ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE INFORMÁTICA Y ELECTRÓNICA
ESCUELA DE INGENIERÍA EN SISTEMAS

"MODELO BASADO EN LAS TÉCNICAS DE MINERÍA DE DATOS APLICADA A LA DETECCIÓN DE ATAQUES EN LAS REDES DE DATOS DE LA FACULTAD DE INFORMÁTICA Y ELECTRÓNICA"

TESIS DE GRADO

Previa a la obtención del título de

INGENIERO EN SISTEMAS INFORMÁTICOS

Presentado por

MARTHA YOLANDA CARRANZA SUICA

ROSLEY AMPARO NARANJO BARRAGÁN

RIOBAMBA - ECUADOR

2014

Agradezco a Dios por darme la conveniencia de vivir y guiarme en cada paso que doy, por darle la fortaleza necesaria y bendecirme con el cariño y apoyo de todas aquellas personas que han estado a mi lado.

A mi madre por su amor y ejemplo por inculcarme siempre buenos principios por ser mi fortaleza en cada momento, a mi hermana y hermanos por su incondicional apoyo. El apoyo en mis estudios y en cada objetivo que me he planteado, de no ser así no hubiese sido posible alcanzar mi sueño.

A mis amigos de manera especial a mi compañera de tesis y demás familiares ya que me brindan el apoyo, la alegría y me dan las fuerzas necesaria para seguir adelante.

Un agradecimiento especial al Ing. Alberto Arellano nuestro tutor y a la Ing. Ivonne Rodríguez miembro de tesis por la colaboración, paciencia y apoyo que me ha brindado durante el desarrollo de este trabajo.

Martha Yolanda Carranza Suica

Agradezco a Dios por ser el pilar fundamental en la consecuencia de mis objetivos, el amor de mis padres que siempre han estado junto a mí en todo momento por darme la vida, guiarme y apoyarme, a mi hermana y hermano por su entusiasmo y motivación, a mi sobrino que con su existencia llena de luz mi vida.

A las personas que saben lo que me ha costado llevar acabo esta tesis, que de una u otra forma siempre supieron brindarme palabras de aliento para completar una de las etapas de mi vida profesional en especial a mi compañera de tesis por su amistad y paciencia .Un agradecimiento especial a nuestro director de tesis Ing. Alberto Arellano y miembro de tesis Ing. Ivonne Rodríguez por ser guías en el desarrollo de nuestra tesis

Rosley Amparo Naranjo Barragán

Dedico este proyecto de tesis a Dios y mis padres. A Dios el que me ha dado la vida y la fortaleza necesaria para continuar cuando a punto de caer he estado.

A mis padres, por darme la vida y guiarme siempre de manera especial a mi padre que físicamente no está junto a mí sé que desde el cielo donde él se encuentra siempre me guio y me dio su bendición, A mi madre por ser mi ejemplo de perseverancia, amor y lucha constante a seguir.

A mi hermana y hermanos quienes han sabido formarme con buenos sentimientos, hábitos y valores, lo cual me impulsa a luchar para alcanzar mis objetivos.

A mis profesores, gracias por su tiempo, por su sustento así como por los conocimientos que me transmitieron en el avance de mi formación profesional, en especial al Ing. Alberto Arellano, por haber guiado el desarrollo de este trabajo y llegar a la culminación del mismo.

Martha Yolanda Carranza Suica

Dedico este trabajo a mis padres Eduardo y Teresa, a mi hermana Piedad, a mi hermano Javier, a mi sobrino Eduardo Javier, a mi tío Simón quienes con amor y paciencia estuvieron junto a mí durante mi formación académica permitiendo así llegar a la meta.

A mis maestros por su formación y amigos, quienes me ayudaron a culminar con éxito esta tesis.

Rosley Amparo Naranjo Barragán

FIRMAS RESPONSABLES Y NOTAS

NOMBRES	FIRMA	FECHA
Ing. Gonzalo Samaniego Erazo **DECANO DE LA FACULTAD DE INFORMÁTICA Y ELECTRÓNICA.**	_____	_____
Ing. Jorge Huilca. **DIRECTOR DE LA ESCUELA DE INGENIERÍA EN SISTEMAS.**	_____	_____
Ing. Alberto Arellano. **DIRECTOR DE TESIS.**	_____	_____
Ing. Ivone Rodríguez **MIEMBRO DE TESIS.**	_____	_____
DIRECTOR DEL CENTRO DE DOCUMENTACIÓN	_____	_____

NOTA: _____

RESPONSABILIDAD DEL AUTOR

Nosotras, Rosley Amparo Naranjo Barragán y Martha Yolanda Carranza Suica somos las responsables de las ideas, doctrinas y resultados expuestos en esta Tesis y el patrimonio intelectual de la misma pertenecen a la Escuela Superior Politécnica de Chimborazo.

_____ _____

Rosley Amparo Naranjo Barragán Martha Yolanda Carranza Suica

ÍNDICE DE ABREVIATURAS

ACID: Atomicity, Consistency, Isolation, Durability

C&R: Modelo de árbol de clasificación y regresión

CGI: Interfaz de entrada común

CPU: Unidad Central de Procesamiento

DCERPC: Distributing Computing Environment Remote Procedure Call

DDoS: Denegación de servicio distribuido

DESITEL: Departamento de sistemas y telemática

DTIC: Dirección de Tecnologías de Información y Comunicación

DNS: Sistema de nombres de dominio

DoS: Denial of Service

EDP: Procesamiento electrónico de datos

EE.UU: Estados Unidos

ESPOCH: Escuela Superior Politécnica de Chimborazo

FIE: Facultad de Informática y Electrónica

FTP: Protocolo de Transferencia de Archivos

GPL: Licencia Pública General

GRI: Inducción de reglas generalizado

H-IDS: Sistema de detección de intrusiones en el host

HTTP: Protocolo de transferencia de hipertexto

ICMP: Protocolo de Mensajes de Control de Internet

IDS: Sistemas de Detección de Intrusos

INEC: Instituto Nacional de Estadística y Censos

IP: Identidad en la red de un ordenador

IPS: Sistema de prevención de intrusos

ISP: Proveedores de servicio de internet

KDD: Knowledge Discovery from Databases

KNIME: Konstanz Information Miner

N-IDS: Sistema de detección de intrusiones de red

PC: Computadora Personal

R2L: Acceso no autorizado desde una máquina remota

RAM: Memoria de acceso al azar

SMB: Bloque de mensajes de servido

SQL: Lenguaje de consulta estructurado

SSH: Intérprete de órdenes segura

SSL: Capa de conexión segura

TCP: Protocolo de control de transmisión

TI: Tecnología de información

TIC: Tecnologías de la Información y Comunicación

TSL: Transport Layer Security

TLS: Seguridad en la Capa de Transporte

U2R: Acceso no autorizado mediante escalamiento de privilegios

UDP: Protocolo de datagrama de usuario

URLs: Localizador de recurso uniforme

WEKA: Entorno Waikato para el Análisis del Conocimiento

XML: Lenguaje de marcas extensible

ÍNDICE GENERAL

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

INTRODUCCIÓN

“Hoy en día, los datos son el activo más valioso para una organización y por lo que se deben proteger debido a que son sensibles dentro de las redes locales y en internet Existe la posibilidad de que hayan personas o sistemas maliciosos que busquen la manera de llegar a la información a través de distintas formas y ocasionen daños graves Los delincuentes informáticos sacan provecho de las vulnerabilidades y atacan los sistemas computacionales, el registro de flujo de navegación en la red y el intercambio de correo, entre otros”. [1] “Las intromisiones se pueden producir de varias formas: atacantes que acceden a los sistemas desde Internet, usuarios autorizados del sistema que intentan ganar privilegios adicionales para los cuales no están autorizados y usuarios autorizados que hacen un mal uso de los privilegios que se les han asignado. También, se puede entender por intrusión a una violación de la política de seguridad del sistema. En todo caso, cualquier definición de intrusión es necesariamente imprecisa, al igual que los requisitos de política de seguridad no siempre se traducen en un conjunto totalmente definido de acciones”. [2] “La auditoría informática es la encargada de informar de eventos que puedan tener lugar en un sistema informático y pueda ser considerado como una intrusión. Por tanto, la utilidad de un sistema de detección como mecanismo previo de alarma radica en facilitar la distinción entre un acceso normal y habitual al sistema, que puede salir de la puesta en marcha de servicios ofrecidos al externo (entendiendo como externo cualquier otro sistema ajeno al que ofrece los servicios), de un intento de violar de algún modo dichos servicios, e incluso de aquellos que no debieran ser públicos, como parte del ataque a dicho sistema. Así, se podrá proporcionar conocimiento de la puesta en marcha de un ataque sobre el sistema antes

que dicho ataque tenga éxito, con lo que se podrán poner en marcha las medidas necesarias para evitar el impacto". [3] De ahí el objetivo de este trabajo que es la prevención de intrusos a la red de información de la facultad FIE de la ESPOCH mediante la construcción de un modelo que se basa en las técnicas de minería de datos.

Este proyecto de investigación está estructurado en cinco capítulos.

En el **Capítulo I,** se tratará sobre el marco referencial, en el cual se encuentra descrita de manera general los antecedentes, la justificación del proyecto de tesis, los objetivos a alcanzar y la hipótesis a demostrar con el desarrollo de la misma.

En el **Capítulo II,** se detallarán las definiciones conceptuales de la minería de datos, las herramientas junto con la técnica adecuada para realizar la tarea de la minería de datos, así como también los diferentes tipos de ataques que se realizan a una red de información y los distintos IDS.

En el **Capítulo III,** se enfoca al desarrollo del modelo que se propone.

En el **Capítulo IV,** se realizará la implementación del modelo propuesto en la red de datos de la FIE y en el prototipo planteado para las distintas pruebas.

En el **Capítulo V,** aquí se mostrará todos los resultados obtenidos de la aplicación del modelo propuesto y la comprobación de la hipótesis.

El presente trabajo finaliza emitiendo las conclusiones y recomendaciones que son resultado del trabajo finalizado.

CAPÍTULO I

MARCO REFERENCIAL

1.1 Antecedentes

"La tecnología de internet es indispensable en las tareas cotidianas del hombre. Tanto es así que sorprende con la gran cantidad de servicios ofrecidos, y con los modernos adelantes en movilidad y tecnología se puede acceder fácilmente a Internet a través de los dispositivos móviles como las tables y los teléfonos celulares, es por eso que las personas constantemente se están conectando a la información, la misma que se encuentra guardada en enormes servidores por lo que cuya información deben ser protegidas de cualquier intrusión por lo que la seguridad en este sector es crítica". [4]

LAS CIFRAS DEL USO DE INTERNET CRECEN EN EL PAÍS

Encuesta del Inec

El 31,4% de las personas utilizó Internet en el último año. Los necesitados aumentaron su acceso. La mayor parte del uso es por razones de comunicación como se puede apreciar en la Figura I 1

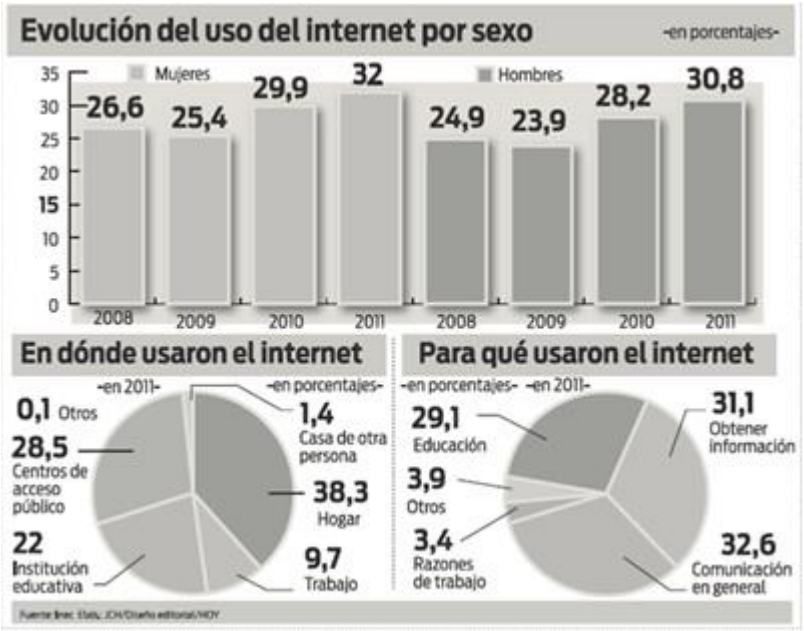


Figura I 1: Uso del internet en el Ecuador
Fuente: Instituto Nacional de Estadística y Censos (INEC).

La última encuesta del Instituto Nacional de Estadística y Censos (INEC) sobre el acceso de los ecuatorianos a las Tecnologías de la Información y Comunicación (TIC), ejecutada en diciembre de 2011, manifestó que el 24,7% de las viviendas tiene una computadora de escritorio y el 9,8% tiene una portátil. Esto es, el 35% de los 3 815 000 hogares que existen en el país. El sondeo realizado en 579 centros poblados abarcó una muestra de 21 768 viviendas, según Byron Villacís, director del INEC. [5]

Ataque a una computadora en un concepto general, un ataque de computadora es cualquier actividad maliciosa realizada sobre el sistema de una computadora o la asistencia que esta provee. Un ejemplo claro de intrusiones de computadoras son los virus, uso de una computadora por un individuo no autorizado, negación de servicio por abuso de una característica, escanear puertos TCP/UDP para reunir información acerca de ese ordenador, o un ataque físico al hardware de una computadora, entre los principales ataques tenemos.

- Modos de Ataque a una red
- Ataque a una estación de trabajo

Según el trabajo realizado por Juan Astudillo "Un sistema de prevención de intrusos (IPS) es un sistema que realiza el control de paso en una red informática para amparar a los sistemas informáticos de intrusiones o acceso no permitido. La tecnología de prevención de intrusos es considerada por algunos como una extensión de los Sistemas de Detección de Intrusos, pero en realidad es otro tipo de control de acceso, más cercano a las tecnologías cortafuegos, la mayoría de los IPS son implementaciones basadas en las versiones IDS por lo que se suele asociar a estos dos dispositivos como uno solo". [6]

"A diferencia de los IDS, los sistemas de prevención de intrusos no sólo alerta al administrador encargado ante la detección de una intrusión, sino que establece políticas de seguridad para resguardar el dispositivo o la red de un ataque". [6]

"Las amenazas contra la seguridad basadas en la red de datos han inducido desfalcos de identidad y estafa financiero generalizados, El correo no esperado, los virus y el spyware causan graves inconvenientes a compañías y clientelas. Una contravención de

seguridad puede producir un daño irremediable a la reputación o la imagen de marca de una corporación. En los Estados Unidos, las dificultades con la seguridad en Internet amenazan con retrasar la adopción de un sistema de expedientes médicos electrónicos a nivel nacional". [7]

"Los ataques actuales hacia la información es un ejercicio beneficioso y a menudo están controlados por las asociaciones del crimen organizado". [7]

"La tecnología de seguridad en Internet sigue avanzando, y está pasando de tener un enfoque pasivo y puntual basado en productos, a obtener planteamientos activos de punta a punta basados en reconocimiento, contención y cuarentena. Además, los vendedores de servicios de Internet (ISP) están luchando en seguridad y los ISP dirigidos al consumidor ofrecen seguridad de Internet dentro de su paquete de servicios". [7]

"Data mining, conocida como la búsqueda de patrones dentro de las enormes bases de datos manipulando para ello métodos estadísticos y de aprendizaje establecido en computadora, está empezando a agrandarse en nuestro país.

