a necesitan utilizar la data para el SIEM.

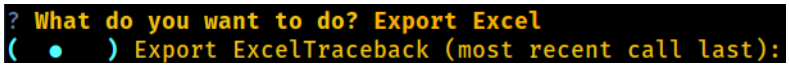


Figura 5.9: Vista opción Export Excel para modo manual

5.1.2.6. Write a comment

Este módulo permite escribir un comentario en alguna ip/hostname/domain que ya se encuentre dentro de la base de datos.

El motivo de esta funcionalidad es poder proveer de información rápida sobre el activo en cuestión. Esto con el fin de agilizar el proceso de obtener información sobre un host y para que pueda ser personalizada la experiencia de buscar información del activo.

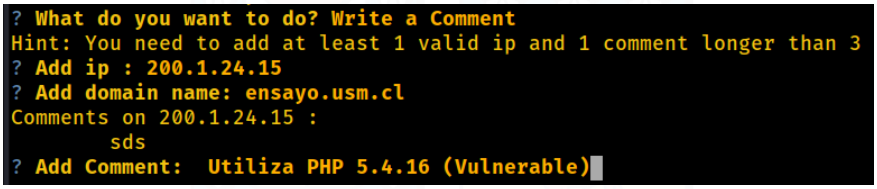


Figura 5.10: Vista opción Write a Comment para modo manual

5.1.2.7. Import Comments

La funcionalidad de este módulo es poder cargar los comentarios generados desde el archivo "comments.csv" a la base de datos ya generada.

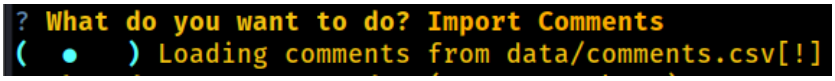


Figura 5.11: Vista opción Import Comments para modo manual

5.1.2.8. Exit

Este módulo es simplemente la opción para salir del programa en la linea de comandos, esto es necesario debido a que debe existir una forma intuitiva de poder cerrar el programa al igual que darle una salida al programa de forma segura y no tener que utilizar una salida como en linux con control+c.

Figura 5.12: Vista opción Exit para modo manual

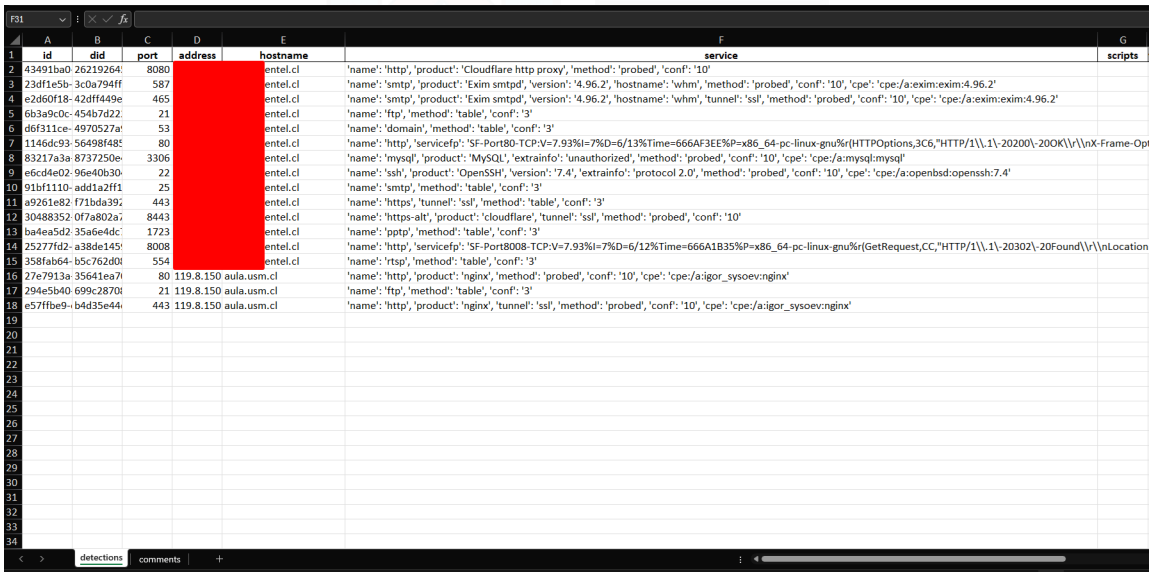
5.1.3. Módulo de Memoria

Para el desarrollo del módulo de memoria se tomaron en cuenta varios factores, como principal funcionalidad es la de almacenar los datos en una base de datos y en forma de excel.

EL motivo del módulo de memoria es poder guardar la data generada al realizar los escaneos de red a las distintas ip/hostname/domain, esto nos entrega distintos beneficios, por ejemplo al momento de ver la información de un activo, este solo tiene que ser consultado ya sea en la data encontrada en la base de datos o en la de excel, esto ahorra mucho tiempo en lo que concierne a la investigación de algún activo o al momento de realizar una prueba de penetration testing.

Otro beneficio importante entregado por el módulo de memoria es el tener un registro histórico tanto de cuando se realizaron las pruebas como también el poder tener almacenado estos datos de los distintos activos con el fin de obtener información sobre alguna posible configuración que se pudo haber tenido en algún momento, esto desde el punto de viste de la ciberseguridad defensiva tiene un gran valor ya que al momento de investigar un escenario de **incident response** es necesario tener información sobre los distintos activos para poder llevar a cabo un buen proceso investigativo.

En el caso del excel tenemos 2 hojas generadas donde la primera corresponde a todas las detecciones generadas por los escaneos de red como se ve en la Figura 5.13



	A	B	C	D	E	F	G
1	id	did	port	address	hostname	service	scripts
2	43491ba0	26219264	8080	[redacted]	entel.cl	'name': 'http', 'product': 'Cloudflare http proxy', 'method': 'probed', 'conf': '10'	
3	23df1e5b	3c0a794ff	587	[redacted]	entel.cl	'name': 'smtp', 'product': 'Exim smtpd', 'version': '4.96.2', 'hostname': 'whm', 'method': 'probed', 'conf': '10', 'cpe': 'cpe:/a:exim:exim:4.96.2'	
4	e2d60f18	42dff449e	465	[redacted]	entel.cl	'name': 'smtp', 'product': 'Exim smtpd', 'version': '4.96.2', 'hostname': 'whm', 'tunnel': 'ssl', 'method': 'probed', 'conf': '10', 'cpe': 'cpe:/a:exim:exim:4.96.2'	
5	6b3a9c0c	454b7d22	21	[redacted]	entel.cl	'name': 'ftp', 'method': 'table', 'conf': '3'	
6	d6f311ce	4970527a	53	[redacted]	entel.cl	'name': 'domain', 'method': 'table', 'conf': '3'	
7	1146dc93	56498f48	80	[redacted]	entel.cl	'name': 'http', 'servicefp': 'SF-Port80-TCP:V=7.93%I=7%D=6/13%Time=666AF3EE%P=x86_64-pc-linux-gnu%r(HTTPOptions,3C6,"HTTP/1\\.1\-20200\-20OK\\r\\nX-Frame-Opt	
8	83217a3a	8737250e	3306	[redacted]	entel.cl	'name': 'mysql', 'product': 'MySQL', 'extrainfo': 'unauthorized', 'method': 'probed', 'conf': '10', 'cpe': 'cpe:/a:mysql:mysql'	
9	e6cd4e02	96e40b30	22	[redacted]	entel.cl	'name': 'ssh', 'product': 'OpenSSH', 'version': '7.4', 'extrainfo': 'protocol 2.0', 'method': 'probed', 'conf': '10', 'cpe': 'cpe:/a:openbsd:openssh:7.4'	
10	91bf1110	add1a2ff1	25	[redacted]	entel.cl	'name': 'smtp', 'method': 'table', 'conf': '3'	
11	a9261e82	f71bda392	443	[redacted]	entel.cl	'name': 'https', 'tunnel': 'ssl', 'method': 'table', 'conf': '3'	
12	30488352	0f7a802a7	8443	[redacted]	entel.cl	'name': 'https-alt', 'product': 'cloudflare', 'tunnel': 'ssl', 'method': 'probed', 'conf': '10'	
13	ba4ea5d2	35a6e4dc	1723	[redacted]	entel.cl	'name': 'pptp', 'method': 'table', 'conf': '3'	
14	25277fd2	a38de145	8008	[redacted]	entel.cl	'name': 'http', 'servicefp': 'SF-Port8008-TCP:V=7.93%I=7%D=6/12%Time=666A1B35%P=x86_64-pc-linux-gnu%r(GetRequest,CC,"HTTP/1\\.1\-20302\-20Found\\r\\nLocation	
15	358fab64	b5c762d0	554	[redacted]	entel.cl	'name': 'rtsp', 'method': 'table', 'conf': '3'	
16	27e7913a	35641ea7	80	119.8.150	aula.usm.cl	'name': 'http', 'product': 'nginx', 'method': 'probed', 'conf': '10', 'cpe': 'cpe:/a:igor_sysoev:nginx'	
17	294e5b40	699c2870	21	119.8.150	aula.usm.cl	'name': 'ftp', 'method': 'table', 'conf': '3'	
18	e57ffbe9	b4d35e44	443	119.8.150	aula.usm.cl	'name': 'http', 'product': 'nginx', 'tunnel': 'ssl', 'method': 'probed', 'conf': '10', 'cpe': 'cpe:/a:igor_sysoev:nginx'	