Compañías en el sector de telecomunicaciones, financiero y de autoservicio están en el proceso de obtener alguna solución tecnológica en este campo, por lo que surge una demanda por recursos humanos con conocimientos en minería de datos". [8]

"Además, al enfrentar un ambiente más profesional las empresas requieren de tecnologías que les admitan adivinar, dentro de un marco probabilística, el comportamiento de sus consumidores y prospectos a fin de ampliar estrategias de encanto o retención. La idea de minería de datos no es nueva. Ya desde los años sesenta

los estadísticos manipulaban términos como data fishing, data mining o data archeology[1]

Con la idea de descubrir correlaciones sin una hipótesis previa en bases de datos con ruido. A inicios de los años ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro, entre otros, emprendieron a fortalecer los términos de data mining y la extracción de conocimiento. A finales de los años ochenta sólo existían un par de compañías dedicadas a esta tecnología; en la actualidad existen más de 100 empresas en el mundo como Snoop Consulting, Data Mining Group, E.T.A.P.A. entre otras que brindan alrededor de 300 soluciones.

Los inteligentes que disputan sobre este tema las forman científicos de más de ochenta países. Esta tecnología ha sido un buen punto de encuentro entre personas pertenecientes al espacio académico y al de los negocios.

El data mining es un conjunto de técnicas formada por fases que integra diversas áreas y que no es aconsejable confundir con un gran sistema.

Durante el proceso de un proyecto de este tipo se usan diversas aplicaciones de software en cada período que pueden ser para observar información de tipo estadístico, de visualización de información o de inteligencia artificial, principalmente.

[1] Explotación de datos

Actualmente existen aplicaciones o herramientas comerciales de data mining muy eficaces que contienen un sinfín de utilerías que proporcionan el desarrollo de un proyecto. Sin embargo, casi siempre acaban complementándose con otra herramienta". [9]

1.2 Justificación del proyecto de tesis

1.2.1 Justificación Teórica

La investigación es conveniente debido a que permitirá solucionar un problema real y actual mediante la utilización de técnicas modernas como es la de minería de datos , también la investigación es de trascendencia social ya que podría ayudar a solucionar los problemas que poseen actualmente empresas que de alguna manera utilizan redes de computadora.

Entre las cuales tenemos como los establecimientos públicos, privados, educativos, salud, comercio entre otras.

"La prevención de los movimientos de intrusos se ejecuta a través de herramientas que atienden el tráfico en la red o en una computadora" [10]. "Entre la cual tenemos la detección basada en firmas. Una firma tiene la capacidad de reconocer una determinada cadena de bytes en cierto contexto, para activar una alerta". [6] "Por ejemplo, las intrusiones contra los servidores Web generalmente toman la forma de URLs". [6] "Por lo tanto se puede indagar utilizando un cierto patrón de cadenas que pueda emparejar ataques al servidor". [6]

"Sin embargo, como este tipo de detección trabaja semejante a un antivirus, el administrador debe comprobar que las firmas estén constantemente actualizadas". [6]

“En base a los fundamentos y técnicas de la minería de datos se pueden diseñar y elaborar modelos que permiten encontrar procedimientos clandestinos de fácil detección a simple vista como lo es la información no evidente desconocida a priori y potencialmente útil en referencia a hechos determinados”. [1]

Con la aplicación de las técnicas de minería de datos se espera poder crear nuevos conocimientos y la posible creación de nuevas reglas que permitan detectar ataques a la red de información. Con la investigación de las técnicas de minería de datos se contribuirá con el descubrimiento de nuevas intrusiones para brindar mayor seguridad a los datos

1.2.2 Justificación Práctica

La Escuela Superior Politécnica de Chimborazo (ESPOCH) se ve en la necesidad de tener seguras sus redes para de esta manera poder cuidar de su información, ya que esta institución brinda servicios como:

- Académico
- Financiero
- Recursos Humanos
- Páginas webs como el Sitio de Salud
- Servicio de Voz IP entre otros.

En la actualidad la red de información de la ESPOCH cuenta con una aceptable protección de la red, más ningún equipo es completamente invulnerable a ataques interiores o exteriores que pueden amenazar la seguridad de los datos en los distintos servicios con los que cuenta la institución. Y se encuentra con la necesidad de revelar la utilidad del data mining en la indagación del tráfico clandestino en la red de

información, con la finalidad de encontrar y pronunciar intrusiones que no son vistos o detectados por el antivirus, cortafuegos, o sistemas de detección de ataques.

En la red de información de la ESPOCH se registran ataques en su mayor parte al servidor de sistema de nombres de dominio (DNS) que han provocado problemas como la lentitud durante la navegación en internet.

En la FIE existen temas de tesis tanto de minería de datos como de IDS (Sistemas de detección de intrusos) pero no se han realizado trabajos de investigación que se haya integrado los dos temas.

“En la investigación se pretende plasmar la aplicación de técnicas de minería de datos partiendo de una base de datos con captura de tráfico de red. La aplicación de técnicas de minería de datos permite ver y realizar procesos dentro de enormes cantidades de datos de una manera predictiva con el objetivo de emitir resultados oportunos, eficientes, eficaces y confiables y así poder entregar la información necesaria que ayude a tomar decisiones ante los posibles ataques”. [1]

En la Figura I 2 podemos ver cómo funciona la detección de intrusos.

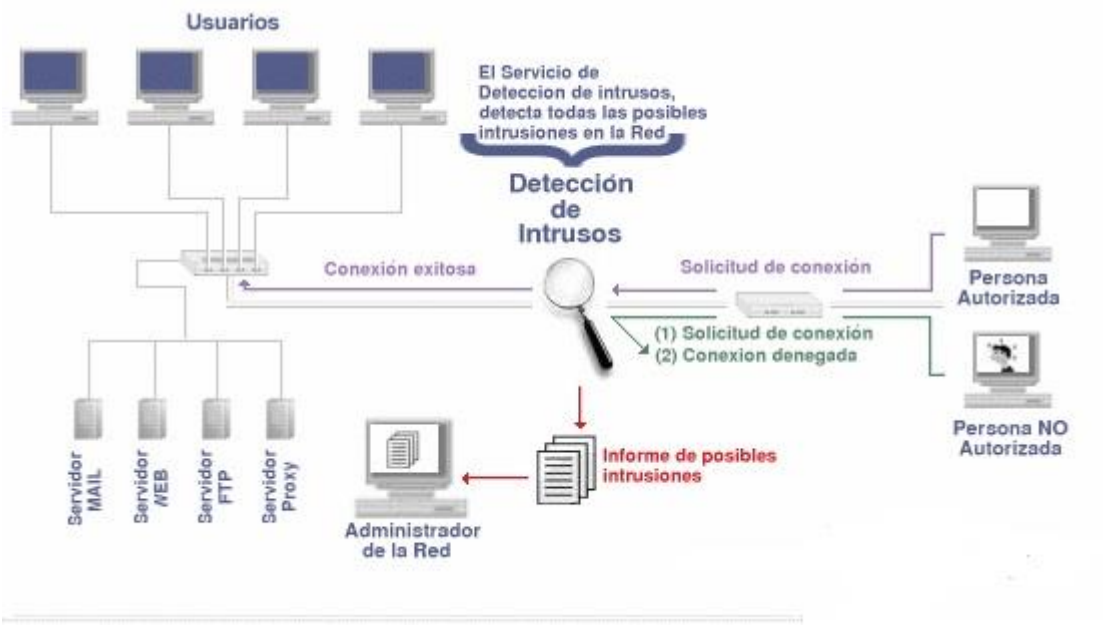


Figura I 2: Detección de intrusos
Fuente: Netsecurity Solutions [1].

Para la aplicación correcta de las técnicas de Data Mining se debe contar con un gran número de registros por esto se realizará el monitoreo de la red de datos de la institución que cuenta con más de 2000 estaciones de trabajo que se conectan simultáneamente a los distintos servicios de la ESPOCH como se muestra en la Figura I 3

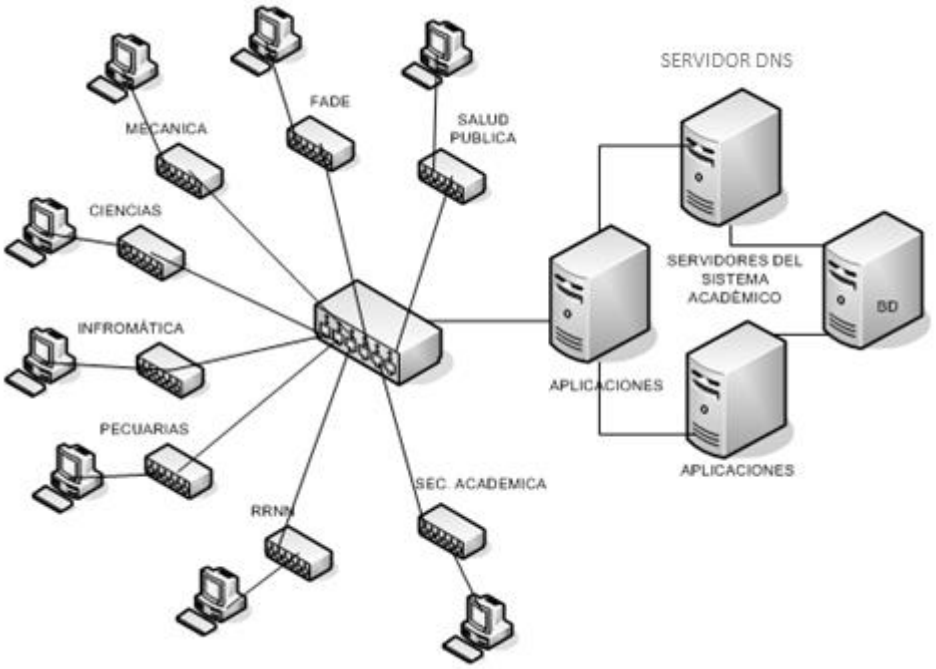


Figura I 3: Esquema de la red institucional de la ESPOCH
Fuente: ESPOCH

Con la información obtenida del monitoreo de la red se aplicará las técnicas de Data Mining, para de esta manera encontrar posibles patrones de los distintos ataques.

En la estructura de la posible solución se propone la implementación de un IDS tradicional y un IDS que aprenderá las nuevas reglas generadas con la aplicación de técnicas de Data Mining.

Las pruebas del modelo basadas en la técnica de Data Mining investigada se realizarán en el departamento de sistemas y telemática (DESITEL).

Los ataques de intrusión se han clasificado en cuatro tipos de ataque los mismos que son:

- "Denegación de servicios (DoS- Denial of service). El objetivo es tratar de interrumpir el funcionamiento de la red, la máquina o el proceso, de tal forma que un servicio o recurso sea inaccesible a los usuarios legítimos". [1]
- "R2L (Remote to local). Acceso no acreditado desde una máquina remota". [1]
- "U2R (User administrador). Acceso no autorizado mediante escalamiento de privilegios de una cuenta de usuario autorizado hasta llegar a superusuario". [1]
- "Indagación y exploración (Probing). Este ataque escanea las redes, en busca de debilidades mediante la recaudación de información, tal como, direcciones IP validas, servicios, sistemas operativos y otros". [1]

Para el análisis se tomaran en cuenta los principales ataques que se realizan a un servidor DNS los que se describen en la Tabla I I.

Tabla I I: Ataques servidores DNS **[1]**.

ATAQUES	DESCRIPCIÓN
DoS (Denial of Service)	Consisten en impedir que los usuarios utilicen el Servicio.
DNS Spoofing	"Realiza la suplantación de identidad por el nombre de dominio. Esto hace que la relación "Nombre de dominio-IP" sea falsa, es decir, resuelve con una dirección IP falsa un cierto nombre DNS o inversamente". [1]
TXID	"Un DNS maneja un ID aleatorio para sus paquetes, pero esto solo es en la primera consulta, después lo aumenta para las continuas consultas de forma que si es posible esnifar el DNS se puede anunciar el ID. En anteriores versiones de Bind esto permitía suplir a un DNS autoritario sobre un DNS víctima y así cambiar su caché a voluntad". [1]

Ataques servidores DNS [1]. (Continuación)

MitM	"Los ataques Man in the Middle (o intermediario) estos ataques se interponen entre el paso de información de dos ordenadores. El atacante lee y altera la información del flujo sin que lo noten las máquinas afectadas. En el caso del servidor DNS el atacante escanea las peticiones DNS de la víctima". [1]

Fuente: Manuel Antolín Ayuso y Miguel Ángel Barcenilla Mancha

Se realizará pruebas de verificación con la ejecución de ataques a otros servicios en el prototipo que se muestra en la Figura I 4.

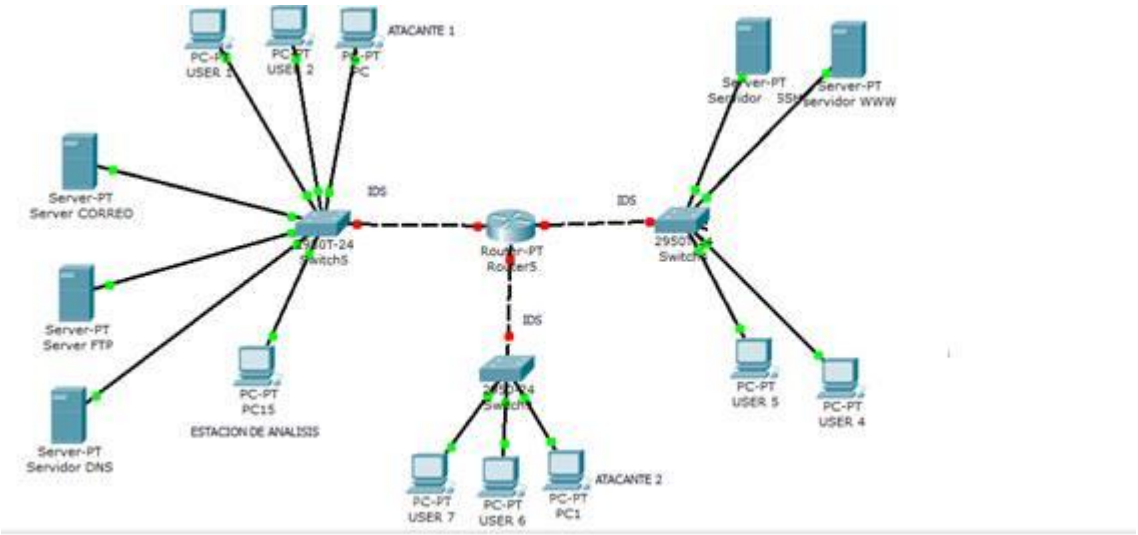


Figura I 4: Prototipo de red
Fuente: Autoras

Para medir el nivel de rendimiento de las propuestas se evaluarán los siguientes parámetros: número de ataques detectados, número de falsos positivos detectados, tiempo de detección de intrusos, tipos de intrusiones detectadas, recursos requeridos en la implementación.

1.3 Objetivos

1.3.1 Generales

Desarrollar un modelo basado en las técnicas de Minería de Datos y su aplicación en la detección de ataques en las redes de datos de la Facultad de Informática y Electrónica

1.3.2 Específicos

1. Estudiar las diferentes técnicas existentes en Data Mining
2. Seleccionar la técnica junto con la herramienta de software libre para Data Mining.
3. Diseñar el modelo utilizando técnicas de Data Mining previamente seleccionadas.
4. Aplicar el modelo para la detección de anomalías en el tráfico de la red de datos de la Facultad de Informática y Electrónica y en el prototipo de red de datos diseñada en el laboratorio de la Academia Cisco.
5. Evaluar los resultados de la aplicación del modelo.

1.4 Hipótesis

Implementar el modelo basado en técnicas de Data Mining permitirá la detección del tráfico anómalo de la red de datos de la Facultad de Informática y Electrónica.

1.5 Métodos y técnicas

1.5.1 Métodos

Para el desarrollo de esta investigación se aplicará el método científico, el mismo que ayudará a seguir una secuencia ordenada de acciones.

- Planteamiento del problema

- Formulación de la hipótesis
- Levantamiento o recopilación de la información
- Análisis e interpretación de resultados
- Comprobación de la hipótesis
- Difusión de resultados

Método Analítico.- Con la ayuda de este método se pretende el estudio en partes del objeto de investigación.

Método Inductivo.- Nos permitirá generar una conclusión de manera general a partir de hechos particulares.

1.5.2 Técnicas

Las diferentes técnicas a utilizarse para el desarrollo de este trabajo son las siguientes.

- Revisión de Artículos Científicos.
- Observación.
- Pruebas.

CAPÍTULO II

MARCO TEÓRICO CONCEPTUAL

2.1 Introducción

En este capítulo se dará a conocer los conceptos básicos para el desarrollo de este tema investigativo. Cuya finalidad es tener en claro los fundamentos teóricos que se necesita sobre la tecnología de minería de datos, ataques o intrusiones a una red de datos, motores de detección de intrusos y herramientas de minería de datos. Se expondrá el enfoque, características importantes, ventajas, desventajas, los tipos, así como también imágenes que explican su funcionamiento. Un elemento clave en esta sociedad de la informática ha sido las diferentes técnicas y tecnologías que van evolucionando para contrarrestar los diferentes ataques producidos a las redes de información dando paso así a la delincuencia informática.

2.2 Minería De Datos

2.2.1 Antecedentes

"La existencia de voluminosas bases de datos conteniendo grandes cantidades de datos, que exceden en mucho las capacidades humanas de reducción y análisis a fin de obtener información útil, actualmente son una realidad en muchas organizaciones. Debido a esto, frecuentemente, las decisiones importantes se toman en base a la intuición y experiencia del decisor más que considerando la rica información almacenada". [1]

"La minería de datos nace como una tecnología que pretende ayudar a entender el contenido de una base de datos. De manera general, los datos son la materia prima bruta. De acuerdo a la situación el usuario agrega algo especial a los datos convirtiéndose así en información. Con la aplicación de técnicas se encuentran modelos, que conjuntamente con la información representa un valor agregado y se genera conocimiento". [11]

2.2.2 Definición de minería de datos

Según el sitio oocites "se puede definir la Minería de Datos como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. Es decir, la tarea fundamental de la Minería de Datos es encontrar modelos comprensibles a partir de los datos". [12]

Según el sitio de Microsoft "La minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado

complejas o porque hay demasiados datos, estos patrones y tendencias se pueden recopilar y definir como un modelo de minería de datos". [13]

Según el sitio Microsoft, "La minería de datos suele describirse como "el proceso de extraer información válida, auténtica y que se pueda procesar de las bases de datos de gran tamaño". En otras palabras, la minería de datos deriva patrones y tendencias que existen en los datos. Estos patrones y tendencias se pueden recopilar y definir como un modelo de minería de datos". [14]

La definición del sitio sinnexus "El datamining (minería de datos), es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto". [15]

"Básicamente, el datamining surge para intentar ayudar a comprender el contenido de un repositorio de datos. Con este fin, hace uso de prácticas estadísticas y, en algunos casos, de algoritmos de búsqueda próximos a la Inteligencia Artificial[2] y a las redes neuronales". [15]

[2] La Inteligencia Artificial es una combinación de la ciencia del computador, fisiología y filosofía, tan general y amplio como eso, es que reúne varios campos (robótica, sistemas expertos, por ejemplo), todos los cuales tienen en común la creación de máquinas que pueden pensar.

De acuerdo a las definiciones mencionadas anteriormente se define como minería de datos al proceso de encontrar un conocimiento a partir de volúmenes grandes de datos. Este proceso ayudará a la toma de decisiones

Proceso de descubrimiento de conocimiento en bases de datos

El KDD (Knowledge Discovery from Databases) "es el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia, comprensibles a partir de los datos". [16]

"El objetivo fundamental del KDD, es encontrar conocimiento útil, válido, relevante y nuevo sobre una determinada actividad mediante algoritmos, dadas las crecientes órdenes de magnitud en los datos como se puede ver en la Figura II 5.

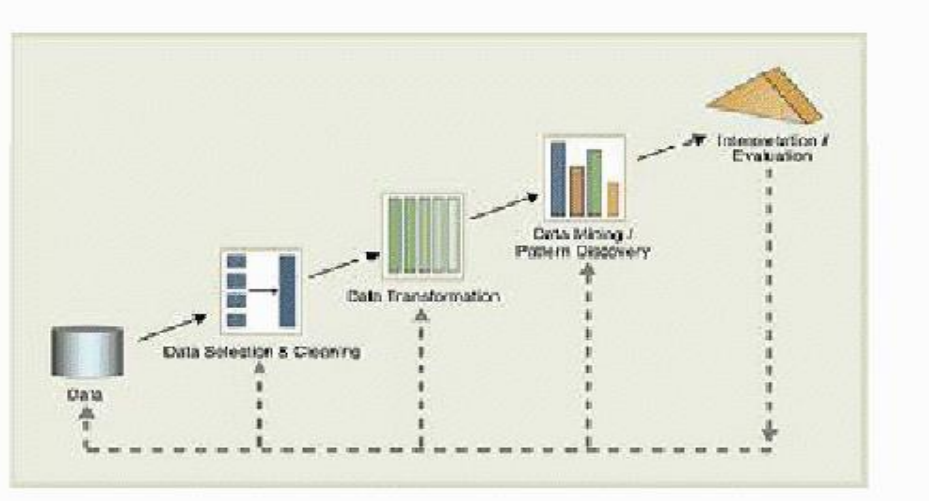


Figura II 5: Proceso del KDD (Knowledge Discovery from Databases)
Fuente: Edixon Rosillo

Al mismo tiempo hay un profundo interés por presentar los resultados de manera visual o al menos de manera que su interpretación sea muy clara. El resultado de la

exploración deberá ser interesante y su calidad no debe ser afectada por ruido en los datos". [17]

2.2.3 El proceso de extracción de conocimiento

"La Extracción de conocimiento está principalmente relacionada con el proceso de descubrimiento conocido como KDD, que se refiere al proceso no-trivial de descubrir conocimiento e información potencialmente útil dentro de los datos contenidos en algún repositorio de información. No es un proceso automático, es un proceso iterativo que exhaustivamente explora volúmenes muy grandes de datos para determinar relaciones, es un proceso que extrae información de calidad que puede usarse para dibujar conclusiones basadas en relaciones o modelos dentro de los datos la Figura II 6 ilustra las etapas del proceso KDD". [18]

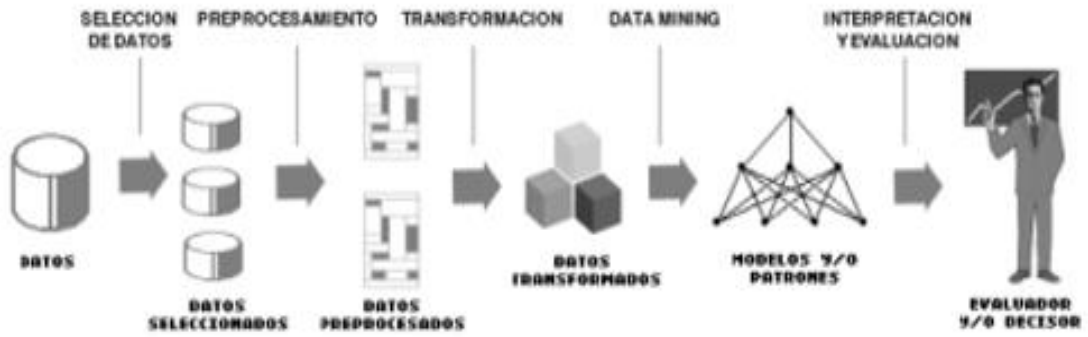


Figura II 6: Proceso KDD
Fuente: webmining.cl

2.2.4 Fases del proceso de extracción del conocimiento

1. "Análisis o entendimiento del negocio. Incluye la comprensión de los objetivos y requerimientos del proyecto desde una perspectiva empresarial, con el fin de convertirlos en objetivos técnicos y en una planificación". [1]

2. "Análisis o entendimiento de los datos. Comprende la recolección inicial de datos, en orden a que sea posible establecer un primer contacto con el problema, identificando la calidad de los datos y estableciendo las relaciones más evidentes que permitan establecer las primeras hipótesis". [1]

3. "**Preparación de los datos.** Incluye las tareas generales de selección de datos a los que se va a aplicar la técnica de modelado (variables y muestras), limpieza de los datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato". [1]

4. "**Modelado.** Se seleccionan las técnicas de modelado más apropiadas para el proyecto de minería de datos específico". [1]

5. "**Evaluación de resultados.** Se evalúa el modelo, no desde el punto de vista de los datos, sino del cumplimiento de los criterios de éxito del problema". [1]

6. "**Explotación o despliegue de resultados.** Normalmente los proyectos de minería de datos no terminan en la implantación del modelo sino que se debe documentar y presentar los resultados de manera comprensible en orden a lograr un incremento del conocimiento. Además, en la fase de explotación se debe asegurar el mantenimiento de la aplicación y la posible difusión de los resultados". [1]

2.2.5 Técnicas de modelado de minería de datos

"Las técnicas se las considera un enfoque conceptual para extraer la información de los datos para su implementación se hace uso de los distintos algoritmos. Los algoritmos representan la manera de aplicar una técnica de manera detallada". [19]

Segmentación

Según Vallejo y Tenelanda, "segmentación dividen los datos en segmentos o conglomerados de registros que tienen patrones similares de campos de entrada. Como sólo se interesan por los campos de entrada, los modelos de segmentación no contemplan el concepto de campos de salida o destino". [20]

Según el trabajo de Tirnauca, "segmentación es discernir diferentes subconjuntos de los datos en base a características similares (los segmentos o clusters), aplicaciones: tratar del mismo modo datos parecidos e identificar rasgos aproximadamente comunes a un segmento de una población". [21]

Según Microsoft, "segmentación, dividen los datos en grupos, o clústeres, de elementos que tienen propiedades similares". [22]

Según el trabajo de Santiagozapatakdd, "segmentación también llamado agrupamiento, permite identificar gran similitud entre grupos de elementos y muchas diferencias entre otros, de esta manera se pueden segmentar los datos de acuerdo al objetivo de estudio". [19]

Asociación

Según Vallejo y Tenelanda, "Asociación encuentran patrones en los datos en los que una o más entidades, como eventos, compras o atributos, se asocian con una o más entidades. Los modelos construyen conjuntos de reglas que definen estas relaciones". "Aquí los campos de los datos pueden funcionar como entradas y destinos. Podría encontrar estas asociaciones manualmente, pero los algoritmos de reglas de asociaciones lo hacen mucho más rápido, y pueden explorar patrones más complejos". [20]

Según Microsoft, "asociación buscan correlaciones entre diferentes atributos de un conjunto de datos. La aplicación más común de esta clase de algoritmo es la creación de reglas de asociación". [22]

Según el trabajo Santiago Zapatakdd,"asociación establece posibles relaciones o correlaciones entre sucesos o hechos que aparentemente son independientes; son utilizadas cuando el objetivo es exploratoria buscando relaciones dentro del conjunto de datos". [19]

Clasificación

Según Vallejo y Tenelanda, "Utilizan el valor de uno o más campos de entrada para predecir el valor de uno o más resultados o campos de destino". [20]

Según Microsoft, "clasificación predicen una o más variables discretas, basándose en otros atributos del conjunto de datos". [22]

Según el trabajo de Santiago Zapatakdd," clasificación es un proceso cuyo fin es dividir el conjunto de datos que se consideran mutuamente excluyentes. Generando así que cada uno de los miembros se encuentra más cerca de uno similar y a su vez lejos de otros, donde la distancia se mide de acuerdo a las variables especificadas". [19]

A continuación se describen algunos de algoritmos más utilizados de clasificación.

ZeroR:- "es uno se de los algoritmos más sencillos, asigna a todos los resultados a la clase mayoritaria [23], implementado en WEKA calcula la media en el caso de tener una clase numérica o la moda en caso de una clase simbólica" [19].

Ridor:- "es la implementación de un aprendizaje regla ondulación hacia abajo propuesto por Gaines y Compton. Se genera una regla por defecto con el menor (ponderado) de tasa de error. Se genera la mejor excepción dentro de las excepciones.

Las excepciones son un conjunto de reglas que predicen categorías distintas de la predeterminada". [24]

Part:- "crea una serie de reglas utilizando los atributos más significativos para cada tipo de ataque .este clasificador, a pesar de que es bastante simple, da muy buen resultado". [23]

Decision Table:- "se resume un conjunto de datos con una tabla de decisión, que contiene el mismo número de atributos que contiene el conjunto de datos original" [24], "consiste en seleccionar subconjuntos de atributos y calcular su precisión para predecir o clasificar los ejemplos; una vez seleccionado el mejor de los subconjuntos, la tabla de decisión estará formada por los atributos seleccionados" [19].

ConjuntiveRule:- "implementa un solo aprendiz regla conjuntiva que puede predecir para las etiquetas de clase numérica y nominales. Una regla consta de antecedentes AND y consecuentes (valor de la clase) para la clasificación. En este caso el consecuente es la distribución de las calases disponibles en el conjunto de datos". [24]

DecisionStump:- "utiliza un único atributo para construir un árbol de decisión, la elección del único atributo se realiza basándose en la ganancia de información su implementación es compleja ya que admite tanto atributos numéricos como simbólicos y clases de ambos tipos también". [19]

C4.5 :-" el algoritmo conocido de WEKA es una implementación del algoritmo C4.5 uno de los más utilizados es un refinamiento del modelo generado con OneR" [23] , "el algoritmo utiliza la técnica codicioso y es una variante de ID3, que determina en cada paso el atributo más predictivo, y divide un nodo basado en este atributo. Cada nodo representa un punto de decisión sobre el valor de algún atributo; se ocupa de los

atributos numéricos de colocar donde se deben colocar los umbrales para las divisiones de decisión" [24].

RandomForest:- "es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos" [24].

BayesNet;-"algoritmo de aprendizaje utilizando diferentes algoritmos de búsqueda y medida de calidad. Se utiliza una clase base, proporciona estructuras de datos (estructuras de la red, distribuciones de probidad condicionada, etc)". [25]

Naive Bayes:- "es un clasificador probabilístico basado en aplicar el teorema de bayes (a partir de estadística bayesiana) con fuertes (ingenuas) independencias de supuestos. Naive Bayes asume que la presencia (o ausencia) de una característica particular de una clase no está relacionada con la presencia (o ausencia) de cualquier otra característica". [26]

En esta investigación más adelante se aplicara herramienta de minería de datos para seleccionar el algoritmo adecuado.