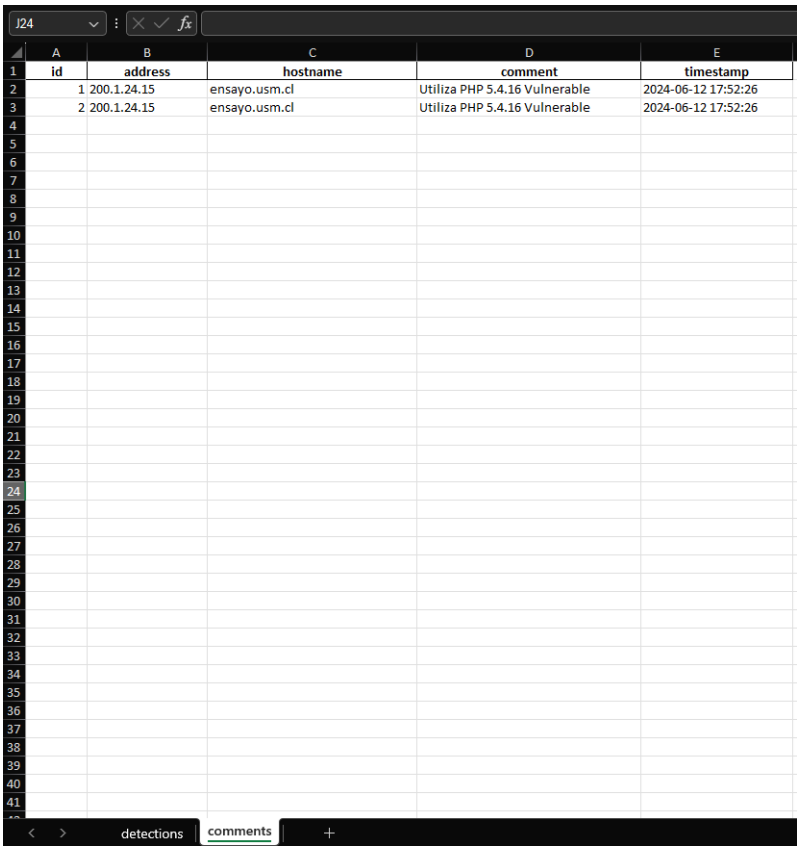
detections comments

Figura 5.13: Vista detecciones para excel generado

Es importante mencionar que los datos más relevantes guardados para las detecciones son:

- **address:** Corresponde a la dirección ip.
- **port:** Corresponde al puerto utilizado para el servicio en cuestión.
- **hostname:** Corresponde al nombre de dominio.
- **service:** Corresponde al servicio encontrado mediante los escaneos.
- **scan_type:** Corresponde al tipo de escaneo realizado por el módulo de reconocimiento y enumeración para obtener la información.

Por otra parte tenemos la hoja correspondiente a los comentarios realizados de forma manual por el usuario haciendo uso de la interfaz CLI.



	A	B	C	D	E
1	id	address	hostname	comment	timestamp
2	1	200.1.24.15	ensayo.usm.cl	Utiliza PHP 5.4.16 Vulnerable	2024-06-12 17:52:26
3	2	200.1.24.15	ensayo.usm.cl	Utiliza PHP 5.4.16 Vulnerable	2024-06-12 17:52:26

Figura 5.14: Vista seccion comentarios para excel generadp

Por último dentro del módulo de memoria tenemos la estructura de la base de datos

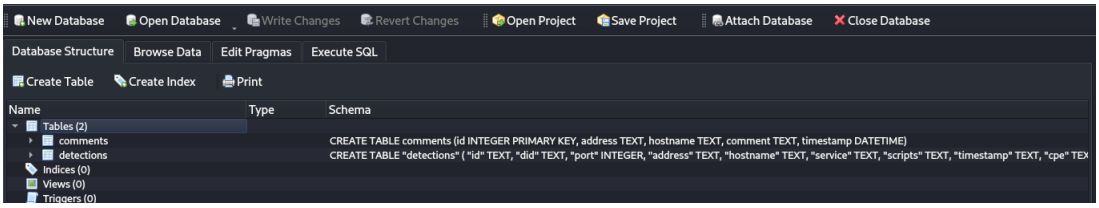


Figura 5.15: Vista ejecución modo manual

5.1.4. Módulo de Reconocimiento y enumeración

Este es el módulo encargado de realizar los escaneos de forma automatizada o manual para los distintos activos que sean necesarios escanear. Este modo permite distintos tipos de modos para realizar pruebas, estos tipos de escaneos fueron elegidos en base a la importancia y regularidad en la que son utilizados por actores maliciosos al tratar de buscar información y probar las seguridades de los servicios entregados.

A continuación se mencionan los modos disponibles con una breve descripción, además de las tácticas y técnicas designadas por MITRE ATT&CK asociadas a cada una de ellas:

- **Network Scan:** Escaneo de red a los 1000 puertos más conocidos. La "Táctica" utilizada es "Reconocimiento" y la "Técnica" correspondiente es el "Escaneo Activo (T1595)"
- **Full Network Scan:** Escaneo de red realizado a todos los puertos de un activo (65535 puertos). La "Táctica" utilizada es "Reconocimiento" y la "Técnica" correspondiente es el "Escaneo Activo (T1595)"
- **Web Scan:** Escaneo de red a los puertos conocidos que tienden a alojar servicios web. La "Táctica" utilizada es "Reconocimiento" y la "Técnica" correspondiente es el "Escaneo Activo (T1595)"
- **Password Spray:** Escaneo de red a los 1000 puertos más conocidos. La "Táctica" utilizada es "Credenciales de acceso" y la "Técnica' correspondiente es el "Brute Force (T1110)"
- **SMB Scan:** Escaneo de red básico al servicio SMB. La "Táctica" utilizada es "Reconocimiento" y la "Técnica" correspondiente es el "Escaneo Activo (T1595)"

Al momento de realizar los escaneos de red estos van generando información referente a los mismos, donde estos se van almacenando en 3 archivos importantes que son el archivo "{scan_mode}-{thread_id}-{scan_address}-parallel.xml" "main.log" y "big.log", donde el primero se encuentra en la carpeta ./data/tmp y las 2 siguientes se encuentran en la carpeta ./logs/ como se muestra en la Figura 5.16.

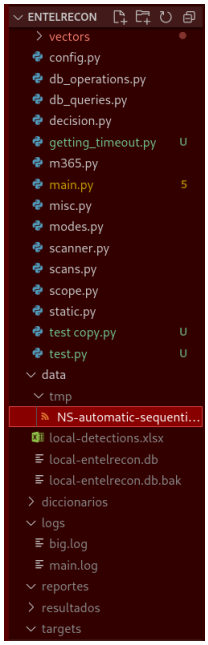


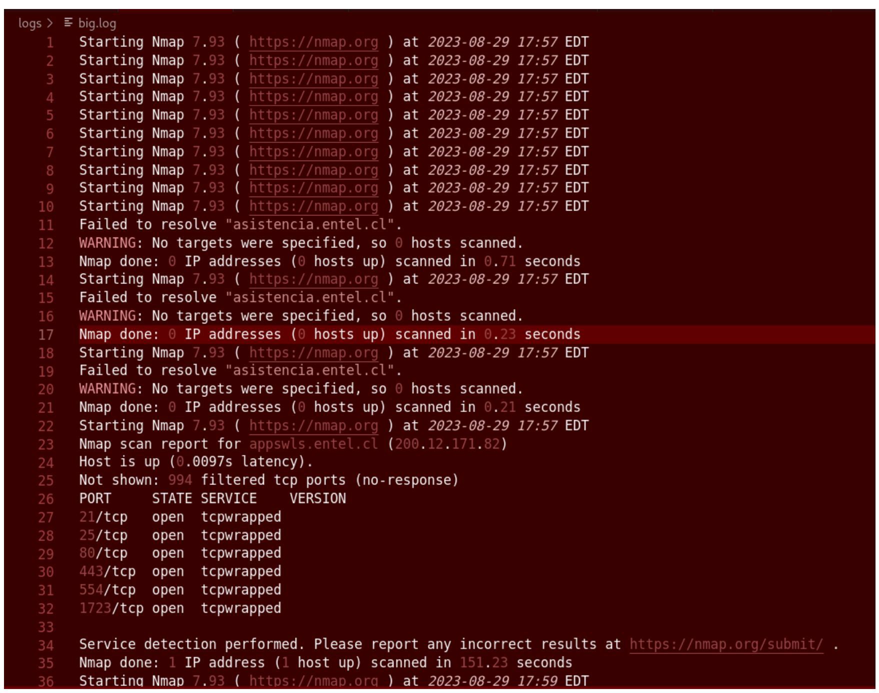
Figura 5.16: Vista de la jerarquía de los archivos del programa

{scan_mode}-{thread_id}-{scan_address}-{scan_type}.xml: Es un archivo que se utiliza para guardar toda la información relacionada con un escaneo para un activo en específico. El motivo de este archivo es después poder procesarlo para extraer la data útil y posteriormente poder hacer la inserción de estos datos en la base de datos local. Al final de ingresar la data dentro de la base de datos local este archivo se elimina, debido a que toda la información de los escaneos se guardará en el archivo "big.log". En este caso el scan_mode hace referencia al modo de escaneo que será utilizado, thread_id corresponde al identificador que tiene el thread que está ejecutando el escaneo, scan_address corresponde a la dirección o el activo que es escaneado del archivo targets.txt.

```
data > tmp > NS-automatic-sequential.xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE nmaprun>
<?xml-stylesheet href="file:///usr/bin/../share/nmap/nmap.xsl" type="text/xsl"?>
<!-- Nmap 7.93 scan initiated Wed Jun 26 18:49:52 2024 as: nmap -Pn -sV -T4 -oX ../data/tmp/NS-automatic-sequential.xml hp.onway.entelocea
<nmaprun scanner="nmap" args="nmap -Pn -sV -T4 -oX ../data/tmp/NS-automatic-sequential.xml hp.onway.entelocean.com" start="1719442192" sta
<scaninfo type="connect" protocol="tcp" numservices="1000" services="1,3-4,6-7,9,13,17,19-26,30,32-33,37,42-43,49,53,70,79-85,88-90,99-10(
```