2.3 Herramientas de minería de datos

2.3.1 Antecedentes

"El proceso de extracción de patrones a partir de datos se llama minería de datos. Es reconocida como una herramienta esencial de los negocios modernos, ya que es capaz de convertir los datos en inteligencia de negocios dando así una ventaja de información. Actualmente, es ampliamente utilizado en las prácticas de perfil, como vigilancia, comercialización, descubrimientos científicos, y detección de fraudes". [27]

La minería de datos tiene:

- **“Clasificación** – Consiste en sistematizar una estructura familiar con la finalidad de aplicarla en los nuevos datos”. [27]
- **“Agrupamiento** – Consiste en formar conjuntos de datos que contengan similares características”. [27]
- **“Aprendizaje de reglas de asociación** – Encuentra características similares entre las variables”. [27]
- **“Regresión** Busca la función indicada que forme los datos con el mínimo error”. [27]

Las herramientas de minería de datos que se pueden obtener de forma gratuita son las siguientes:

ORANGE

“Orange es una potente aplicación diseñada para programadores expertos o principiantes y está destinada a realizar análisis y visualización de datos. También puedes utilizar el programa para minería de datos mediante la programación visual o programación en Python”. [28]

“El programa está repleto de características para el análisis de datos y también incluye componentes para el aprendizaje de máquina. La utilidad del programa puede ampliarse mediante complementos para minería de texto y bioinformática”. [28]

RAPIDMINER

“Microsystem, en su calidad de Partner Oficial de Rapid-i, empresa que desarrolla los sistemas RapidMiner y RapidAnalitics, cuenta con Certificaciones a nivel experto que le permiten brindar soluciones basadas en dichas herramientas.

RapidMiner es una herramienta de Minería de Datos ampliamente usada y probada a nivel internacional en aplicaciones empresariales, de gobierno y academia. Implementa más de 500 técnicas de pre-procesamiento de datos, modelación predictiva y descriptiva, métodos de prueba de modelos, visualización de datos, etc. Rapid-i, con base en Dormunt – Alemania, nació en 2006 como Spin-Off de la Universidad de Dortmund, donde se inauguró la primera versión del software en 2001. RapidMiner ha sido utilizada en más de cuarenta países y en compañías como Ford, Honda, E.ON, Nokia, IBM, Cisco, Hewlett Packard, Elexso, Akzo Nobel, PharmaDM, Bank of America, Merrill Lynch, entre muchas otras". [28]

JHEPWORK

"Diseñado para los científicos, ingenieros y estudiantes, jHepWork es un país libre y de código abierto de análisis de estructura de datos que se crea como un intento de hacer un análisis de entorno de datos usando paquetes de código abierto con una interfaz de usuario comprensible y para crear una herramienta competitiva para programas comerciales. Esto se hace especialmente para las parcelas científicas interactivas en 2D y 3D y contiene numérica bibliotecas científicas implementadas en Java para funciones matemáticas, los números al azar, y otros algoritmos de minería de datos. jHepWork se basa en un lenguaje de programación de alto nivel Jython, pero Java codificación también se puede utilizar para llamar a bibliotecas jHepWork numérica y gráfica". [29]

KNIME

"KNIME (Konstanz Information Miner) es de uso fácil y comprensible, y de fuente abierta de integración de datos, procesamiento, análisis, y la plataforma de exploración. Se ofrece a los usuarios la capacidad de crear de forma visual los flujos de datos o

tuberías, ejecutar selectivamente algunos o todos los pasos de análisis, y luego estudiar los resultados, modelos y vistas interactivas. KNIME está escrito en Java y está basado en Eclipse y hace uso de su método de extensión para apoyar plugins proporcionando así una funcionalidad adicional. A través de plugins, los usuarios pueden añadir módulos de texto, imagen, y el procesamiento de series de tiempo y la integración de varios otros proyectos de código abierto, tales como el lenguaje de programación de R, WEKA, el Kit de desarrollo de la Química, y LIBSVM". [29]

WEKA

"Escrito en Java, Weka (Waikato Enviroment for Knowledge Analysis) es una conocida suite de software para el aprendizaje y la máquina que soporta varias tareas de minería de datos típicos, especialmente los datos del proceso previo, el agrupamiento, clasificación, regresión, visualización y selección de características. Sus técnicas se basan en la hipótesis de que los datos están disponibles en un único archivo plano o una relación, donde se etiqueta cada punto de datos por un número fijo de atributos. WEKA proporciona acceso a bases de datos SQL utilizando Java Database Connectivity y puede procesar el resultado devuelto por una consulta de base de datos. Su interfaz de usuario principal es el Explorer, pero la misma funcionalidad que se puede acceder desde la línea de comandos a través de la interfaz basada en componentes de flujo de conocimientos". [29]

2.4 Ataques a las redes de información

2.4.1 Introducción

La seguridad en la informática cada día va tomando un gran avance, debido a las nuevas situaciones con las que se encuentran y sobre todo a las nuevas tecnologías y

plataformas que va apareciendo. Con la facilidad de conectarse a través de las redes sociales, la sociedad va encontrándose con nuevos espacios el cual le permite mejorar la productividad de cada uno de las personas, con esto también la aparición de nuevas amenazas para los sistemas de información.

“Debido a que el uso de Internet se halla en incremento, cada vez más empresas aprueban a sus socios y proveedores acceder a sus sistemas de información. Por lo tanto, es fundamental conocer qué recursos de la compañía necesitan amparo para así controlar el acceso al sistema y los derechos de los usuarios del sistema de información. Los mismos procedimientos se aplican cuando se permite el acceso a la compañía a través de Internet”. [30]

“Además, debido a la tendencia que crece hacia un estilo de vida nómada de hoy en día, el cual admite a los trabajadores conectarse a los sistemas de información casi a partir de cualquier lugar, se pide a los empleados que lleven consigo parte del sistema de información fuera de la infraestructura segura de la compañía”. [30]

“La amenaza simboliza el tipo de acción que tiende a ser perjudicial, mientras que la vulnerabilidad (también conocida a veces como falencias (flaws) o brechas (breaches)) constituye el grado de exposición a las amenazas en un contexto particular”. [30]

“Las contramedidas que se deben realizar no sólo son soluciones técnicas, sino también reflejan la capacitación y la toma de conocimiento por parte del usuario, además de reglas visiblemente definidas”. [30]

“Para que un sistema sea seguro, se debe considerar las posibles amenazas y por lo tanto, saber y prever el curso de acción del enemigo”. [30]

2.4.2 Antecedentes

“El Informe de Seguridad Anual 2014 de Cisco, revela que las amenazas diseñadas para aprovechar la confianza de los usuarios en sistemas, aplicaciones y redes personales han alcanzado niveles asombrosos. De acuerdo con el reporte, la falta de casi un millón de profesionales expertos en seguridad a nivel mundial está impactando las habilidades de las organizaciones de monitorear y asegurar las redes, mientras las vulnerabilidades y amenazas en general alcanzaron sus niveles más altos desde el año 2000.

Las conclusiones del informe ofrecen una imagen clara de los desafíos con respecto de los retos en seguridad que evolucionan rápidamente y que enfrentan hoy las empresas, los departamentos de TI y los individuos. Los métodos de los atacantes incluyen robo socialmente planeado de contraseñas y acreditaciones, infiltraciones escondidas a simple vista, y explotación de la confianza requerida para transacciones económicas, servicios de gobierno e interacciones sociales”. [31]

“Mayor sofisticación y proliferación del panorama de amenazas. Los simples ataques que causaban daño que podía ser contenido han dado lugar a operaciones de crimen organizado cibernético que es sofisticado, bien fundado, capaz de generar daño económico y de reputación a víctimas en el sector público y privado”. [31]

“Mayor complejidad de las amenazas y soluciones debido al rápido crecimiento de la adopción de dispositivos móviles inteligentes y la computación en la nube que brindan una mayor superficie de ataque como nunca antes. Los nuevos dispositivos y las nuevas arquitecturas de las infraestructuras ofrecen a los atacantes oportunidades de explotar debilidades no anticipadas y bienes defendidos inadecuadamente”. [31]

“Los criminales cibernéticos han aprendido que aprovechar el poder de la infraestructura de Internet rinde muchos más beneficios que simplemente ganar acceso a computadoras o dispositivos individuales. Estos ataques en la infraestructura buscan ganar acceso a servidores de web estratégicamente posicionados, servidores de nombre (nameservers) y centros de datos —con el objetivo de multiplicar los ataques en gran cantidad de bienes individuales provistos por estos recursos. Al tener como objetivo la infraestructura de Internet, los atacantes merman la confianza en todo lo conectado o habilitado por esta”. [31]

2.4.3 Definiciones

“Un ataque informático es un intento organizado e intencionado causado por una o más personas para causar daño o problemas a un sistema informático o red. Los ataques en grupo suelen ser hechos por bandas llamados "piratas informáticos" que suelen atacar para causar daño, por buenas intenciones, por espionaje, para ganar dinero, entre otras”. [32]

“Un ataque informático consiste en aprovechar alguna debilidad o falla en el software, en el hardware, e incluso, en las personas que forman parte de un ambiente informático; para obtener un beneficio, por lo general de condición económica, causando un efecto negativo en la seguridad del sistema, que luego pasa directamente en los activos de la organización”. [32]

“Según (Taype Espinoza Hebert Humberto) ataques informáticos defino como ataques realizados a través de una computadora con la finalidad de robar información o simplemente por diversión.

Según (EvilFingers) Un ataque informático consiste en aprovechar alguna debilidad o falla (vulnerabilidad) en el software, en el hardware, e incluso, en las personas que forman parte de un ambiente informático; a fin de obtener un beneficio, por lo general de índole económico, causando un efecto negativo en la seguridad del sistema, que luego repercute directamente en los activos de la organización". [33]

De acuerdo a las definiciones se dice que un ataque de la forma que se realice pretende hacer daño y obtener un beneficio que en la mayoría concuerda que es económico.

2.4.4 Tipos de ataques a las redes de información

2.4.4.1 Ataques de denegación de servicio (DDos)

"En Internet, un ataque de denegación de servicio (DDoS) es el que se realiza cuando una cantidad considerable de sistemas atacan a un objetivo único, provocando la denegación de servicio de los usuarios del sistema afectado. La sobrecarga de mensajes entrantes sobre el sistema objetivo fuerza su cierre, denegando el servicio a los usuarios legítimos". [34]

"En un ataque DDoS típico el hacker (o si lo prefiere cracker) empieza buscando una vulnerabilidad en un sistema informático y creando un master para el DDoS. Desde este master el sistema identifica y se comunica con otros sistemas que pueda utilizar. El atacante usa las herramientas de cracking disponibles en Internet sobre cientos o miles de equipos. Con un solo comando el atacante puede controlar todas estas máquinas para que lancen un ataque simultáneo sobre un objetivo concreto. La avalancha de paquetes provoca el error de denegación de servicio". [34]

"Mientras que la prensa tiende a centrarse en las victimas de los ataques DDoS, lo cierto es que hay más afectados por estos ataques – como son todos los sistemas afectados y

controlados por el intruso. Aunque los propietarios de estos equipos no siempre están al tanto de la debilidad de sus equipos si es cierto que pueden sufrir de errores, problemas de funcionamiento y degradación del servicio. Tanto los propietarios como los usuarios del sitio afectado sufren los efectos del ataque. Yahoo, Buy.com, RIAA o la oficina de Copyright de Estados Unidos son algunas de las víctimas de estos ataques DDoS. Unos ataques que también pueden provocar mayores daños por encadenamiento. Por ejemplo, en octubre de 2012 un ataque masivo DDoS dejo a todo el país de Myammar desconectado". [34]

"Al equipo que cae bajo el control del intruso se le llama bot o zombie. Al grupo de ordenadores afectados se le llama botnet o armada zombi. Tanto Kaspersky Labs como Symantec consideran los botnets y no el spam, los virus ni los gusanos como la mayor amenaza de seguridad en internet". [34]

2.4.4.2 Ataque R2L acceso no autorizado desde una máquina remota

"Cuando un atacante que no dispone de cuenta alguna en una máquina, logra acceder (tanto como usuario o como root) a dicha máquina. En la mayoría de los ataques R2L, el atacante entra en el sistema informático a través de Internet". [35]

"Cuando un atacante que no dispone de cuenta alguna en una máquina, logra acceder (tanto como usuario o como root) a dicha máquina. En la mayoría de los ataques R2L, el atacante entra en el sistema informático a través de Internet. Hay varias maneras en que un atacante puede lograr su objetivo [Kendall, 99].

Algunos ataques explotan el desbordamiento de búfer causado por el software de servidor de red "imap, named, sendmail". Los ataques de " ftp_write, xsnoop y guest" tratan de explotar la debilidad o la mala configuración de las políticas de seguridad del

sistema. El ataque "xlock" utiliza ingeniería social para tener éxito, el atacante debe suplantar a los operadores humanos que proporcionan sus contraseñas de los protectores de pantalla que en realidad son caballos de Troya". [35]

2.4.4.3 Ataque U2R acceso no autorizado mediante escalamiento de privilegios de una cuenta de usuario autorizado hasta llegar a superusuario

"Este tipo de ataque se da cuando un atacante que dispone de una cuenta en un sistema informático es capaz de elevar sus privilegios explotando vulnerabilidades en los mismos, un agujero en el sistema operativo o en un programa instalado en el sistema. Hay varios tipos de ataques U2R [Ken98] La más común es el ataque "buffer_overflow" que se produce cuando un programa copia una gran cantidad de datos en un búfer de memoria estática sin comprobar si el tamaño de esta última es suficiente, lo que provocará un desbordamiento. Los datos desbordados se almacenan en la pila de sobrecarga del sistema, cubriendo así las siguientes instrucciones para ser ejecutadas. Mediante la manipulación cuidadosa de los datos almacenados en la pila, un atacante puede provocar la ejecución de código en el sistema operativo que le ayudará a conseguir lo que quiere. Otra clase de ataques U2R explotan los programas que proporcionan información sobre el medio en el que se ejecutan, un buen ejemplo de este tipo de ataque es el ataque "lodamodule". Otra clase de ataques U2R explotan los programas que tienen una mala gestión de los archivos temporales. Algunos ataques U2R explotan la vulnerabilidad debido a las condiciones competitivas explotables durante la ejecución de un solo programa, dos o más programas se ejecutan simultáneamente [Gar96]. A pesar de que una programación controlada podría eliminar todas estas vulnerabilidades, tales errores están presentes en todas las versiones de UNIX y Windows de Microsoft disponibles hoy en día". [35]

2.4.4.4 Monitorización

“Este tipo de ataques escanean las redes tratando de identificar direcciones IP válidas y recoger información acerca de ellas (servicios que ofrecen, sistemas operativos que usan). A menudo, esta información provee al atacante una lista de vulnerabilidades potenciales que podrían ser utilizadas para llevar a cabo ataques a los servicios y a las máquinas escogidas. Estos ataques son los más frecuentes, y a menudo son precursores de otros ataques. Un atacante con un mapa de las máquinas y servicios disponibles en una red puede utilizar esta información para encontrar todos los puntos débiles de esta última. Algunas de estas herramientas de análisis "satan, saint, mscan” permiten que incluso un hacker principiante, pueda revisar rápidamente cientos o miles de máquinas en una red”. [35]

“Este tipo de ataque se ejecuta para observar a la víctima y su sistema, con el objetivo de construir sus vulnerabilidades y posibles formas de acceso futuro. Se puede presentar como”: [32]

Shoulder Surfing

“Otro tipo de ataque relacionado con la ingenuidad de los usuarios del sistema (pero también con el control de acceso físico) es el denominado shoulder surfing. Consiste en `espiar' físicamente a los usuarios, para obtener generalmente claves de acceso al sistema. Por ejemplo, una medida que lamentablemente utilizan muchos usuarios para recordar sus contraseñas es apuntarlas en un papel pegado al monitor de su PC o escribirlas en la parte de abajo del teclado; cualquiera que pase por delante del puesto de trabajo, sin problemas puede leer el login, password e incluso el nombre de máquina a la que pertenecen. Esto, que nos puede parecer una gran tontería, por desgracia no lo es,

y se utiliza más de lo que muchos administradores o responsables de seguridad piensan; y no sólo en entornos `privados' o con un control de acceso restringido, como pueda ser una sala de operaciones de un centro de cálculo, sino en lugares a los que cualquiera puede llegar sin ninguna acreditación: personalmente, hace unos años pude leer claramente `post-it' pegados a los monitores de los PCs utilizados por el personal de información de unos grandes almacenes de Valencia, en los que aparecían el nombre de usuario, la clave y el teléfono de varios sistemas de la empresa; cualquiera que se acercase al mostrador podía leer y memorizar esta información sin problemas. El shoulder surfing no siempre se ve beneficiado por la ingenuidad de los simples usuarios de un equipo; en determinadas ocasiones son los propios programadores (gente que teóricamente ha de saber algo más sobre seguridad que el personal de administración o de atención al público) los que diseñan aplicaciones muy susceptibles de sufrir ataques de este tipo. Por ejemplo, en ciertas aplicaciones especialmente algunas que se ejecutan sobre MS Windows, y que son más o menos antiguas muestran claramente en pantalla las contraseñas al ser tecleadas. Cualquiera situado cerca de una persona que las está utilizando puede leer claramente esa clave; un perfecto ejemplo de lo que No se debe hacer nunca". [36]

Decoy (Señuelos)

Los Decoy son programas diseñados con la misma interface que otro original. En ellos se imita la solicitud de un logeo y el usuario desprevenido lo hace. Luego, el programa guardará esta información y dejará paso a las actividades normales del sistema. La información recopilada será utilizada por el atacante para futuras "visitas". Una técnica semejante es aquella que, mediante un programa se guardan todas las teclas

presionadas durante una sesión. Luego solo hará falta estudiar el archivo generado para conocer nombres de usuarios y claves. [37]

Scanning (Búsqueda)

El Scaneo, como método de descubrir canales de comunicación susceptibles de ser explotados, lleva en uso mucho tiempo. La idea es recorrer (scanear) tantos puertos de escucha como sea posible, y guardar información de aquellos que sean receptivos o de utilidad para cada necesidad en particular. Muchas utilidades de auditoría también se basan en este paradigma.