Figura 5.17: Archivo generado xml

big.log: Este se encuentra encargado de almacenar toda la información referente a los escaneos realizados. La idea de este archivo es tanto para tener un respaldo sobre lo que se obtiene en los escaneos como un archivo único que podría llegar a ser utilizado por alguna herramienta del tipo "SIEM" como "splunk" o un stack de tecnologías utilizado para el monitoreo de la información en redes y hosts como lo es "ELK" donde se podría analizar esta data para encontrar posibles brechas de seguridad, además de poder tener la posibilidad de tener un registro de todo lo realizado durante el periodo del escaneo de la red.



```
logs > big.log
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Failed to resolve "asistencia.entel.cl".
WARNING: No targets were specified, so 0 hosts scanned.
Nmap done: 0 IP addresses (0 hosts up) scanned in 0.71 seconds
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Failed to resolve "asistencia.entel.cl".
WARNING: No targets were specified, so 0 hosts scanned.
Nmap done: 0 IP addresses (0 hosts up) scanned in 0.23 seconds
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Failed to resolve "asistencia.entel.cl".
WARNING: No targets were specified, so 0 hosts scanned.
Nmap done: 0 IP addresses (0 hosts up) scanned in 0.21 seconds
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:57 EDT
Nmap scan report for appswls.entel.cl (200.12.171.82)
Host is up (0.0097s latency).
Not shown: 994 filtered tcp ports (no-response)
PORT     STATE SERVICE    VERSION
21/tcp   open  tcpwrapped
25/tcp   open  tcpwrapped
80/tcp   open  tcpwrapped
443/tcp  open  tcpwrapped
554/tcp  open  tcpwrapped
1723/tcp open  tcpwrapped

Service detection performed. Please report any incorrect results at https://nmap.org/submit/ .
Nmap done: 1 IP address (1 host up) scanned in 151.23 seconds
Starting Nmap 7.93 ( https://nmap.org ) at 2023-08-29 17:59 EDT
```

Figura 5.18: Archivo big log

main.log: Está encargado de guardar información relevante y de alta importan sobre el escaneo a la hora de querer ver que sucedió con algún escaneo a algún activo en específico, como por ejemplo las hora de inicio y de término del escaneo, con esto podemos extraer información para realizar debugging y ver posibles errores o casos de borde en los distintos escaneos.

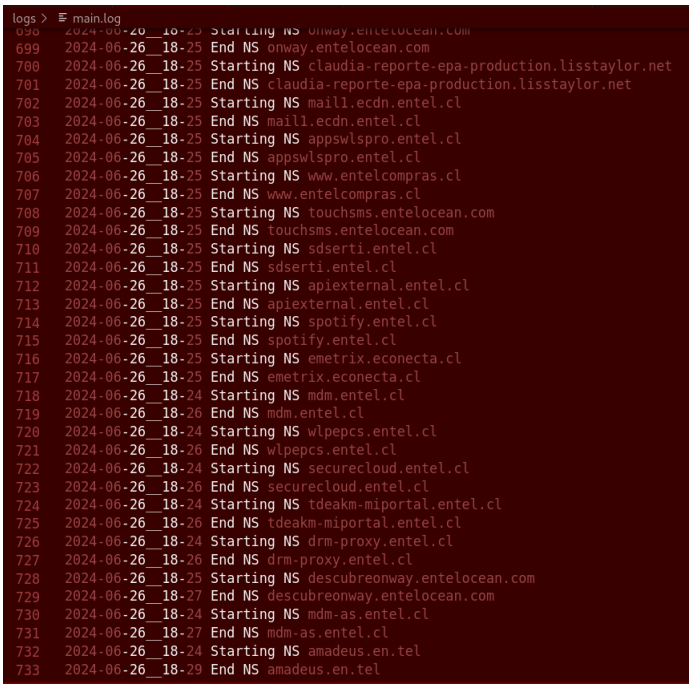


Figura 5.19: Archivo main log

Un archivo importante para los escaneos es el archivo “scans.py” el cual se muestra a continuación:

```
AVAILABLE_SCANS = [
    {
        "long_name": "Network Scan",
        "short_name": "NS",
        "description": "TCP SYN Scan on Top 1000 Most Common Used Ports",
        "cmd": ["nmap", "-Pn","-sV", f"-T{NMAP_TIMMING_TEMPLATE}"],
        "timeout": 1800
    },
    {
        "long_name": "Full Network Scan",
        "short_name": "FNS",
        "description": "TCP SYN Scan on all 65,535 Ports",
       "cmd": ["nmap", "-Pn","-sV", f"-T{NMAP_TIMMING_TEMPLATE}","-p-"],
       "timeout": 1800
    },
    {
        "long_name": "Web Scan",
        "short_name": "WS",
        "description": "Search for interesting and default web URLs on the specified ports",
        "cmd": ["nmap", "-Pn","-sV", f"-T{NMAP_TIMMING_TEMPLATE}","--script", "http-enum","-p", f"{WEB_PORTS}"],
        "timeout": 1800
    },
```

Figura 5.20: Vista de archivo scans.py

El objetivo de usar este archivo es poder tener información sobre los escaneos en general que pueden llegar a ser requeridos por el programa. Como podemos observar en el código tenemos una lista de diccionarios, donde es posible recorrer cada uno de los escaneos disponibles por la herramienta. La llave “long_name” nos muestra el nombre completo del escaneo, “short_name” muestra el nombre abreviado del escaneo para poder ser usado como validador en la ejecución del programa al pasar los parámetros por linea de comando, la llave description sirve para mostrar información sobre lo que realiza el escaneo, cmd corresponde a las flags a utilizar al ejecutar el escaneo nmap y finalmente timeout corresponde al tiempo

asignado a cada uno de los escaneos para tener como referencia cuando uno se demore más del tiempo indicado descartarlo y seguir con el siguiente por alguna posible demora.

Existe otro archivo que nos ayuda a configurar distintos parámetros de los escaneos, este es el archivo "scope.py" que se muestra a continuación:

Figura 5.21: Vista del archivo scope.py

Este archivo como se ve tiene distintas variables que pueden ser modificables por el operario de la herramienta. Como vemos tenemos la variable "MAX_TARGETS_THREADS" que es responsable de decir cual es la cantidad máxima de threads que se pueden ejecutar en el modo paralelo, el "NMAP_TIMMING_TEMPLATE" que corresponde al valor agregado a una flag de la herramienta "nmap" la cual fluctúa enteros entre 0 y 5 donde 0 es hacer el escaneo a una velocidad muy lenta y 5 es a una velocidad muy rápida de escaneo lo cual puede generar mucho ruido, después tenemos "WEB_PORTS" que corresponden a los puertos web que se van a escanear en caso de elegir la opción de Web Scan, la opción de "EXTRA_NMAP_OPTIONS" es para tener la opción de poder agregar flags extras a la herramienta al momento de ejecutar los escaneos, "UPDATE_CVES" sirve para conectarse a la **API** de nvdlib y poder hacer un match con los CVE, finalmente tenemos "RECALCULATE_VECTORS" que nos sirve para poder recalcular los vectores de ataque creados en base a una metodología adoptada por la organización.

Es importante mencionar que el módulo de reconocimiento y enumeración automático tiene 2 modos distintos de operación, en primera instancia se tiene el modo "Secuencial" y el modo "Paralelo".

A continuación se hará una descripción de los modos "Secuencial" y "Paralelo".

5.1.4.1. Modo Secuencial

Este modo no hace uso del paralelismo, por lo tanto realiza escaneos de forma progresiva a medida que va recorriendo linea por linea dentro del archivo "targets.txt" donde tenemos los activos que queremos escanear y posterior al escaneo realizado, este guarda los resultados obtenidos en la base de datos local finalizando la ejecución del escaneo y del modo en cuestión.

El motivo y razón de haber creado un módulo secuencial es para poder apreciar de forma sencilla el comportamiento de algunos scans en diferentes activos. Esto debido a que al no utilizar paralelismo es más fácil de ver resultados de los escaneos realizados en los archivos "main.log" y "big.log".

A continuación se muestra un simple diagrama de como funciona a grandes rasgos el escaneo secuencial

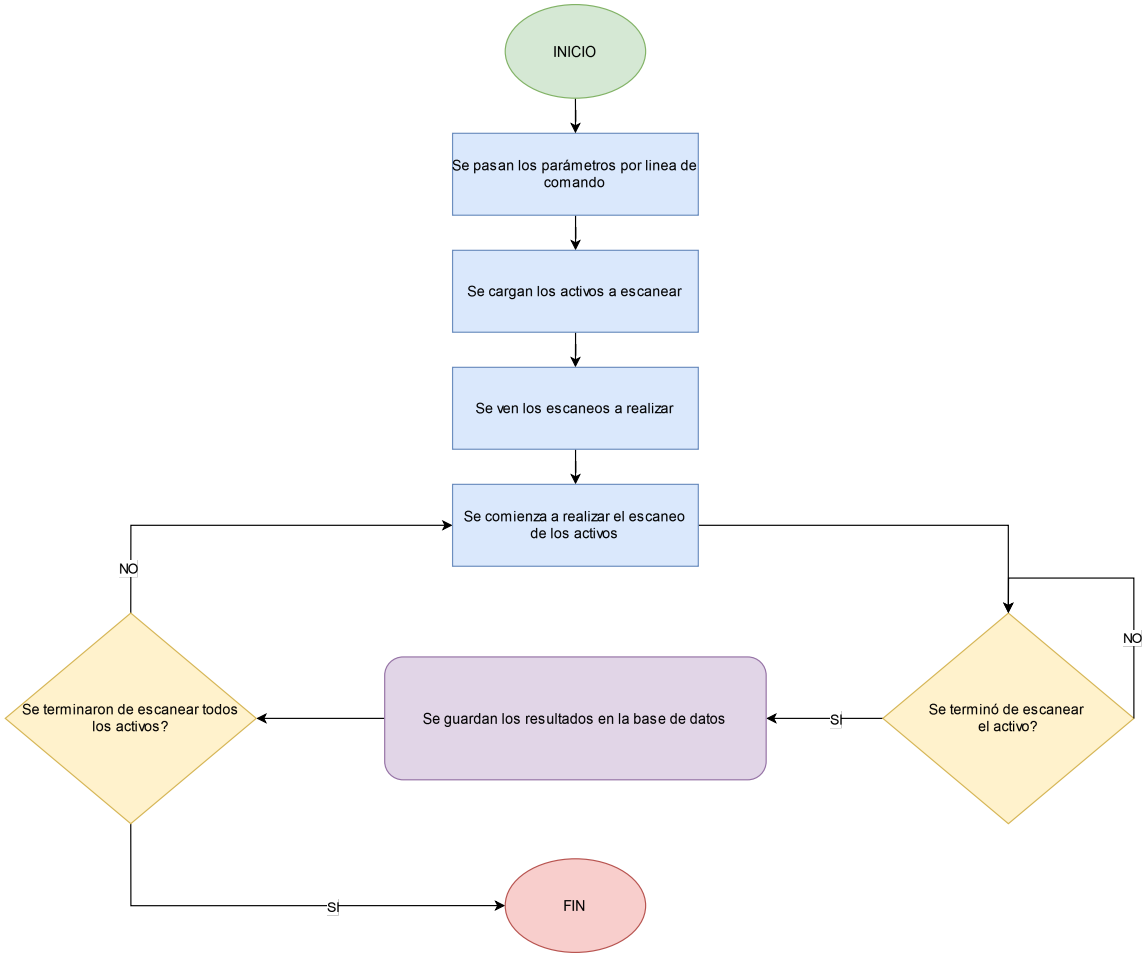


Figura 5.22: Diagrama del scan secuencial

En el flujo del diagrama se puede ver que se empieza tomando los parámetros ingresados por la linea de comando, posteriormente se cargan los activos a escanear por el archivo "targets.txt", se ven los tipos de escaneos a realizar, ya sea network scan, password spray, etc. Posteriormente se comienzan a escanear los activos en base a los tipos de escaneos y se espera hasta que termine de escanear el activo, una vez terminado todos los escaneos relacionados al activo estos se guardan en la base de datos local y al finalizar la operación en la base de datos se verifica si se terminaron de escanear todos los activos, si no es el caso, entonces se sigue con el siguiente activo a escanear, si es el caso y este fue el último, entonces se termina la ejecución del escaneo secuencial.

A continuación se muestra la ejecución del modo secuencial en la “cmd” de “kali” utilizando el comando “htop” de linux para poder mostrar los procesos que están corriendo en memoria donde podemos demostrar la ejecución de este módulo.

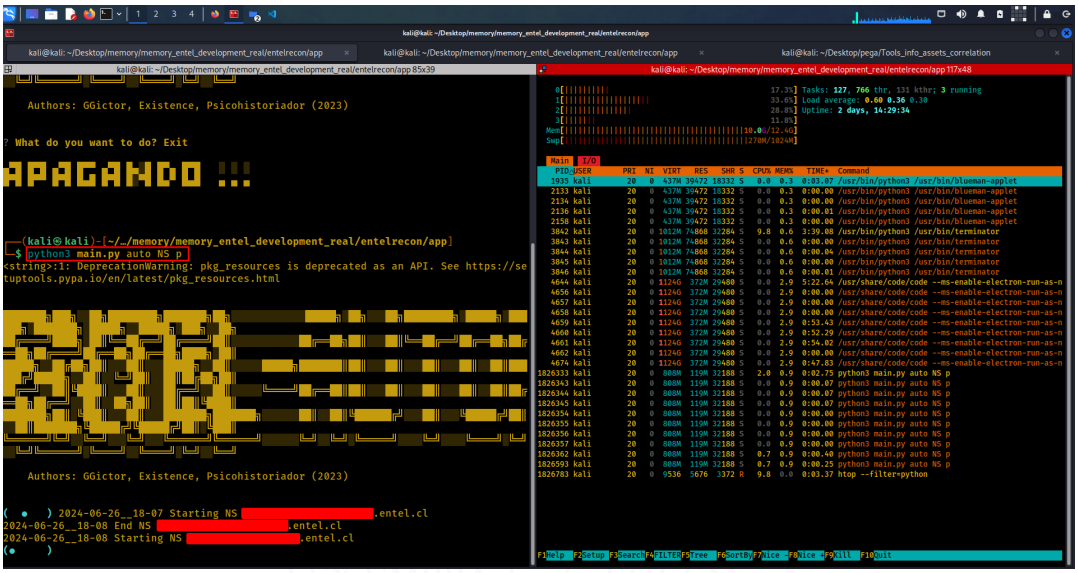


Figura 5.23: Ejecución del scan secuencial

Como se aprecia en la imagen anterior en el costado izquierdo de la imagen podemos ver la ejecución del comando “python3 main.py auto NS p”, donde ejecutamos el programa utilizando python3, le decimos que se ejecute de forma automática que hace referencia el módulo de reconocimiento y enumeración, NS que como comentamos anteriormente hace referencia al modo de escaneo Network Scan (indexar) y finalmente tenemos una “p” que indica que el escaneo será en paralelo, si bien es cierto el escaneo que estamos realizando es un escaneo secuencial, dejo este parámetro a propósito debido a que por diseños de la aplicación si queremos utilizar el modo paralelo el indicador “p” que hace referencia al modo paralelo tiene que venir inmediatamente después de la ejecución del programa (después del main.py), por lo tanto la ejecución en este caso al no tener la flag “p” en el posterior a la ejecución del main.py no se considera y por lo tanto da como resultado la ejecución en modo secuencial.

En lo que respecta al costado derecho de la Figura 5.23 podemos ver la ejecución del comando htop para ver los procesos, en este caso podemos apreciar que se ejecuta el programa en memoria varias veces debido a que el programa utiliza subprocesos para poder realizar los escaneos de la red.

5.1.4.2. Modo Paralelo

Este modo hace uso del paralelismo dentro de su lógica, al igual que en el caso del modo secuencial los activos a escanear se encuentran en el archivo "targets.txt", la diferencia radica en que en este modo se hace uso de **threads** para poder así ejecutar varios escaneos al mismo tiempo y así poder escanear muchos activo al mismo tiempo, de esta forma el ahorro en tiempo al momento de escanear una gran cantidad de activos es notable al final del escaneo, pero es aquí donde empiezan a haber problemas al momento de realizar los escaneos.

En el caso del escaneo en paralelo debemos tomar bastantes factores en cuenta al momento de desarrollarlo, en este caso tenemos problemas en la lectura y escritura de archivos, esto sucede al ejecutar muchas veces las funciones que hacen el escaneo correspondiente dentro del código. Tomando en consideración el uso del paralelismo nos interesa saber cuantos threads vamos a querer utilizar para realizar los escaneos, de esta forma podemos tener un control sobre la cantidad de ejecuciones en paralelo y al mismo tiempo de recursos que vaya a utilizar la máquina que ejecuta la herramienta de reconocimiento. Otro problema que puede llegar a suceder es el problema de los "DeadLock" y del "Starvation" dentro del programa.

El "DeadLock" sucede cuando un proceso nunca obtiene los recursos necesarios para continuar con su ejecución, esto puede suceder cuando tengamos el problema de escritura en la base de datos, al requerir algún recurso por más de algún proceso puede llegar a suceder que se produzca este problema, con el uso de semáforos podemos tener un control sobre la sección crítica del programa que en este caso sería la escritura dentro de la base de datos.

El "Starvation" ocurre cuando cuando existe un proceso que está en ejecución y se queda esperando recursos sin nunca recibirlos debido a que hay uno o muchos otros procesos utilizándolo por ende puede caber la posibilidad de que nunca lo reciba dependiendo de como funciona la cola de prioridad dentro de los procesos. En este caso puede suceder cuando todos los threads se encuentren realizando escaneos y se demoren mucho a tal punto que puede que el escaneo se quede pegado, de esta forma va a haber un thread que se quedará esperando a que alguno se desocupe pero puede caber la posibilidad de que nunca pueda seguir con su ejecución. Para esto lo que se ha realizado es tanto en la escritura de la base de datos como en los escaneos dejar un tiempo asociado a cada semáforo para que si sobrepasa ese tiempo de ejecución este deja de ejecutarse y pasa al siguiente escaneo.

Para resolver estos problemas hacemos uso de los "semáforos". Los semáforos son una herramienta que ayuda a la sincronización de procesos, donde su uso nos ayuda bastante dentro de nuestro caso donde tenemos que tener un control sobre el acceso a recursos dentro del código, cabe decir evitar las condiciones de carrera. También nos ayuda al control del acceso a recursos, eso gracias a la funciones acquire() y release(), donde la primera nos ayuda a poder restar un valor dentro del semáforo y el release() le suma en 1 el valor al semáforo, si es que el semáforo llega al valor 0, significa que bloquea el uso del acquire() del siguiente que lo necesite y lo deja en espera a que se haga un release() dentro del programa.

Los semáforos nos ayudan a evitar el "Starvation" y el "Deadlock" gracias a que podemos tener control sobre la entrada y salida de recursos, al igual que podemos colocar un temporizador dentro de estos para así evitar que se queden esperando de forma indefinida.

A continuación en la figura 5.24 se muestra un diagrama que muestra el flujo de ejecución del modo paralelo.