El Scaneo de puertos pertenece a la Seguridad Informática desde que era utilizado en los sistemas de telefonía. Dado que actualmente existen millones de números de teléfono a los que se pueden acceder con una simple llamada, la solución lógica (para encontrar números que puedan interesar) es intentar conectarlos a todos.
La idea básica es simple: llamar a un número y si el módem devuelve un mensaje de conectado, grabar el número. En otro caso, la computadora cuelga el teléfono y llama al siguiente número. [37]

Snooping–Downloading

“Los ataques de esta categoría tienen el mismo objetivo que el Sniffing: obtener la información sin modificarla. Sin embargo los métodos son diferentes. Aquí, además de interceptar el tráfico de red, el atacante ingresa a los documentos, mensajes de correo electrónico y otra información guardada, realizando en la mayoría de los casos un downloading (copia de documentos) de esa información a su propia computadora, para luego hacer un análisis exhaustivo de la misma. El Snooping puede ser realizado por

simple curiosidad, pero también es realizado con fines de espionaje y robo de información o software. Los casos más resonantes de este tipo de ataques fueron: el robo de un archivo con más de 1700 números de tarjetas de crédito desde una compañía de música mundialmente famosa, y la difusión ilegal de reportes oficiales reservados de las Naciones Unidas, acerca de la violación de derechos humanos en algunos países europeos en estado de guerra". [37]

2.4.5 Descripción de los ataques

"(computer attack). Intento organizado y deliberado de una o más personas para causar daño o problemas a un sistema informático o red". [32]. Los ataques en grupo suelen ser hechos por bandas de piratas informáticos por diversión, para causar daño, buenas (relativamente buenas) intenciones, espionaje, obtención de ganancias, etc. Los blancos preferidos suelen ser los sistemas de grandes corporaciones o estados, pero ningún usuario de internet u otras redes está exento. Probablemente el primer ataque informático masivo de la historia se produjo un viernes 13 de 1989, cuando una revista informática regaló disquetes de promoción, pero estaban infectados con un virus. El virus afectó a cientos de empresas y particulares. Actualmente los ataques suelen afectar principalmente a internet, pero también afectan otras redes como las de telefonía, redes Wi-Fi, redes de área local de empresas, etc". [38]

2.4.6 Indicios de una intrusión

"Cuando la gente oye hablar de sistemas de detección de intrusiones, generalmente los asocia a ´´alarmas de ladrones para ordenadores o redes´´. Es fácil entender un concepto como este usando comparaciones sencillas. En realidad, la explicación es bastante aproximada, y los usuarios que no se dedican a la seguridad no necesitan saber más. Sin

embargo, los expertos en seguridad no pueden cometer el error de conformarse con algo tan trivial, sin tener conocimiento alguno sobre la historia de estos sistemas". [39]

2.5 Sistema de detección y prevención de intrusos IDS/IPS

2.5.1 Sistema de detección de intrusos

"El término IDS (Sistema de detección de intrusiones) hace referencia a un mecanismo que, silenciosamente, escucha el tráfico en la red para detectar actividades anormales o sospechosas, y de este modo, disminuir el riesgo de intrusión". [40]

Existen dos claras familias importantes de IDS:

- "El grupo **N-IDS** (Sistema de detección de intrusiones de red), que certifica la seguridad dentro de la red". [40]
- "El grupo **H-IDS** (Sistema de detección de intrusiones en el host), que garantiza la seguridad en el ordenador". [40]

"Un N-IDS requiere un hardware exclusivo. Éste forma un sistema que puede confirmar paquetes de información que recorren por una o más líneas de la red para expresar si se ha producido una actividad maliciosa o anormal. El N-IDS pone uno o más de los adaptadores de red exclusivos del sistema en modo promiscuo. Éste es una especie de modo "invisible" en el que no tienen dirección IP. Tampoco tienen una serie de protocolos situados". [40]

"Es común localizar varios IDS en diferentes partes de la red. Por lo general, se instalan sondas fuera de la red para aprender los posibles ataques, así como también se ubican sondas internas para examinar solicitudes que hayan pasado a través del firewall o que se han realizado desde dentro como se puede observar en la Figura II 7." [40]

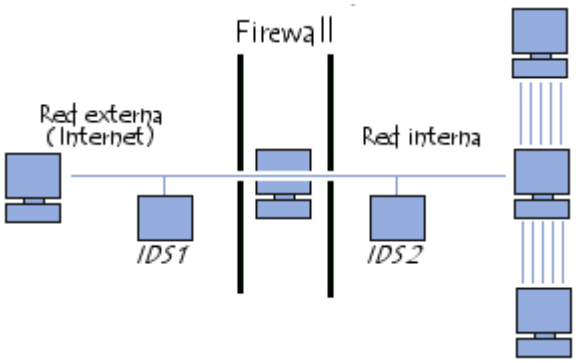


Figura II 7: Uso de los IDS
Fuente: es.kioskea.net [40]

"El H-IDS se localiza en un host particular. Por lo tanto, su software cubre una extensa gama de sistemas operativos como Windows, Solaris, Linux, HP-UX, Aix, etc". [40]

"El H-IDS actúa como un daemon o servicio estándar en el sistema de un host. Tradicionalmente, el H-IDS examina la información particular acumulada en registros (como registros de sistema, mensajes, lastlogs y wtmp) y también captura paquetes de la red que se introducen/salen del host para poder confirmar las señales de intrusión (como ataques por denegación de servicio, puertas traseras, troyanos, intentos de acceso no autorizado, ejecución de códigos malignos o ataques de desbordamiento de búfer)". [40]

2.5.2 Sistemas de prevención de intrusos

Los sistemas de prevención de intrusiones (IPS) protegen a la red ante una amplia gama de ataques a la seguridad.

- "Detección de firmas de estado: las firmas sólo se emplean a segmentos notables del tráfico de red determinados por el contexto de protocolos adecuado, con lo que los falsos positivos se reducen al mínimo". [41]
- "Detección de anomalías de protocolo: el uso de protocolos se comprueba frente a un RFC publicado para descubrir violaciones o abusos, y así se resguarda de forma

proactiva la red frente a intrusiones e incluso frente a vulnerabilidades no descubiertas". [41]

- "Detección de anomalías de tráfico: las reglas heurísticas facilitan detección a partir de patrones de tráfico inesperados que pueden proponer reconocimiento o ataques. Este sistema de prevención de intrusiones comunica de forma proactiva actividades de reconocimiento y bloquea ataques de denegación de servicio distribuidos (DDoS)". [41]
- "Administración basada en funciones: logran asignarse más de 100 actividades diferentes como permisos exclusivos para distintos administradores, lo que aligera las operaciones del negocio al apartar y aplicar de forma lógica las funciones de varios administradores". [41]
- "Las funciones del sistema de prevención de intrusiones se estrechan a las operaciones empresariales: habilitar el alejamiento lógica de dispositivos, políticas, informes y otras actividades de gestión para concentrar dispositivos en función de prácticas empresariales". [41]

2.5.3 Modos de detección de ataques de los IDS/IPS

Detección basada en firmas

"Una firma posee la capacidad de reconocer una determinada cadena de bytes en cierto contexto, para agilizar o activar una alerta. Por ejemplo, los ataques hacia los servidores Web generalmente toman la forma de URLs. Por lo tanto se puede averiguar haciendo uso de un patrón de cadenas que pueda identificar claramente a los ataques al servidor Web. Sin embargo, como este tipo de detección marcha parecido a un antivirus, el administrador debe comprobar que las firmas estén constantemente actualizadas". [41]

Detección basada en políticas

"En este tipo de detección, el IPS necesita que se expongan muy claramente las políticas de seguridad. Por ejemplo establecer que hosts pueden tener comunicación con determinadas redes. El IPS examina el tráfico fuera del perfil permitido y lo descarta. Necesita de un gran trabajo por parte del administrador de red y es menos propenso a Falsos Veredictos, pero esto no es tan completo o flexible como la detección por firmas". [41]

Detección basada en anomalías.

"Radica en descubrir condiciones anormales de la red. Para ello el dispositivo debe ingresar primero en un modo de auto aprendizaje para averiguar umbrales de normalidad y para que el administrador pueda perfeccionar dado los Falsos Veredictos. Este tipo de detección es propensa a crear muchos falsos positivos, ya que es sumamente difícil establecer y calcular una condición 'normal'. En este tipo de detección existen dos opciones": [41]

Detección Estadística de anormalidades: "El IPS examina el tráfico de red por un periodo específico de tiempo y crea una línea base de comparación. Cuando el tráfico varía excesivamente con respecto a la línea base de comportamiento se genera una alarma". [41]

Detección no Estadística de anormalidades: "En este tipo de detección, es el administrador quien precisa el patrón 'normal' de tráfico. Sin embargo, debido a que con este enfoque no se ejecuta un análisis dinámico y real del uso de la red, es susceptible a generar muchos falsos positivos". [41]

Detección Honey Pot

“Aquí se utiliza un 'distractor'. Se asigna como honey pot un dispositivo que pueda lucir como atractivo para los atacantes. Los atacantes utilizan sus recursos para tratar de ganar acceso en el sistema y dejan intactos los verdaderos sistemas. Mediante esto, se puede monitorear los métodos utilizados por el atacante e incluso identificarlo, y de esa forma implementar políticas de seguridad acordes en nuestro sistemas de uso real”. [41]

2.5.4 Diferencia entre IPS e IDS

“Es común percibir en el concepto del IPS como “un IDS con capacidades de bloqueo”. Esta noción es correcta, mas no el concepto general de los IPS. Las diferencias entre estos dos dispositivos no se localizan únicamente en su funcionalidad sino asimismo en sus prioridades y características”. [41]

“Un IDS es una herramienta de detección pasiva, no requiere estar en medio del tráfico para examinar los paquetes, le basta con recoger una copia del mismo. La caída del sistema resulta en un dolor de cabeza para el analista de red que perdió información valiosa durante el tiempo de fallo. El IPS, por otro lado, está en medio del tráfico tomando todos los paquetes y decidiendo cual puede o no transitar, la caída del sistema es mucho más catastrófica que en los IDS dado que la red entera se puede caer junto al IPS si es que no existe un sistema de recuperación de fallos configurado en la solución”. [41]

“En desempeño, los IPS necesitan estar en equipos más robustos debido a que de su velocidad de procesamiento de paquetes dependerá el throughput máximo al que estará limitado dicha red”. [41]

“En las redes comúnmente se dan ráfagas de tráfico llamadas “bursts”. Durante estos picos de tráfico, los IDS logran recopilar en buffers de memoria los paquetes que no alcanza a procesar, para examinarlas cuando el tráfico se estabilice. Los IPS, no pueden almacenar en buffers tantos paquetes como los IDS debido a que esto crecería la latencia del dispositivo y por ende de la red”. [41]

Durante ráfagas, los IPS tienen que descartar paquetes.

“En cuanto a detección, los IDS logran y suelen ejecutar errores; falsos positivos si es que alerta un evento que en realidad no es peligroso, y falsos negativos si es que no alerta un evento que en realidad sí es ataque”. En los IDS, los falsos negativos son más perjudiciales, pues hurtan información de amenazas al analista de red. En cambio con los IPS, un falso positivo resulta en más que una alerta falsa, sino que acaba en un bloqueo de paquetes inofensivos.

Los módulos de detección de anomalías son muy buenos para descubrir actividades sospechosas que por alguna razón no pueden generalizarse en una regla. El problema con estos módulos es que pueden inventar falsos positivos y como ya se explicó, los IPS no pueden cometer errores. Por esta razón este tipo de módulos son fuertemente recomendados únicamente para los motores IDS.

La siguiente tabla resume lo analizado en esta sección sobre las diferencias entre los Sistemas de Detección y Prevención de Intrusos como se puede ver en la Tabla II II [41].

Tabla II II: Diferencia entre IDS e IPS

	IPS	**IDS**
	“Inline, Bloque Automático”	“Mirror, Alertas para Analistas”
“Estabilidad”	“Caída del sistema es	“Caída del sistema quita

	catastrófica para la red"	información al analista de red. No es algo crítico"
"Desempeño"	"Solicita mayor capacidad de procesamiento. Puede producir cuellos de botella".	"La falta de procesamiento puede ser compensada con buffers de mucha memoria. Nunca producirá cuellos de botella".
"Precisión de Falsos Positivos"	"Produce bloqueos de paquetes. Problema con aplicaciones".	"Carga trabajo innecesaria Para el analista en busca de falsas alarmas".
"Precisión de Falsos Negativos"	"Paquetes maliciosos ingresan a la red. No es tan crítico como en el caso de los IDS".	"Ataques resultan completamente invisibles y pueden retornar a ocurrir. Pérdida de información Para el analista".

Fuente: ESCUELA SUPERIOR POLITECNICA DEL LITORAL [41]

2.6 Motores de detección de intrusos

2.6.1 Antecedentes

"Cuando la gente oye hablar de Sistemas de Detección de Intrusiones, generalmente los asocia a "alarmas de ladrones para ordenadores o redes". [39]

"La detección de intrusiones es el fruto de la aplicación del Procesamiento Electrónico de Datos (EDP) a las auditorías de seguridad, utilizando mecanismos de identificación de patrones y métodos estadísticos. Es una parte imprescindible en las modernas tecnologías de seguridad de redes". [39]

2.6.2 Snort

"Snort: es una fuente de prevención de intrusiones de red abierta y el sistema de detección (IDS / IPS) desarrollado por Sourcefire . La combinación de los beneficios de la firma, el protocolo y la inspección anomalía basado en Snort es el mayor despliegue IDS / IPS de tecnología en todo el mundo. Con millones de descargas y cerca de 400.000 usuarios registrados, Snort ha convertido en el estándar de facto para IPS". [42]

"Snort es un sniffer de paquetes y un detector de intrusos basado en red (se monitoriza todo un dominio de colisión). Es un software muy flexible que ofrece capacidades de almacenamiento de sus bitácoras tanto en archivos de texto como en bases de datos abiertas como lo es MySQL[3]. Implementa un motor de detección de ataques y barrido de puertos que permite registrar, alertar y responder ante cualquier anomalía previamente definida. Así mismo existen herramientas de terceros para mostrar informes en tiempo real (ACID) o para convertirlo en un Sistema Detector y Preventor de Intrusos". [43]. Este IDS realiza un lenguaje de creación de reglas flexibles, potentes y sencillas. Durante su instalación ya nos suministra de cienes de filtros o reglas para backdoor, DDoS, finger, FTP, ataques web, CGI, Nmap. [43]

"Puede marchar como sniffer (podemos ver en consola y en tiempo real qué ocurre en nuestra red, todo nuestro tráfico), registro de paquetes (admite almacenar en un archivo los logs para su posterior análisis, un análisis offline) o como un IDS normal (en este caso NIDS). Cuando un paquete concuerda con algún patrón establecido en las reglas de configuración, se logea. Así se sabe cuándo, de dónde y cómo se produjo el ataque". [43].

[3] Es un sistema de gestión de bases de datos relacional, multihilo y multiusuario con más de seis millones de instalaciones

“Aun cuando tcpdump es considerada como una herramienta de auditoría muy útil, no se considera un verdadero IDS puesto que no estudian ni señala paquetes por anomalías. tcpdump ilustra toda la información de paquetes a la salida en pantalla o a un archivo de registro sin ningún tipo de análisis. Un verdadero IDS analiza los paquetes, marca las transmisiones que sean potencialmente maliciosas y la acumula en un registro formateado, así, Snort maneja la biblioteca estándar libcap y tcpdump como registro de paquetes en el fondo”. [43]

“Snort está disponible bajo licencia GPL, gratuito y marcha bajo plataformas Windows y UNIX/Linux. Dispone de una gran cantidad de filtros o patrones ya predefinidos, así como actualizaciones constantes ante casos de ataques, barridos o vulnerabilidades que vayan siendo detectadas a través de los distintos boletines de seguridad”. [43]

“La característica más apreciada de Snort, además de su funcionalidad, es su subsistema flexible de firmas de ataques. Snort tiene una base de datos de ataques que se está actualizando constantemente y a la cual se puede añadir o actualizar a través de la Internet. Los usuarios pueden crear 'firmas' basadas en las características de los nuevos ataques de red y enviarlas a la lista de correo de firmas de Snort, para que así todos los usuarios de Snort se puedan beneficiar. Esta ética de comunidad y compartir ha convertido a Snort en uno de los IDS basados en red más populares, actualizados y robustos”, como muestra la Figura II 8 Captura de la consola del sistema”. [43]

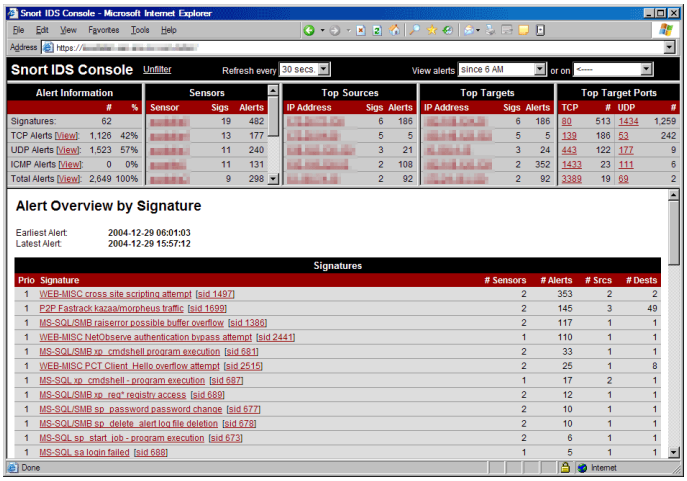


Figura II 8: Consola del Snort
Fuente: jgdasir2.files.wordpress.com

2.6.3 Suricata

Características de la herramienta Suricata

"Multi-Threading: Esta característica es uno de los fuertes de suricata pues las soluciones software libre actual de IDS/IPS y algunas de fabricantes, son uni-threaded. Multi-Threading consiste en procesar los paquetes en uno o más Hilos (Threads), y de esta manera se puede aprovechar el procesamiento multinúcleos de los actuales procesadores, haciendo que cada núcleo del procesador se encargue del procesamiento de uno o más hilos. Esto no es posible con soluciones como snort que son uni-threaded". [41]

"Estadísticas de Rendimiento: Este módulo se encarga de contar elementos de rendimiento como nuevas tramas/sec, duración, etc; y almacena esta información para presentarlos como estadísticas al administrador de alguna manera, ya sea vía logs, mensajes SNMP, vía web, etc". [41]

“Detección de Protocolos Automático: El motor de suricata tiene palabras claves para algunos protocolos como: IP, TCP, UDP, ICMP, HTTP, TLS, FTP y SMB. Esto quiere decir que se puede detectar una ocurrencia dentro de un stream de datos, sin importar el puerto en donde ocurre. Esta característica es importante para el control y detección de malware”. [41]

“Descompresión Gzip: Con la ayuda de la librería HTP es posible descomprimir un archivo en Gzip para examinarlo en busca de patrones de ataque”. [41]

“Independencia de la librería HTP: La librería HTP es un proyecto independiente a suricata e integrado efectivamente a suricata. Puede ser utilizado por otras aplicaciones como: proxies, filtros; y hasta el módulo mod_security de Apache”. [41]

“Métodos de Entrada Estándar: Soporte para NFQueu, IPFRing y LibPcap standard para la captura de tráfico”. [41]

“Unified2 Output: Soporte para métodos y herramientas de salida estándar Unified2. Este tipo de salida binario busca reducir carga al Suricata en cuanto a parseo de la información de salida, dejándole el trabajo a soluciones externas como Barnyard quien agarra los binarios, los interpreta y los almacena según donde configure el administrador”. [41]

Fast IP Matching: El motor de suricata puede usar automáticamente un preprocesador especial para validar más rápidos las reglas que hagan coincidencia únicamente de IP; por ejemplo, RBN, o las listas de ip de “EmergingThreats”. [41]

“HTTP Log Module: Las peticiones de HTTP pueden retornar una respuesta con el formato de log de apache para monitoreo y registro de actividad. Es posible usar Suricata únicamente como HTTP sniffer con este módulo”. [41]

“**IP Reputation:** Consiste en compartir información de direcciones IP de mala reputación con otras organizaciones y soluciones de seguridad, para eliminar falsos positivos. Este módulo se encarga de colectar, almacenar, actualizar y distribuir el conocimiento de la reputación de direcciones IP. Esta calificación puede ser positiva o negativa y clasifica a las IP en categorías”. [41]

“Funcionaría bajo una estructura “Hub and Spoke” donde en una base de datos central (Hub) se almacenaría y procesaría toda la información recolectada por los IDS para luego ser compilada y distribuida a la base de datos de los IDS clientes (Spokes)”. [41]

“La implementación técnica está dividida en tres componentes: La integración con el motor, el Hub, o base de datos central, que redistribuirá las reputaciones y el protocolo de comunicación entre Hubs y sensores (IDS‟s)”. [41]

“Graphics Card Acceleration: usando CUDA y OpenCL se puede utilizar el poder de procesamiento de las tarjetas gráficas para acelerar el IDS y quitarle carga al procesador principal para ganar rendimiento. Este módulo ya está incluído desde la versión 0.9.x de suricata pero aún sin los resultados esperados. Aún en desarrollo”. [41]

“Windows Binaries: Suricata también puede correr en Windows, sobre versiones mayores a XP, aunque no es muy recomendable hacerlo pues es una solución inicialmente desarrollada para Linux que se caracteriza por ser más estable que Windows”. [41]

“Flowint: Permite la captura, almacenamiento y comparación de datos en una variable global, es decir que permitirá comparar datos de paquetes provenientes de streams no relacionadas también. Se puede usar para un buen número de cosas útiles como contar ocurrencias, sumar o restar ocurrencias, activar una alarma al obtener un número x de ocurrencias, etc”. [41]

“**Módulo de capa de aplicación SSH:** El interpretador SSH interpreta sesiones SSH y detiene la detección e inspección del flujo de datos después de que la parte de encriptación ha sido inicializada. Este módulo implementado en la versión estable 1.0.2, se concentra en reducir el número de paquetes que necesita inspección al igual que los módulos SSL y TSL (7)”. [41]

Más sobre el Multi-threading de Suricata

“Como se ha mencionado, Suricata funciona a base de multi-hilos, usa múltiples núcleos del CPU para procesar paquetes de manera simultánea. Si está en un CPU con un solo núcleo los paquetes serán procesados uno por uno. Existen 4 módulos por hilo de CPU: Adquisición de paquete, decodificación de paquete, capa de flujo de datos, detecciones y salidas. El módulo de adquisición de paquete lee el paquete desde la red. El módulo de decodificación interpreta el paquete y se encarga de gestionar a qué stream pertenece qué paquete; el módulo de capa de flujo realiza 3 tareas:

La primera, realiza lo que se conoce como „tracking“ o rastreo de flujo, que asegura que todos los pasos que se están siguiendo tienen una conexión de red correcta. La segunda: el tráfico de red TCP viene en paquetes, por lo tanto el motor de este módulo reconstruye el stream original. Finalmente la capa de aplicación será inspeccionada, tanto el flujo HTTP como DCERPC (Distributing Computing Environment Remote

Procedure Call) será analizado. Los hilos de detección compararan firmas, pueden existir varios hilos de detección que pueden trabajar simultáneamente. En el módulo de detecciones y salidas, todas las alertas y eventos serán analizados". [41]

FODA SURICATA

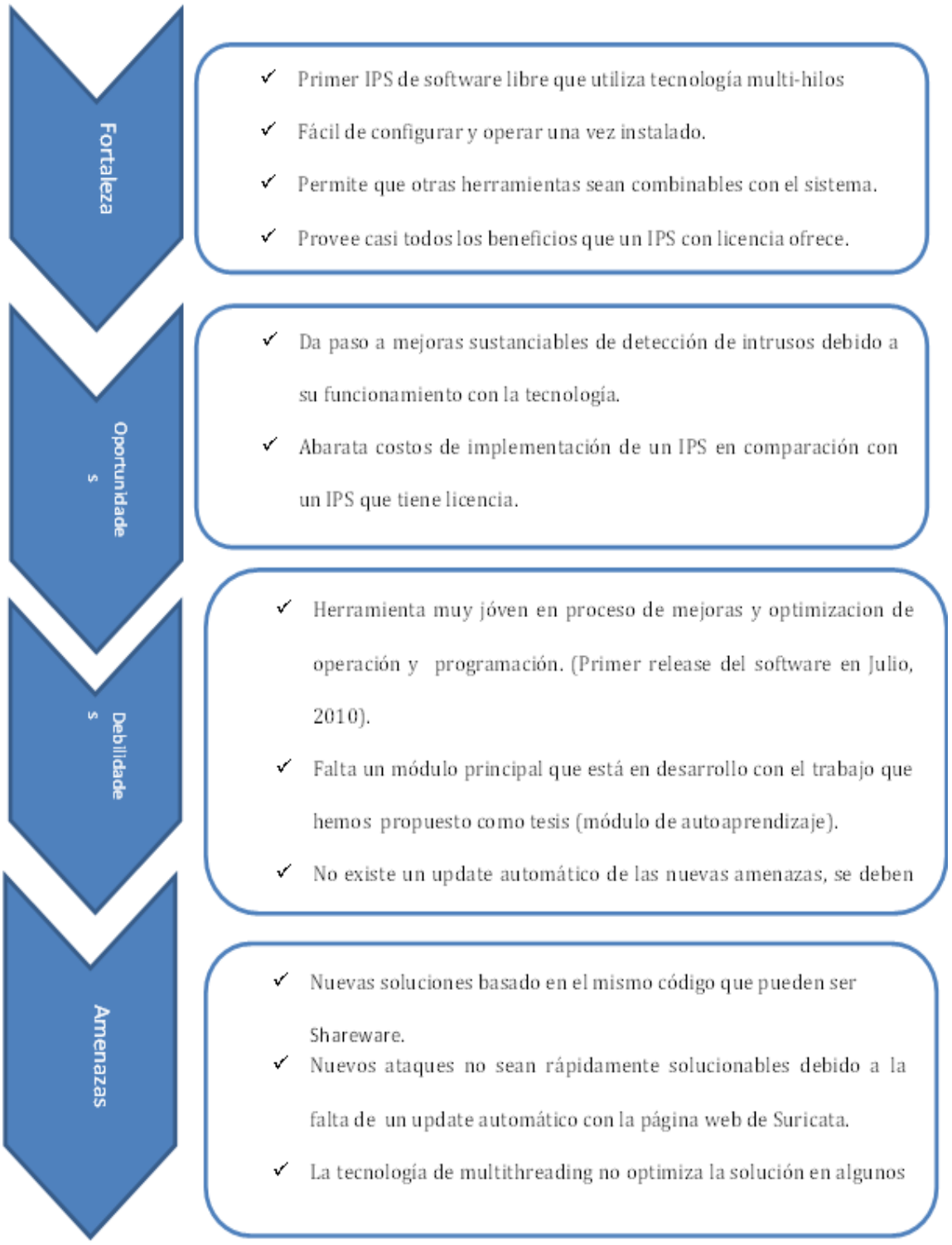


Figura II 9: Foda Suricata
Fuente: ESPOL

2.7 Selección del IDS

En este modelo se hace uso de la herramienta de Suricata y se trata de "una solución reciente y gratuita, es entendible que las características de las soluciones propietarias van a superar a las de este IDS/IPS". [41] "La desventaja de poder adquirir una solución propietaria es que se requiere de una gran inversión dependiendo del fabricante, desempeño, robustez y la inversión va a variar desde los 5000 hasta los 200.000 dólares.

De acuerdo a la investigación realizada por unos tesistas de la Escuela Superior Politécnica del Litoral que compara algunas características de Suricata con otras soluciones propietarias en general". [41] . Resultados que se ilustra en la Tabla II III.

Tabla II III: Comparación de Suricata con IPS/IDS"s propietarias

Características	Soluciones comerciales IDS/IPS	Suricata
Multi-Threading	X	Si
Soporte para IPV6	Cisco, IBM Stonesoft	Si
IP Reputation	Cisco	Si
Detección automática de protocolos	No	Si
Aceleración con GPU	No	Si
Varaibles Globales/Flowbit	No	Si
GeoIp	No	Si
Análisis Avanzado de HTTP	No	Si
HTTP Access Logging	No	Si
SMB Access Logging	No	Si
Anomaly Detection	Si	No
Alta Disponibilidad	Si	No
GUI de Administración	Si	No
Gratis	No	Si

Fuente: Escuela Superior Politécnica del Litoral

"Cuando de soluciones de código abierto se trata existen los que solamente trabajan como IDS (Detección de Intrusos), siendo Bro uno de los más populares.

En cuanto a Detección y Prevención de intrusos, el más conocido es SNORT, un IDS/IPS desarrollado por Sourcefire". [41]

"Actualmente suricata posee características únicas que no se encuentran en snort (Tampoco en otros IPS) y de hecho, muchas veces se menciona a suricata como una actualización o mejora basada en snort". [41]

La tabla compara características de Suricata con Bro y Snort que son de código abierto como se puede ver en la Tabla II III.