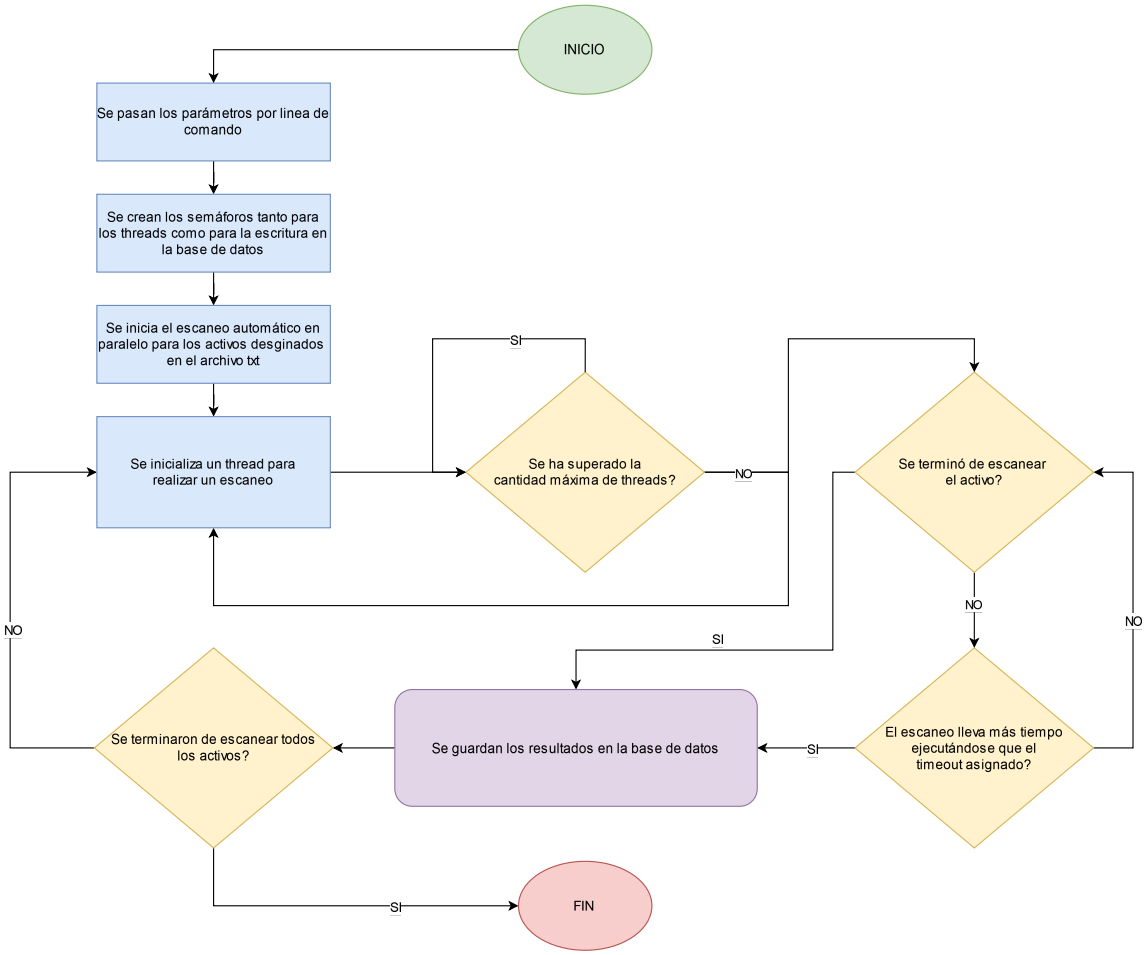


Figura 5.24: Diagrama del scan en paralelo

Como se observa en el diagrama anteriormente mencionado es posible observar que al iniciar el programa se toman los parámetros por la linea de comando para saber cuales serán los escaneos a realizar y cual será el modo de operación que en este caso está orientado al paralelo. Una vez que son pasados los parámetros se determina cual va a ser el modo de operación del escaneo, posterior a esto viene el proceso de creación de los semáforos para poder tener el límite de threads a ejecutarse en paralelo para realizar los escaneos, al igual que se tomó la decisión de crear un "semáforo mutex" que será utilizado para gestionar el acceso y escritura dentro de la base de datos para así tener un control sobre ese proceso. Como se aprecia en la Figura 5.25 podemos ver la creación de los 2 semáforos, donde el "thread_semaphore" corresponde al semáforo encargado de poder marcar el límite de escaneos en paralelo a realizar (Cantidad de threads a ejecutar en paralelo), en este caso utilizando la variable "MAX_TARGETS_THREADS" viene dada por la variable del mismo nombre del archivo "scope.py" de la figura 5.21 mientras que "thread_semaphore_database" corresponde al semáforo mutex que va a estar encargado de dar la entrada y salida de la sección crítica de escritura dentro de la base de datos.

```
else:
    try:
        thread_semaphore = threading.Semaphore(MAX_TARGETS_THREADS)
        thread_semaphore_database = threading.Semaphore()
```

Figura 5.25: Código asociado a la creación de semáforos

Posteriormente a tener todo listo para poder realizar el escaneo es cuando se asigna la variable de un semáforo a cada thread para poder ser llamado desde dentro del thread creado al igual que un activo que viene dado por el archivo "targets.txt" ya mencionado. En el proceso de ejecución de los threads tenemos varias condiciones para el manejo del flujo de este, donde el primero es el ver si se ha superado la cantidad máxima de threads en ejecución, esto se puede verificar gracias al semáforo creado, ya que este lleva la contabilidad de estos, si es positivo y se ha superado la cantidad máxima de threads, entonces queda en espera a que alguno de los threads termine, si no es el caso, entonces ejecuta el escaneo correspondiente y el programa principal ejecuta otro thread. Para el caso en el que el scan no haya terminado se verifica si este ha superado el tiempo máximo de ejecución asignado en un comienzo, en este caso si no lo cumple, entonces sigue el tiempo de ejecución normalmente, pero en caso de que si haya terminado, entonces se ejecuta la escritura en la base de datos y si se termina el escaneo con normalidad, entonces sigue el flujo esperado donde también se escriben los resultados en la base de datos, cabe mencionar que los tiempos asignados a cada uno de los escaneos se encuentra en un archivo llamado "scans.py" que ya describimos anteriormente, en cambio el tiempo relacionado con la escritura dentro de la base de datos se encuentra directamente ingresada dentro del código.

Finalmente al terminar de escribir el resultado en la base de datos se tiene la última condicional que es si se terminaron de escaner todos los activo. Si no es el caso, entonces se inicializa un nuevo thread asociado a un activo a escanear y vuelve al ciclo anteriormente mencionado, si ya no se terminaron de escanear todos los activos, entonces se da fin el modo paralelo.

Es importante considerar que el proceso que ejecuta el thread es el mismo que se ejecuta en el modo secuencial, cabe decir que va a hacer uso del archivo "{scan_mode}-{thread_id}-{scan_address}-parallel.xml" "main.log" y "big.log" y va a seguir el mismo flujo que el secuencial con la diferencia que esta vez se ejecutan en paralelo, por lo tanto como se ha descrito anteriormente es necesario hacer uso de threads y semáforos para evitar posibles problemas relacionados con el paralelismo.

A continuación en la Figura 5.26 se puede ver la ejecución del modo en paralelo.

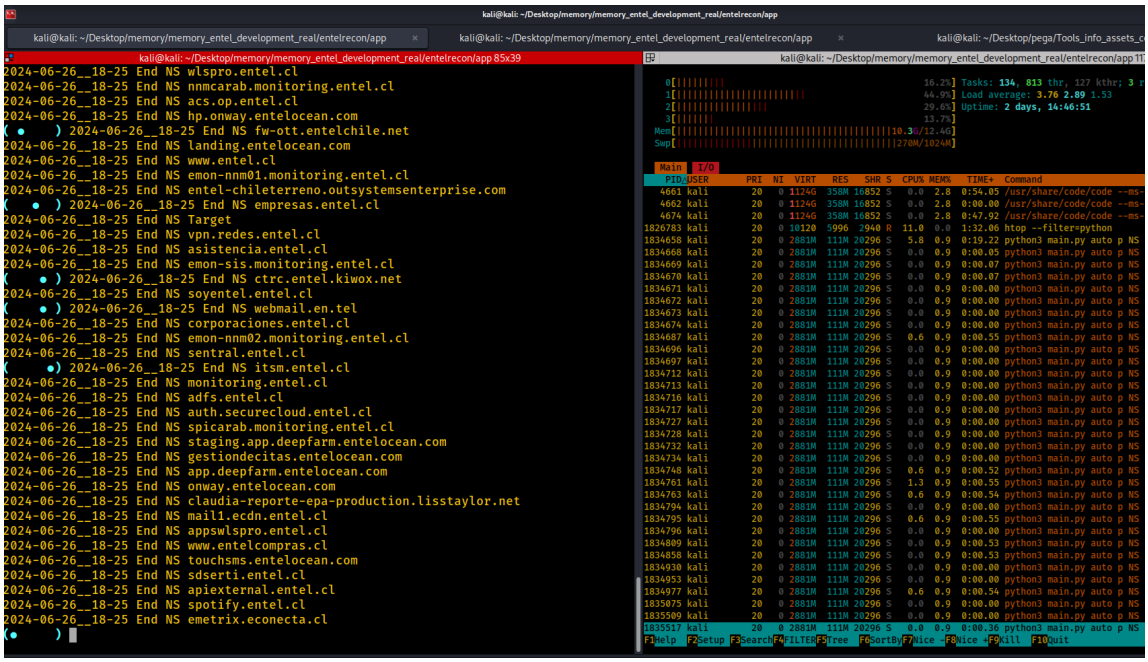


Figura 5.26: Ejecución del scan paralelo

En la imagen anterior se puede apreciar en el costado izquierdo de la linea de comandos la ejecución del modo en paralelo, donde vemos el inicio y término de escaneo de los distintos activos encontrados dentro del archivo "targets.txt" al mismo tiempo que en costado derecho podemos ver los distintos procesos que está ejecutando el sistema gracias a las configuraciones entregadas para el límite máximo de threads.

5.1.5. Módulo de Decisión

Este módulo es el encargado de mostrar cual es el riesgo asociado a cada uno de los activos en base a los resultados obtenidos por los escaneos, para esto es necesario definir que consideramos como riesgo. En el caso particular de la ciberseguridad no existe una sola forma de calcular el riesgo, es por esto que para nuestro caso en particular tomamos el riesgo como:

$$\text{riesgo} = \text{probabilidad de ocurrencia} \times \text{impacto}$$

Donde la probabilidad de ocurrencia es la probabilidad de que ocurra el evento, mientras que el impacto hace referencia a un valor numérico que represente que tan dañino puede llegar a ser para la organización. Estos valores son arbitrarios y son asignados según los dolores de la organización.

Antes de mostrar el funcionamiento general del módulo es importante conocer 2 archivos clave para este proceso, el archivo "config.py" y el archivo "decision.py".

El archivo "config.py" para el caso del módulo de decisión tiene la particularidad de guardar el valor del riesgo asociado a distintos hallazgos encontrados gracias a los escaneos realizados con anterioridad.

```
VECTOR_RISK_LEVELS = {"password_policy": 14.003,
                      "cve": 5.67,
                      "insecure_shared_folder": 15.75,
                      "web_login": 10.843,
                      "suspicious_url": 8.714,
                      "default_credentials": 19.6875,
                      "possible_directory_traversal": 10.9845,
                      "dbms": 23.625
                     }
```

Figura 5.27: Vista de archivo config.py para diccionario VECTOR_RISK_LEVELS

Como se aprecia en la imagen anterior, podemos ver el diccionario “VECTOR_RISK_LEVELS” el cual como mencionamos anteriormente tiene un valor numérico que corresponde al riesgo dependiendo del hallazgo encontrado en los escaneos para cada activo.

El archivo “decision.py” es el encargado de extraer la información obtenida por los escaneos en cada activo a partir de lo guardado en la base de datos local y poder verificar si se encuentra un vector disponible asociado a algún riesgo en algunas de las categorías vistas en el diccionario de la Figura 5.27.

A continuación se muestra el diagrama correspondiente al módulo de decisión.

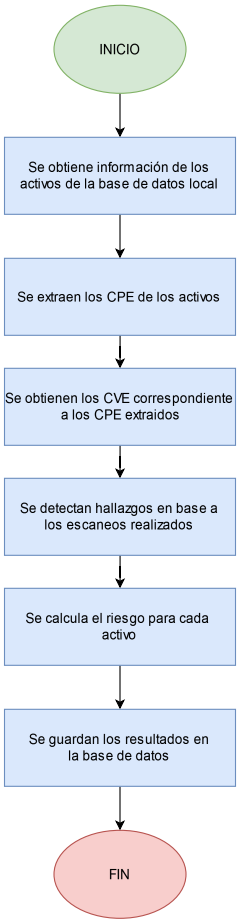


Figura 5.28: Diagrama del módulo de decisión

Aquí es posible ver como se comporta a grandes rasgos el módulo de decisión, empezamos obteniendo la información de los activos obtenida y guardada en la base de datos local. Una vez obtenemos esta información es necesario extraer los CPE y posteriormente conectarse con la base de datos del NIST donde se hace match entre los CPE obtenidos y los CVE para así obtener vulenrabilidades asociados a los hallazgos de los escaneos. Posterior a este proceso se detectan hallazgos encontrados en base a los escaneos realizados como por ejemplo alguna base de datos encontrada, un web server, un servicio ftp, etc. Cada uno de estos hallazgos se considera como un positivo y por tanto se considera su valor del riesgo asociado al diccionario "VECTOR_RISK_LEVELS". Al tener todos los hallazgos, entonces ahora se suman los riesgos obtenidos como se muestra en la imagen a continuación:

```
95  def evaluate_row_risk(row):
96      risk = 0
97      for vector_name in VECTOR_RISK_LEVELS:
98          if vector_name in row and row[vector_name]:
99              risk += VECTOR_RISK_LEVELS[vector_name]
100     return risk
```

Figura 5.29: Vista de archivo config.py para función de calculo de riesgo

Aquí se verifica si dentro de los hallazgos se encuentra algún riesgo y estos se suman devolviendo el resultado final de riesgo asociado a ese activo en particular. Finalmente se crea una columna nueva dentro de la base de datos asociada al riesgo si es que no se encontraba ya creada y se agrega el riesgo calculado asociado a cada uno de los activos, así de esta forma terminando con la ejecución del módulo de decisión.

5.1.6. Módulo de Business Intelligence

Finalmente se tiene el último módulo que es el de business intelligence. Para este caso se tomó como opción hacer utilización de "PowerBI" como la herramienta de business intelligence, debido a su facilidad de uso y de implementación.

Para poder utilizar el módulo de business intelligence es necesario subir la información de la base de datos a un sharepoint corporativo donde solo los especialistas tengan acceso a esta data y puedan hacer uso de los dashboard creados.

Una de las particularidades de los dashboards de "PowerBI" es que son interactivos, por tanto si elegimos alguna de las opciones que se ven dentro del dashboard, toda la interfaz se actualizará a los match que haga con esos datos en particular.

Sin menospreciar la presentación de los datos dentro del dashboard es que decidimos crear un dashboard que sigue la paleta de colores de la organización consideracndo la importancia de la experiencia de usuario al utilizar la herramienta en cuestión.

Como objetivo final dentro de todo el proceso de desarrollo se pensó en hacer 2 vistas utilizando la herramienta, una con un enfoque más ofensivo y otro con un enfoque desde el punto de vista defensivo, donde a una se le nombró como la vista de la "Superficie de Ataque" y a la otra como la vista de los "Vectores de Ataque". A continuación se describen ambos dashboards.

5.1.6.1. Vista dashboard Vectores de Ataque

La creación de este dashboard va orientado con un enfoque más ofensivo, tomando en consideración los resultados obtenidos por los escaneos realizados gracias al módulo de reconocimiento y enumeración.

En la imagen a continuación se expone el dashboard correspondiente a los vectores de ataque.

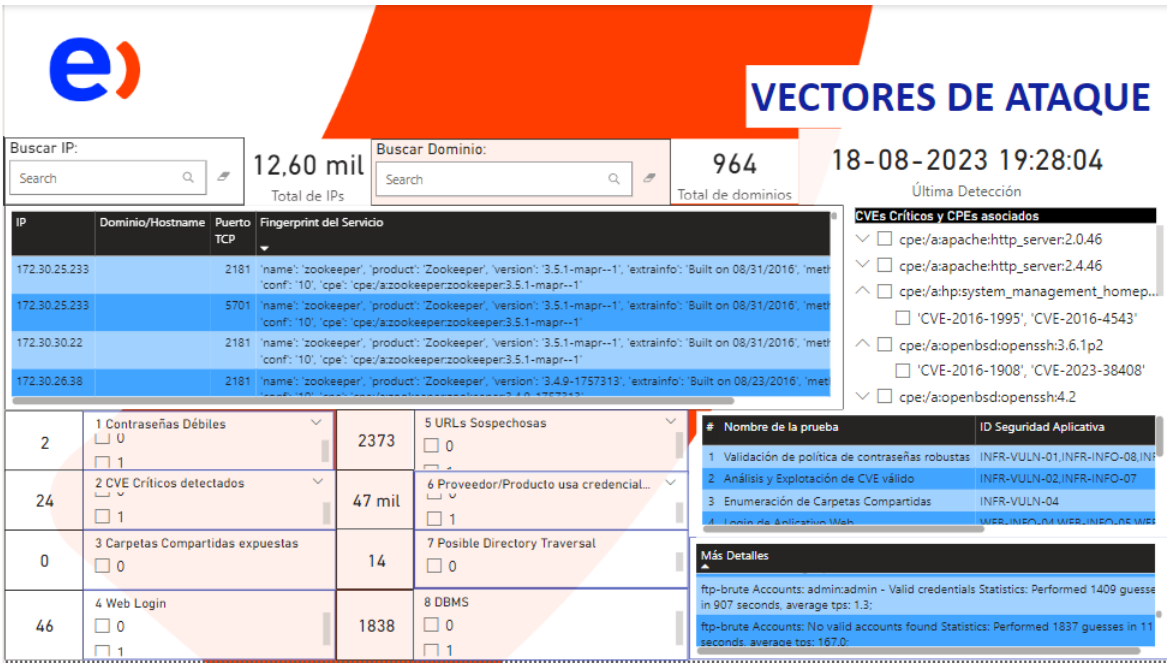


Figura 5.30: Dashboard Vectores de Ataque

Se tiene buscadores en la parte superior del dashboard en base a alguna ip o algún dominio en particular que sea de interés de buscar, al igual que la cantidad total de ip y de dominios según el filtro utilizado, además de una fecha y hora del último escaneo registrado dentro de la base de datos usada para el dashboard, además en la parte del centro del dashboard, abajo de las búsquedas por ip y dominio tenemos la ip asociada a un dominio en particular al igual que el fingerprint del servicio para tener información rápida sobre algún activo. En el costado derecho tenemos la columna donde podemos filtrar por algún CVE en particular en caso de conocer alguno de características críticas, se puede filtrar y probar si existe la vulnerabilidad reportada por el dashboard en algún activo en específico. En el costado izquierdo inferior podemos ver los vectores correspondiente a los dominios e ip que están filtrados, si necesitamos por ejemplo filtrar por los activos que tienen algún DBMS encontrado, entonces podemos seleccionar la opción de True y todo el dashboard se actualizará para los activos que cumplan con el criterio, y así de esta forma con las demás opciones existentes. En el costado derecho donde tenemos el "ID del vector" podemos ver el nombre de la prueba a probar en base a ese ID y el ID interno que corresponde a las pruebas a realizar en base a la metodología interna de entel. Finalmente tenemos más detalles que nos entrega información sobre algún host donde se detectó algún vector de caracter crítico como el que se ve en el ejemplo donde vemos credenciales por defecto en un servicio ftp con el usuario y contraseña por defecto admin:admin.

5.1.6.2. Vista dashboard Superficie de Ataque

Como se mencionó en el comienzo de la módulo de business intelligence, este dashboard tiene un enfoque en el ámbito defensivo.

A continuación se muestra el dashboard correspondiente a la Superficie de Ataque.