Tabla II III: Comparación de Suricata con IPS/IDS"s de código abierto

Características	Bro	Snort	Suricata
Multi-Threading	No	X	Si
Soporte para IPV6	Si	Si	Si
IP Reputattion	Algo	No	Si
Detección Automática de Protocolos	Si	No	Si
Aceleración con GPU	No	No	Si
Variables Globales/Flowbits	Si	No	Si
GeoIP	Si	No	Si
Análisis Avanzado de HTTP	Si	No	Si
HTTP Access logging	Si	No	Si
SMB Access Logging	Si	No	Si

Fuente: Escuela Superior Politécnica del Litoral

La gran ventaja que Suricata posee por ser una solución de código abierto, se le puede agregar nuevas funcionalidades al sensor de acuerdo a las necesidades.

CAPÍTULO III

MODELO PROPUESTO

3.1 Introducción

Este capítulo se lo ha dividido en dos partes: La primera se refiere a la selección de la herramienta de minería de datos y las técnicas apropiadas para la detección de intrusos, y la segunda parte la conforma el modelo que se propone para la detección de intrusos, el cual integra la herramienta y la técnica seleccionada anteriormente. El modelo que se propone en este trabajo se basa en la investigación de Liu Wenjun. [44]

3.2 Selección de la Herramienta de Minería De Datos

El procedimiento que se ha realizado para la selección de la herramienta de minería de datos y la técnica a utilizar en esta investigación, es el que se indica en la Figura III 10.

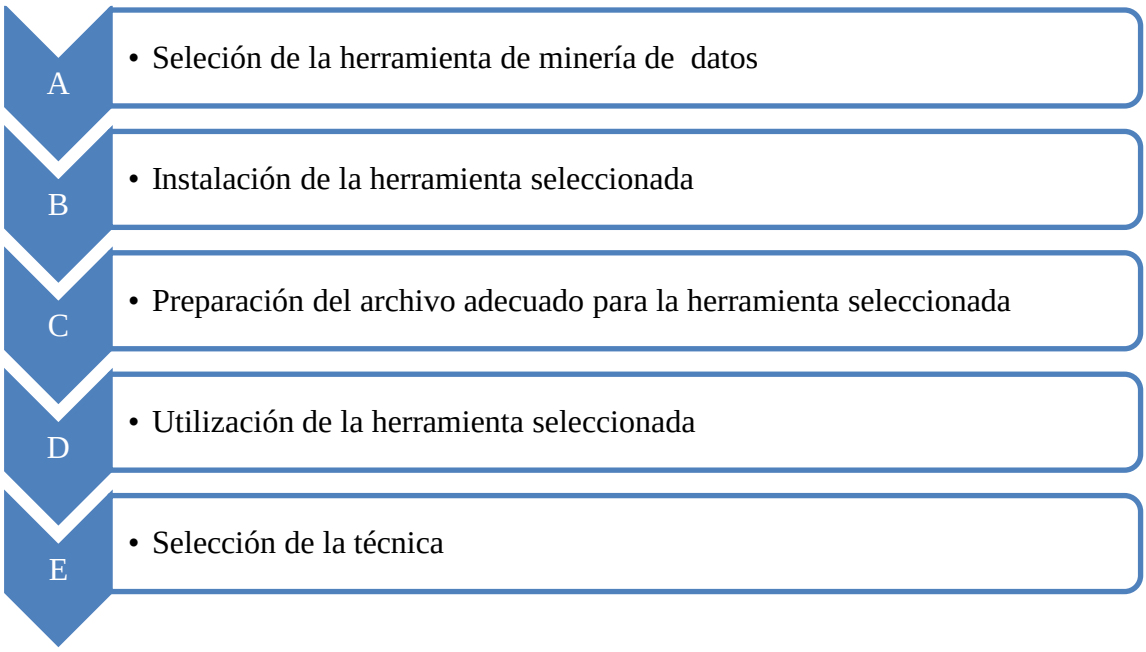


Figura III 10: Procedimiento de selección de herramienta y técnica
Fuente: Autoras

A continuación se describe cada uno de los pasos del procedimiento de selección de herramienta y técnica de minería de datos.

A: Selección de la herramienta de minería de datos.

En base a la revisión de la literatura expuesta en el marco teórico, se ha elegido la herramienta WEKA cuyas características que sustenta esta elección se los expone a continuación.

Entorno Waikato para el análisis del conocimiento (WEKA)

"Weka escrito en Java, es una conocida suite de software para máquinas de aprendizaje que soporta varias tareas típicas de minería de datos, especialmente pre-procesamiento de datos, agrupamiento, clasificación, regresión, visualización y características de selección. Sus técnicas se basan en la hipótesis de que los datos están disponibles en un

único archivo plano o relación, donde cada punto marcado es etiquetado por un número fijo de atributos. WEKA proporciona acceso a bases de datos SQL utilizando conectividad de bases de datos Java y puede procesar el resultado devuelto como una consulta de base de datos. Su interfaz de usuario principal es el Explorer, pero la misma funcionalidad puede ser accedida desde la línea de comandos o a través de la interfaz de flujo de conocimientos basada en componentes". [45]

Características de Weka

- "Permite el acceso a otras instancias de bases de datos por medio de SQL, gracias al JDBC, además puede procesar un resultado generado a base de una consulta hecha a una base de datos". [46]
- Está disponible libremente bajo la licencia pública general de GNU.
- Es muy portable porque está completamente implementado en Java y puede correr en casi cualquier plataforma.
- Contiene una extensa colección de técnicas para pre procesamiento de datos y modelado.
- "Es fácil de utilizar por un principiante gracias a su interfaz gráfica de usuario". [47]

"A continuación se muestra la Tabla III IV con la comparación de la herramienta weka en relación a dos herramientas de licencia propietaria y una de licencia libre". [1]

Tabla III IV: Comparación de herramientas Minería de Datos

CARACTERÍSTICAS	MODELER	SAS ENTERPRISE MINER	TARIYKDD	WEKA
Licencia Libre	No	No	Si	Si
Requiere conocimientos avanzados	No	No	No	No
Acceso a SQL	Si	No	Si	Si
Multiplataforma	No	Si	Si	SI
Requiere base de datos Especializadas	NO	---	No	No
Métodos bayesianos	Si	---	No	Si
Puede combinar modelos	Si	Si	No	Si(no resulta muy eficiente)
Modelos de clasificación	Si	Si	Si	Si
Implementa arboles de decisión	Si	Si	Si	Si
Modelos de regresión	Si	Si	No	Si
Clustering y agrupamiento	Si	Si	No	Si
Interfaz amigable	Si	Si	Si	Si
Permite visualización de datos	Si	Si	Si	Si

Fuente: Diego Vallejo y Germán Tenelanda

Como se puede ver en la Tabla III IV, Weka es una solución completa que reúne potentes características para la explotación de datos, muy similar a otras herramientas propietarias, con la ventaja que Weka es una herramienta gratuita.

Una vez que se ha seleccionado la herramienta de minería de datos que en este trabajo es Weka, se procede con el siguiente paso, instalación de la herramienta.

B: Instalación de la herramienta seleccionada WEKA

Para identificar la técnica adecuada para la detección de intrusos es necesario que Weka este correctamente instalada, para de esta manera elegir la técnica la misma que es parte de la herramienta seleccionada.

El sistema Weka que permite aplicar algoritmos de minería de datos a un conjunto de base de datos y para instalarlo se puede encontrar en el siguiente link http:/www.cs.waikato.ac.nz/ml/weka/. Desde aquí permitirá la instalación de la máquina virtual java y a la vez la instalación de la herramienta.

Una vez iniciado Weka aparecerá una imagen como la siguiente que permite seleccionar la interfaz con la que se va a trabajar. Las posibles interfaces que se puede seleccionar son Explorer, Experimenter, KnowledgeFlow y Simple CLI en nuestro caso seleccionaremos Explorer como se puede observar en la Figura III 11.

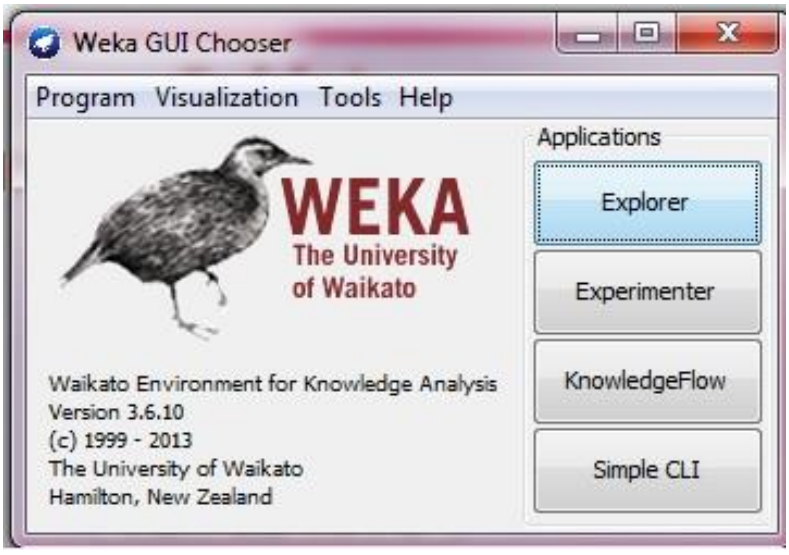


Figura III 11: Primera pantalla del weka
Fuente: Autoras

C: Preparación del archivo adecuado para la herramienta seleccionada WEKA

Weka trabaja con archivos en formato arff, acrónimo de Attribute-Relation File Format. El formato del archivo se explica a continuación.

Cabecera: "En este campo es donde se va a definir el nombre de la relación". [48]

"@relation <nombre-de-la-relacion>" [48]

Donde <nombre-de-la-relacion> es de tipo string, en el caso que se coloque espacios se debe colocar entre comillas.

Declaración de atributos: En esta sección se declaran los atributos que compondrán el archivo de los datos

@attribute <nombre-del-atributo> <tipo>

Donde <nombre-del-atributo> es de tipo string al igual que en nombre de la relación existe la misma restricción anterior los atributos. Los tipos de datos que son aceptados son los siguientes:

- "Numeric: expresa números reales.
- Integer: expresa números enteros.
- Date: expresa fechas, debe ir etiquetado entre comillas y separa con pos guiones o espacios y unidad de tiempo: dd dia, MM mes, yyyy Año, HH horas, mm minutos, ss segundos.
- String: expresa cadena de texto, con las restricciones del tipo string mencionadas anteriormente.
- Enumerado: expresa posibles valores que debe estar entre llaves y separado por comas que puede tomar el atributo

@attribute tiempo { soleado,lluvioso,nublado}”. [48]

Sección de datos: se declaran los datos que integran la relación separados por comas los atributos y con salto de línea las relaciones.

@data

1, 2,3

Para la elección del algoritmo se va a utilizar los datos de entrenamiento de KDD Proporciona el siguiente conjunto de datos:

kddcup.data.gz: Datos de entrenamiento originales (743 MB descomprimido).

kddcup.data_10_percent.gz: Subconjunto del 10% de los datos de entrenamiento (75 MB descomprimido).

kddcup.testdata.unlabeled.gz: Datos de test originales sin etiquetar (430 MB descomprimido).

kddcup.testdata.unlabeled_10_percent.gz: Subconjunto del 10% de los datos de test sin etiquetar (45 MB descomprimido).

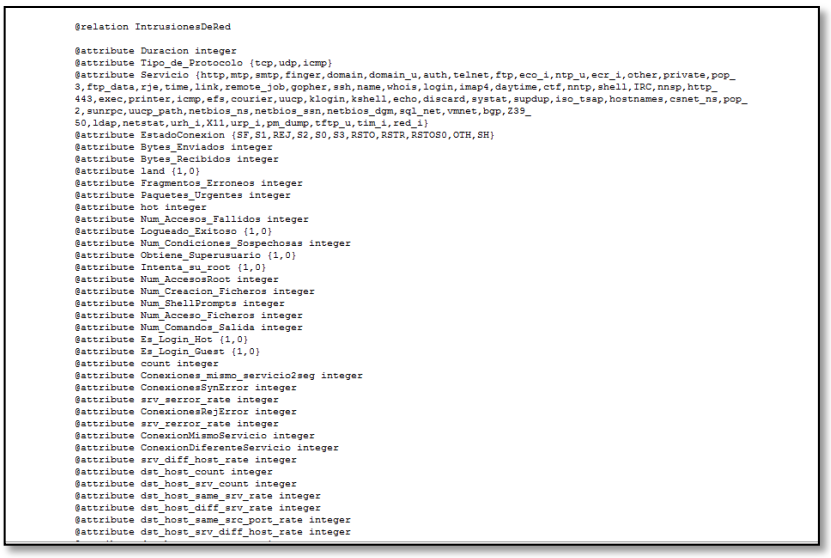
corrected.gz: Subconjunto anterior con los datos correctamente etiquetados (ataques correspondientes especificados) (48 MB descomprimido).

La herramienta weka tiene la capacidad de manejar gran cantidad de datos pero como para su funcionamiento necesita de la ejecución de máquina de java se dispone de una memoria limitada. Tomando en cuenta estas características es conveniente contar con un archivo lo suficientemente grande que sea representativo para el análisis pero que a la vez sea pequeño para que Weka pueda manejarlo sin ningún problema.

De esta manera se tomará en cuenta el archivo del 10% de datos etiquetados y se dispone de un conjunto de 494021 instancias de datos de entrenamiento, que sigue siendo muy grande.

Se probó con diferentes tamaños de conjunto de datos quedando un conjunto de 62666, la elección de estos conjuntos se han hecho muestreando una de cada 8 de las instancias para contar con una muestra homogénea.

El conjunto de datos seleccionados se ha transformado en formato .arff y diseñado la cabecera como se puede ver en la Figura III 12.



```
@relation IntrusionesDeRed

@attribute Duracion integer
@attribute Tipo_de_Protocolo {tcp,udp,icmp}
@attribute Servicio {http,mtp,smtp,finger,domain,domain_u,auth,telnet,ftp,eco_i,ntp_u,ecr_i,other,private,pop_
3,ftp_data,rje,time,link,remote_job,gopher,ssh,name,whois,login,imap4,daytime,ctf,nntp,shell,IRC,nnsp,http_
443,exec,printer,icmp,efs,courier,uucp,klogin,kshell,echo,discard,systat,supdup,iso_tsap,hostnames,csnet_ns,pop_
2,sunrpc,uucp_path,netbios_ns,netbios_ssn,netbios_dgm,sql_net,vmnet,bgp,Z39_
50,ldap,netstat,urh_i,X11,urp_i,pm_dump,tftp_u,tim_i,red_i}
@attribute EstadoConexion {SF,S1,REJ,S2,S0,S3,RSTO,RSTR,RSTOS0,OTH,SH}
@attribute Bytes_Enviados integer
@attribute Bytes_Recibidos integer
@attribute land {1,0}
@attribute Fragmentos_Erroneos integer
@attribute Paquetes_Urgentes integer
@attribute hot integer
@attribute Num_Accesos_Fallidos integer
@attribute Logueado_Exitoso {1,0}
@attribute Num_Condiciones_Sospechosas integer
@attribute Obtiene_Superusuario {1,0}
@attribute Intenta_su_root {1,0}
@attribute Num_AccesosRoot integer
@attribute Num_Creacion_Ficheros integer
@attribute Num_ShellPrompts integer
@attribute Num_Acceso_Ficheros integer
@attribute Num_Comandos_Salida integer
@attribute Es_Login_Hot {1,0}
@attribute Es_Login_Guest {1,0}
@attribute count integer
@attribute Conexiones_mismo_servicio2seg integer
@attribute ConexionesSynError integer
@attribute srv_serror_rate integer
@attribute ConexionesRejError integer
@attribute srv_rerror_rate integer
@attribute ConexionMismoServicio integer
@attribute ConexionDiferenteServicio integer
@attribute srv_diff_host_rate integer
@attribute dst_host_count integer
@attribute dst_host_srv_count integer
@attribute dst_host_same_srv_rate integer
@attribute dst_host_diff_srv_rate integer
@attribute dst_host_same_src_port_rate integer
@attribute dst_host_srv_diff_host_rate integer
```

Figura III 12: Archivo .arff
Fuente: Autoras

D: Utilización de la herramienta seleccionada WEKA

Una vez que se tiene el archivo preparado se procede a la utilización de la herramienta, en donde se aplicarán las diferentes técnicas junto con sus algoritmos propios de Weka. Proceso que a continuación se explica

Para aplicar los diferentes algoritmos de la herramienta se carga el archivo a través de la opción Open File, seleccionamos el archivo y a continuación se visualizará los datos como se puede observar en la Figura III 13.