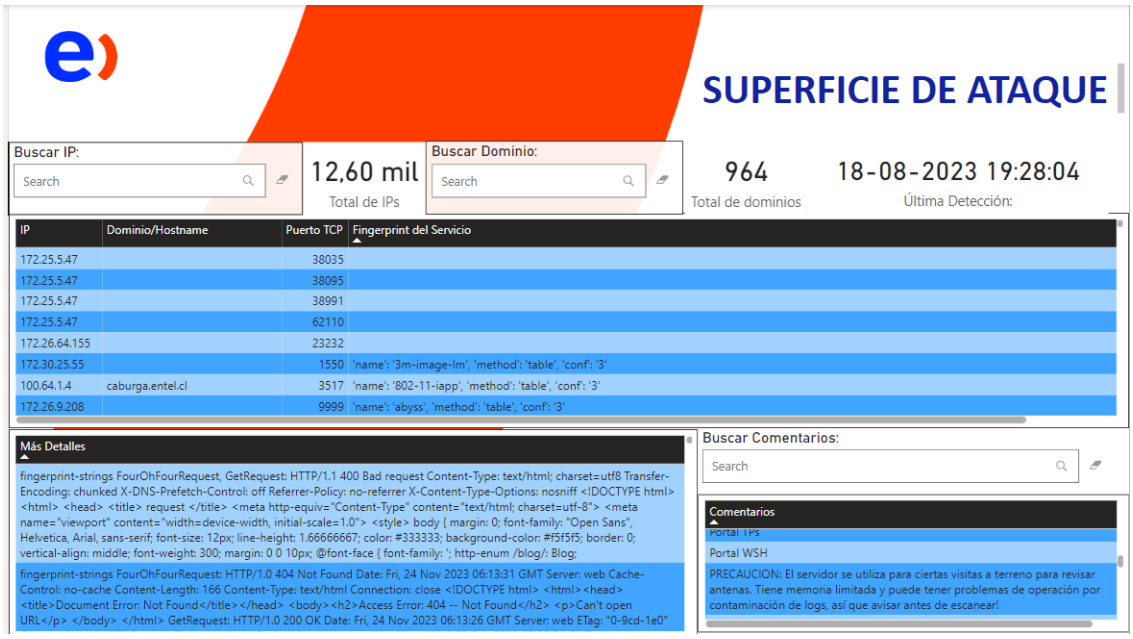


Figura 5.31: Dashboard Superficie de Ataque

En el dashboard es posible ver al igual que en el de “Vectores de Ataque” la búsqueda por IP, busqueda por dominio, cantidad de ips y dominios asociado al filtro utilizado, además de la fecha y hora del último escaneo registrado en la base de datos usada para generar el dashboard. Bajo el buscador de ip es donde se tiene una tabla con la ip,dominio, puerto y el fingerprint del servicio. Esto es de importancia para la ciberseguridad defensiva debido a que si es necesario hacer alguna investigación en algún dominio o ip se puede filtrar la información más reciente sobre este activo en cuestión y ver que servicios tiene expuestos de cara a la red, muchas veces suele suceder que no se tiene unificado todos los planes de trabajo y desarrollos que suceden en una organización, es por esto que aquí se ve el potencial de tener un dashboard unificado con los puertos y servicios que aloja cada una de las ip y dominio de los hosts.

En el costado inferior izquierdo se tiene información general obtenida en este caso por los escaneos respecto a los distintos activos y puertos escaneados que pueden llegar a ser útiles cuando se investigue un servicio de un activo en específico. Finalmente en el costado inferior derecho se tiene un buscador de comentarios al igual que los comentarios encontrados. Los comentarios son muy importantes debido a que son conclusiones o información entregada por algún especialista al hacer uso de la herramienta, por tanto esto ahorra mucho tiempo de investigación al igual que funciona como una forma de comunicación un poco más efectiva debido a que se puede crear un comentario a gusto del especialista donde puede utilizarse una jerga propia de la organización para poder comunicar.

Considerando un caso de uso, donde el equipo defensivo de la organización encuentra una vulnerabilidad crítica liberada como un nuevo CVE, en este caso lo que se quiere dentro de la organización es saber cuales de sus servidores y estaciones de trabajo son posiblemente vulnerables a una vulnerabilidad encontrada, con esta información pueden rápidamente filtrar por los hosts que corran algún servicio en específico y gracias a esto reducir la cantidad de ip’s y dominios a revisar en estos casos, donde se optimiza el tiempo que es crucial al momento de liberarse una vulnerabilidad de estas caracterísiticas.

6 | Resultados

En esta sección se muestran los resultados obtenidos en el transcurso del proyecto. Para esto es necesario considerar que las pruebas realizadas para poder generar una medición y una conclusión frente a lo realizado a lo largo del proyecto se realizaron en el ambiente de Entel, donde se consideraron 22 mil IP's, donde 12.000 de estas corresponden a IP's de Chile y 10.000 corresponden al Perú.

En este proceso de realización de pruebas se realizaron entre estas los siguientes tipos de escaneos:

Network Scan : Gracias al escaneo se obtuvieron como resultado 12.600 IP's activas.

Full Network Scan : Haciendo una comparativa frente al Network Scan se llegó a registrar más de 100 servicios diferentes a los obtenidos gracias al Network Scan.

Password Spray : En las pruebas realizadas a este tipo de escaneo activo se pudo obtener un total de 4 credenciales por defecto asociadas a servicios FTP y SSH, donde se obtuvo 2 para cada uno de estos servicios.

Web Scan : En el caso de los escaneos de servicios web se pudo encontrar un total de más de 2.300 URL's asociadas a posibles directorios que dan inicio a realizar alguna prueba de penetración y pudiesen convertirse en un vector de ataque a utilizar por un ciberdelincuente.

En base a estos resultados realizados por el escaneo en la red interna de la organización se pudo hacer un escaneo efectivo de la superficie de ataque interna entregada por el cliente tanto de Chile como de Perú, obteniendo de esta forma la memoria que se requirió por parte del cliente, además de realizar un escaneo que obtuvo una cantidad de 1.375 IP's escaneadas por hora, donde la prueba duró un total aproximado de 16 horas. Esto muestra una mejora para la organización a la hora de saber los servicios que se encuentran activos en la red interna, Considerando la velocidad a la cual se obtiene la información de los servicios en los activos se puede tener un registro de la red de forma diaria.

En base a los escaneos realizados se obtuvo información de caracter relevante, como por ejemplo:

- 1838 Bases de Datos.
- 2.373 URL's Sospechosas.
- 2 Login ssh y ftp con credenciales por defecto.
- 46 Login Web
- 24 activos con vulnerabilidades críticas y altas.
- 14 posibles directory traversal.

Es importante poder hacer mención de algunos casos de uso mediante el manejo de los distintos módulos de la herramienta que tienen mayor relevancia para el uso en la ciberseguridad defensiva.

6.1. Casos de uso

Los casos de uso a revisar en esta sección van orientados al dashboard de powerBI de Vectores de Ataque y a la generación de data a partir del escaneo en modo automático.

6.1.1. Generación de data para visualizar en powerBI

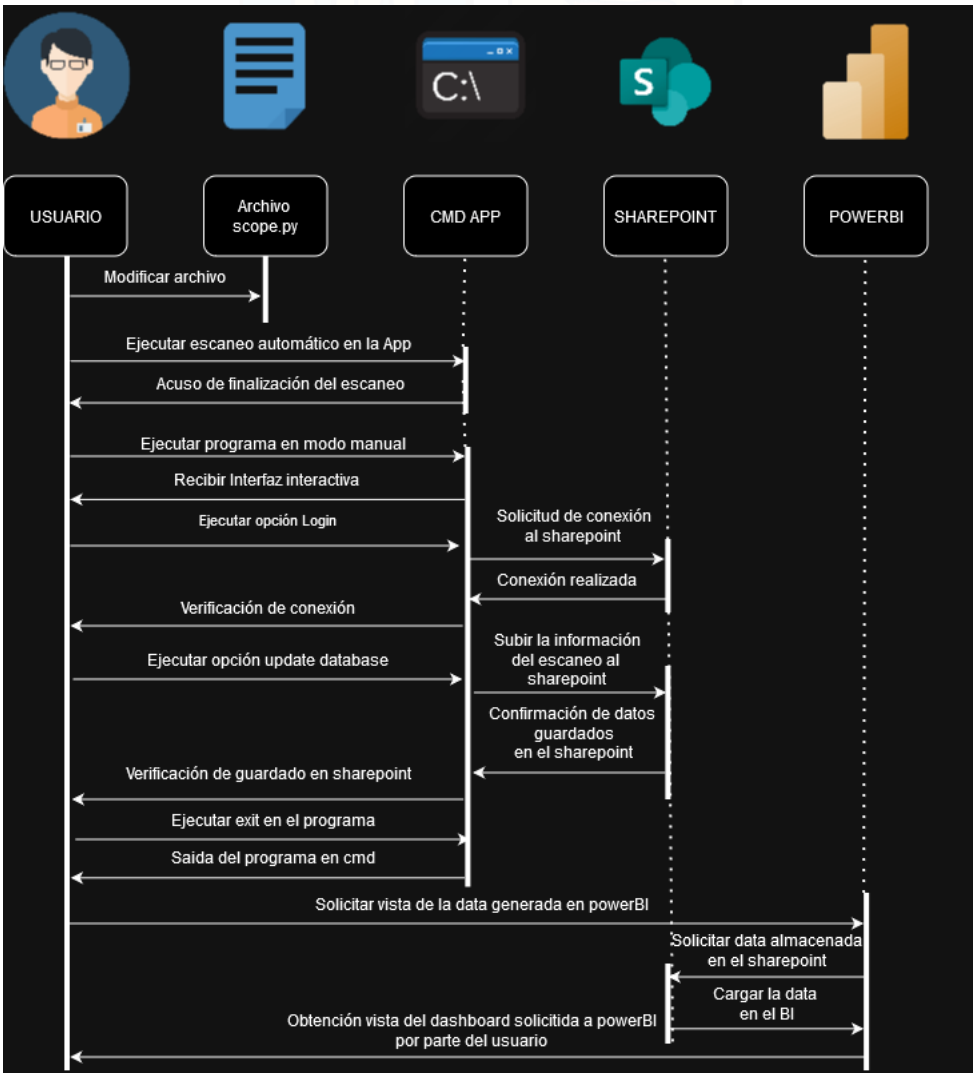


Figura 6.1: Diagrama de interacción para generar data en powerBI

Para entender este proceso debemos empezar desde la configuración del archivo Scope.py (Fig 5.21), donde este contiene las variables a considerar para configurar el modo automático, donde lo que es personalizable son los tipos de escaneos que se realizarán y la cantidad de hilos a utilizar para realizar este proceso.

Posterior a la configuración del archivo hay que ejecutar el programa en modo automático, cabe recordar que como se mencionó en la sección del modo paralelo (Sección 5.1.4.2), pudimos ver que se encuentra el archivo targets.txt el cual tiene la particularidad de poder agregarle las ip y dominios que se quieren escanear antes de realizar la ejecución del programa en modo automático como sigue a continuación:

```
sudo python3 main.py auto p <flags>
```

Las flags corresponden a los distintos tipos de pruebas que se quieran realizar, donde se tiene por ejemplo cualquiera de las opciones que se describen y fueron expuestas en la sección de requisitos de interfaces 4.9.3 y en mayor profundidad lo que hace cada uno de estas flags en la sección donde se describe el módulo de reconocimiento 5.1.4.

Una vez termine la ejecución del escaneo realizado, se debe ejecutar el programa en modo manual, de esta forma debemos ejecutar la opción de login (Sección 5.1.2.3)para poder autenticarse con microsoft 365 y así tener acceso al sharepoint corporativo, posteriormente a esto se va a subir la información recopilada en base al escaneo realizado al sharepoint haciendo uso de la opción update database (Sección 5.1.2.4), las distintas opciones del módulo manual se pueden ver en la Fig 5.2

Al recibir el acuso del programa indicando que ya finalizó la subida de información con éxito al SharePoint procedemos a salir de la aplicación (Fig 5.12)

Finalmente solo queda conectarse directamente al dashboard de powerBI que maneja la organización y poder hacer uso de los dashboards creados tanto con el enfoque ofensivo como defensivo. Esto se indica y se detalla más a fondo en la Sección 5.1.6

Todo esto nos da como resultado la obtención de información relacionada con alguna red de IP y dominios a escanear, esto tiene distintas ventajas como el poder estar haciendo un escaneo constante de alguna gran red y después ver los resultados específicos de esa red como también poder hacer un escaneo puntual en caso de ser necesario por algún evento o requerimiento pedido por gerencia o líderes de equipo dentro del funcionamiento de la gerencia de ciberseguridad.

6.1.2. Investigación e información

Esta sección está orientada a como opera en la realidad el equipo defensivo haciendo uso del dashboard de Superficie de ataque y sus distintas actividades.

6.1.2.1. Búsqueda por IP

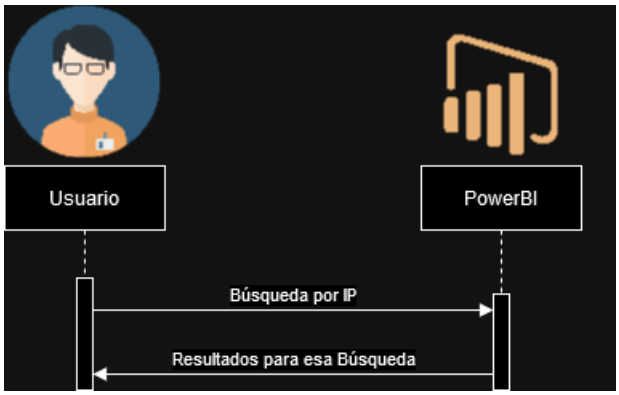


Figura 6.2: Diagrama de interacción para Búsqueda por IP

El usuario hace una búsqueda por IP utilizando el campo de Buscar IP como se vió en la sección anterior, en la Figura 5.31

Esta utilidad puede surgir en base a distintas necesidades:

- Necesidad de gerencia por saber sobre alguna IP en específico.
- Obtener información sobre la IP en alguna auditoría de forma rápida.
- Ver los puertos que tiene abiertos alguna IP en cuestión.
- Ver los servicios que tiene alojados la IP en cuestión.

6.1.2.2. Búsqueda por Dominio

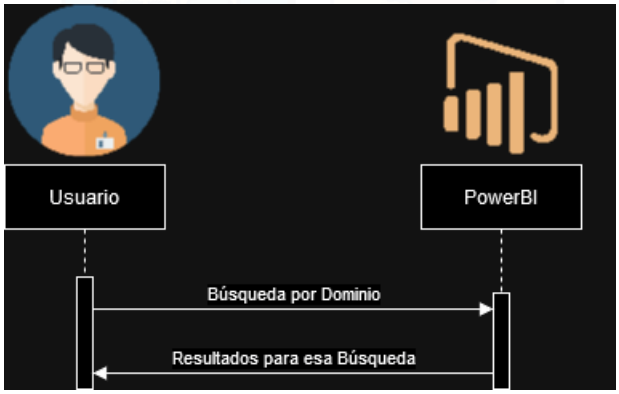


Figura 6.3: Diagrama de interacción para Búsqueda por Dominio

El usuario hace una búsqueda por dominio utilizando el campo de Buscar Dominio en la Figura 5.31

La utilidad de esta parte del dashboard está más asociado a la facilidad de búsqueda, en vez de buscar cual es la ip que se encuentra asociada con el dominio, es posible buscar el dominio directamente en el dashboard y así encontrar la información necesaria.

Las necesidades por las cuales es necesario este buscador son las mismas de la búsqueda por IP.

6.1.2.3. Búsqueda por CVE

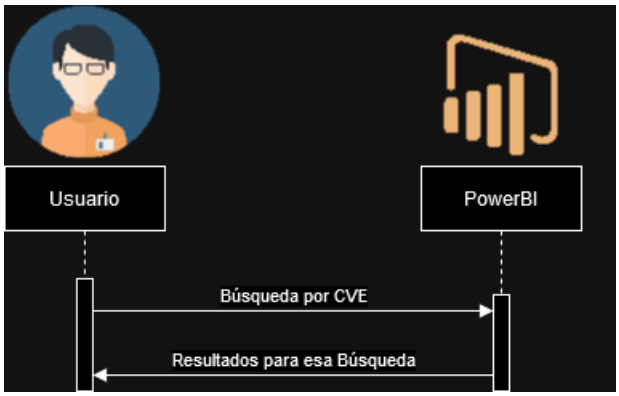


Figura 6.4: Diagrama de interacción para Búsqueda por CVE

El usuario hace una búsqueda por algún CVE encontrado en el campo de Buscar CVE en la Figura 5.31

La utilidad que se le puede ver a esta búsqueda puede ser por distintos motivos:

- Algún nuevo CVE crítico aparece y se revisa si alguna IP es afectada por esta.
- Se ha hecho popular un nuevo CVE por grupos de atacantes que la explotan para poder implantar malware (spyware, ransomware, troyans, etc.)

7 | Conclusión y Trabajos Futuros

Como se observó en el capítulo anterior podemos ver que casi se cumplieron todos los requisitos funcionales propuestos dentro del trabajo de memoria, a excepción del módulo de automatización de entrega de data al SharePoint coorporativo. Dentro de este proceso se pudo entregar memoria a la organización a través de los escaneos que generan información en el tiempo de los distintos activos, además gracias al trabajo realizado es posible realizar un proceso de higienización gracias a la data obtenida. Con respecto al dashboard generado, se puede ahorrar horas hombre al momento de consultar por algún activo y en el momento de un incidente es posible minimizar el tiempo de respuesta debido a la facilidad de consulta de la información por parte del dashboard creado en power BI. Todo esto está realizado con un bien mayor que cumple este trabajo que es el "Conocer la superficie de ataque" y así minimizar esta misma para bajar la probabilidad de ocurrencia de un evento que puede provocar un incidente desde el punto de vista de la ciberseguridad.

Desde el punto de vista de la mejora continua y de la optimización se pueden tomar en cuenta varios puntos de mejora como trabajo futuro, donde se considera:

- Dockerizar la herramienta de tal forma que su instalación sea fácil de realizar en caso de llevar esta herramienta a algún otro host que lo requiera y no tenga conexión a internet para poder descargar las librerías necesarias para su correcto funcionamiento.
- Automatizar el proceso de reconocimiento haciendo uso de un CronJob dentro del host y tener un método para realizar el manejo de la ejecución del programa en caso de que llegue a ejecutarse 2 veces dentro del mismo ambiente.
- Realizar una conexión directa con el SIEM de la organización, esto debido a que hoy en día se tiene la información, pero hay que extraerla del host que corre la herramienta de entel-autorecon e ingresarla al SIEM de forma manual.
- Hacer uso del Machine Learning para crear un clasificador que calcule el riesgo de una forma distinta.
- Realizar pruebas más profundas en el contexto del pentesting Web.
- Realizar otros tipos de scan para darle una mayor utilidad a la herramienta, como por ejemplo ver protocolos como servicios asociados a las conexiones a la red SWIFT, LDAP, IMAP, DNS, etc.

Estas mejoras quedan al pendiente de ser aplicadas en un futuro con el objetivo de mejorar y nutrir la información entregada y la memoria generada por la herramienta en cuestión.

Bibliografía

[1] Decreto política nacional de ciberseguridad
2023-2028 en Chile (https://www.bcn.cl/leychile/navegar?idNorma=1198702) 2.1.6

[2] 2023 Threat Landscape Year in Review: If Everything Is Critical, Nothing Is
(https://blog.qualys.com/vulnerabilities-threat-research/2023/12/19/2023-threat-landscape-year-in-review-part-one) 2.1.6

[3] Paté-Cornell and M. A. Kuypers
A Probabilistic Analysis of Cyber Risks in IEEE Transactions on Engineering Management
vol. 70, no. 1, pp. 3-13, Jan. 2023, doi: 10.1109/TEM.2020.3028526. 2.1.6

[4] Mukhopadhyay, A., Chatterjee, S., Bagchi, K.K. et al.
Cyber Risk Assessment and Mitigation (CRAM) Framework Using Logit and Probit Models for Cyber Insurance
Inf Syst Front 21, 997–1018 (2019).
https://doi.org/10.1007/s10796-017-9808-5 2.1.6

[5] Eling Martin and Jan Hendrik Wirfs
Modelling and Management of Cyber Risk
International Actuarial Association.
https://api.semanticscholar.org/CorpusID:41195093 2.1.6

[6] Hitesh P. Sanghvi and Mohindersinh Dahiya
Cyber Reconnaissance: An Alarm before Cyber Attack
International Journal of Computer Applications. Volume 63, 2013,pp. 36-38.
doi: https://doi.org/10.5120/10472-5202 2.1.6

[7] S. Achleitner, T. F. La Porta, P. McDaniel, S. Sugrim, S. V. Krishnamurthy and R. Chadha
Deceiving Network Reconnaissance Using SDN-Based Virtual Topologies
IEEE Transactions on Network and Service Management, vol. 14, no. 4, pp. 1098-1112, Dec. 2017
doi: 10.1109/TNSM.2017.2724239. 2.1.6

[8] Stefan Achleitner, Thomas La Porta, Patrick McDaniel, Shridatt Sugrim, Srikanth V. Krishnamurthy, and Ritu Chadha. 2016.
Cyber Deception: Virtual Networks to Defend Insider Reconnaissance.
In Proceedings of the 8th ACM CCS International Workshop on Managing Insider Security Threats (MIST '16). Association for Computing Machinery, New York, NY, USA, 57–68.
https://doi.org/10.1145/2995959.2995962 2.1.6