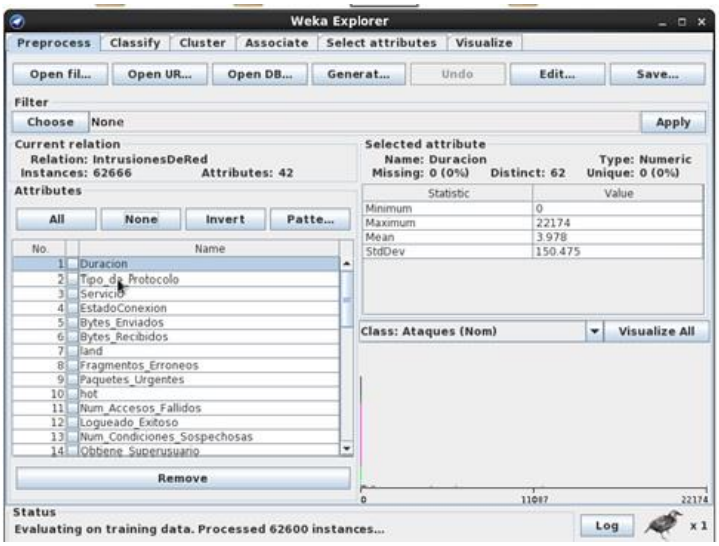


Figura III 13: Weka Explore
Fuente: Autoras

A continuación se realizará la aplicación de los algoritmos más utilizados en Weka orientado a la detección de intrusos, se debe dar clic en la pestaña Clasifity, luego damos clic en el botón Choose que permite seleccionar el algoritmo que se desea aplicar, para visualizar los resultados se da clic en Start como se puede ver en la Figura III 14.

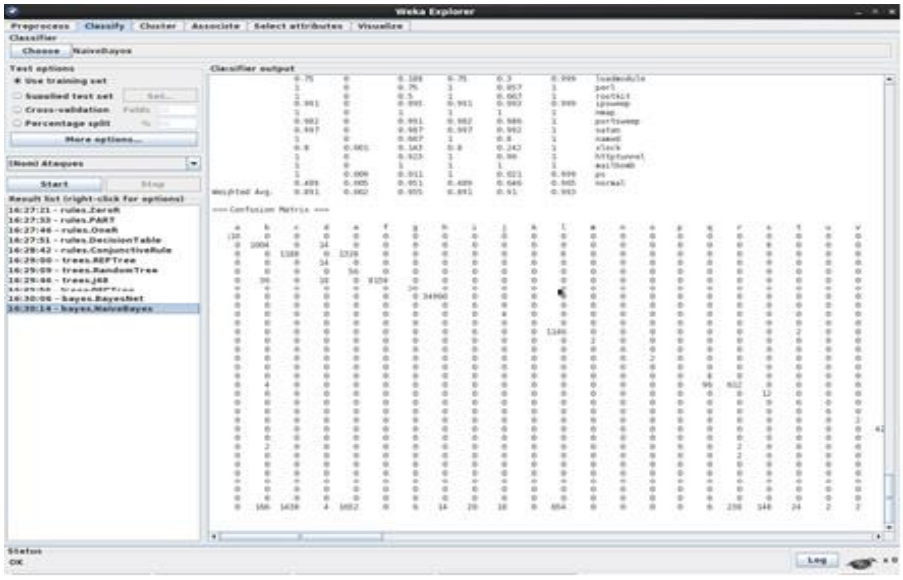


Figura III 14: Resultados de la aplicación de algoritmos
Fuente: Autoras

E: Selección de la técnica

De acuerdo al análisis de los resultados de la aplicación de los algoritmos, se procede a la selección de la técnica adecuada para la detección de intrusos.

Los resultados de cada uno de los algoritmos seleccionados se resumen en la Tabla III V, la cual contiene los siguientes datos: precisión, instancias correctamente clasificadas, instancias incorrectamente clasificadas, tiempo de construcción, factor de estadística Kappa y error absoluto.

Tabla III V: Comparación de los resultados de los respectivos algoritmos

Algoritmos	Precisión	Instancias correcta mente clasificadas	Instancias incorrecta mente clasificadas	Tiempo de construcción	Kappa	Error absoluto
ZeroR	56.17	35198	27468	0.04 s	0	0.04
Ridor	96.73	60620	2046	834.09 s	0.9489	0.02
PART	96.90	60724	1942	5.1 s	0.9515	0.0023
Decision Table	95.80	60040	2626	30.74 s	0.934	0.0058
ConjuntiveRule	71.72	44948	17718	1.53 s	0.5413	0.0209
DecisionStump	71.65	44904	17762	0.52 s	0.5398	0.021
J48	96.87	60706	1960	2.45 s	0.951	0.0024
RandomForest	96.89	60722	1944	3.63 s	0.9514	0.0023
REPTree	96.77	60648	2018	1.57 s	0.9496	0.0024
BayesNet	95.18	59650	3016	1.57 s	0.9253	0.003
Naive Bayes	89.14	55864	6802	0.3 s	0.8337	0.0068

Fuente: Autoras

De acuerdo a la revisión bibliográfica expuesta en el marco teórico se ha elegido la técnica de clasificación y, en base a los resultados obtenidos de las pruebas de los algoritmos de WEKA (Imagen III 14) y que se visualizan en la Tabla III V los algoritmos Part y Random Forest tienen mayor precisión en instancias correctamente clasificadas con el menor error absoluto; uno de los algoritmos que en menor tiempo se construye es ZeroR pero la precisión de instancias correctamente clasificadas es baja. Por lo tanto se llega a la conclusión de que los algoritmos a utilizar en este trabajo son: Part y Random Forest.

3.3 Desarrollo de la propuesta del modelo

El modelo que se propone en este trabajo se basa en los estudios de Lui Wenjun [44] y este a su vez hace referencia al modelo basal de IDS que utiliza minería de datos.

Se hace referencia al estudio de Wenjun porque utiliza las técnicas de minería de datos para crear reglas para la detección de intrusos, el autor define que para optimizar el rendimiento en minería de datos se lo realiza sobre los datos que se consideren

sospechosos ya que según DARPA el 95 % del tráfico es normal. También se define que la solución que plantea en este trabajo integraría la ejecución de minería de datos en tiempo real; sin embargo, al respecto Wenjun opina que esta integración debe mejorar, por lo que expone que: "*en el fututo será importante combinar la tecnología de minería de datos, tecnología de análisis de protocolos y tecnologías de análisis de la relatividad con el fin de disminuir la tasa de distorsión y mejorar la eficiencia de detección de intrusos desconocidos* ".

Los datos sospechosos son guardados en una base de datos y luego son utilizados para aplicar minería de datos permitiendo disminuir el tiempo del procesamiento de datos.

3.3.1 Marco Conceptual

La tecnología de minería de datos es un proceso no elemental, donde hace búsquedas de relaciones, correlaciones, dependencias, asociaciones, modelos, estructuras, tendencias, clases, segmentos, los mismos que se obtienen de grandes cantidades de datos y tiene gran impacto en la detección de intrusos.

"La compleja situación en los sistemas y redes informáticas ha traído consigo el desarrollo de investigaciones con el propósito de crear mecanismos de seguridad como la prevención". [49]

"Uno de los principales exponentes dentro de las herramientas de detección son los Sistemas de Detección de Intrusos (IDS –Intrusion Detection System). Estos sistemas son los encargados de identificar y responder a las actividades maliciosas que tienen como objetivo una computadora o los recursos de red". [49]

"Esta herramienta es relativamente joven pero desde su surgimiento han aparecido una enorme variedad de propuestas que intentan dar solución a este enfoque, tan complicado como rico en posibilidades". [49]

3.3.2 Modelo Propuesto

El modelo que se propone con este trabajo de tesis se lo esquematiza en la Figura III 15.

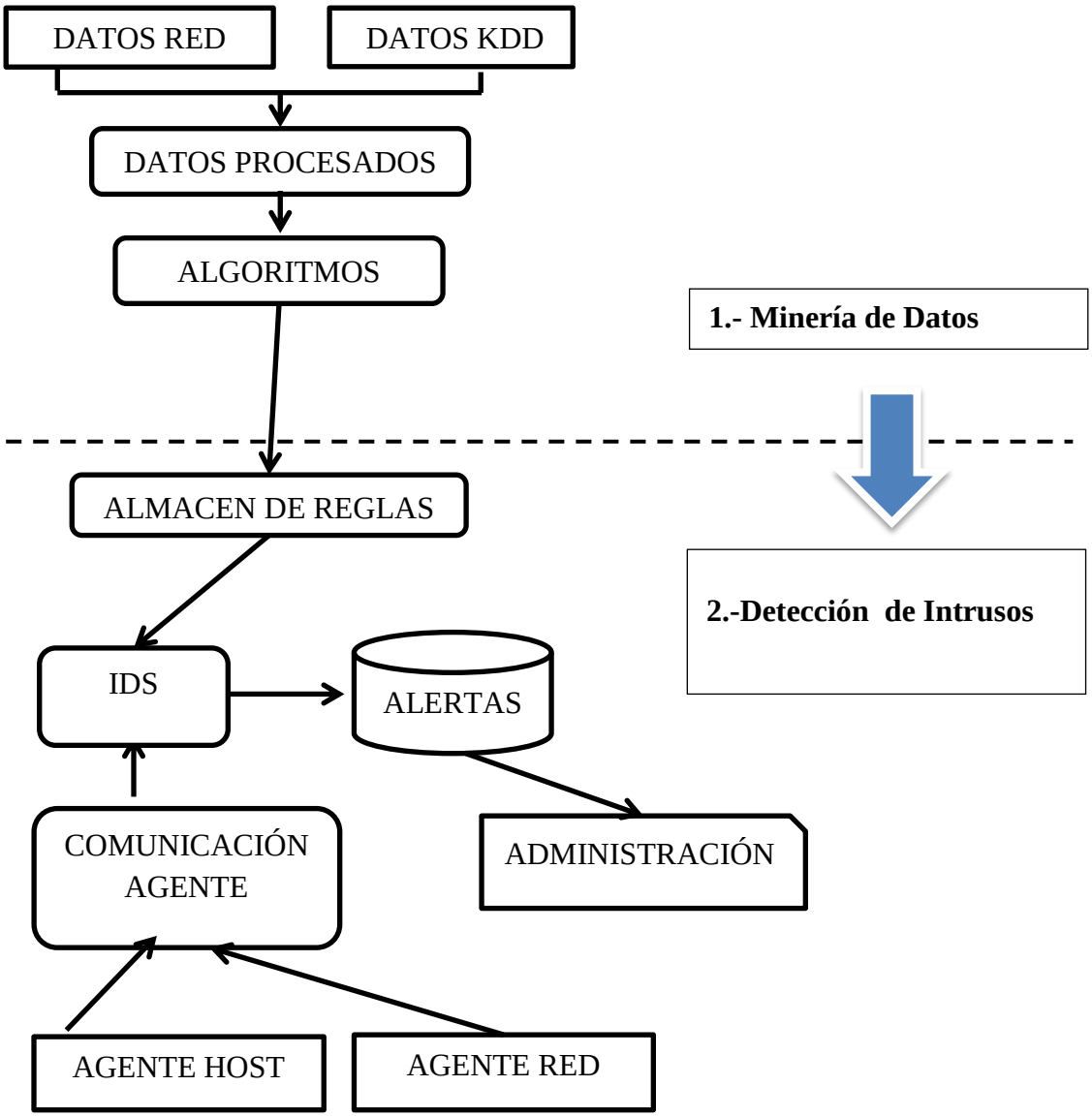


Figura III 15: Modelo Obtenido
Fuente: Autoras

El modelo propuesto esta subdividido en dos componentes que son: Minería de datos y la detección de intrusos los mismos que se detallan a continuación.

COMPONENTE 1: MINERÍA DE DATOS

Este componente está conformado por 4 pasos como se ilustra en la Figura III 15, los mismos que son descritos a continuación. Para una mejor explicación se hace referencia a la Figura III 16.

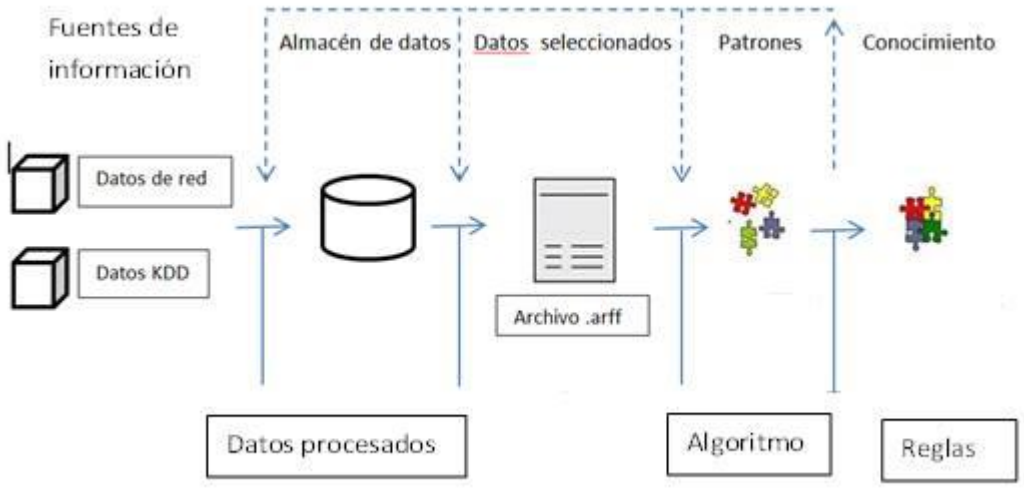


Figura III 16: Componente minería de datos
Fuente: Autoras

A continuación se describe cada uno de los pasos.

Datos de red: Son los datos capturados de la red de información donde se va a implementar el modelo, tomando en cuenta que sean los suficientemente representativos para la correcta aplicación de la tecnología de minería de datos.

Datos KDD: Son un conjunto de datos del concurso KDD de 1999, que tenían como objetivo evaluar el estudio y la investigación en la detección de intrusos.

Datos procesados:

Este a su vez contiene los siguientes pasos:

1.- Entender el entorno de la aplicación, obtener el conocimiento más importante, y los objetivos del usuario final.

2.- Integración y recopilación de datos se analiza las fuentes de los datos la selección de los conjuntos y subconjuntos de muestra. En el modelo, los datos que se van a utilizar son una pequeña selección del conjunto de datos del concurso KDD de 1999, en donde se hizo uso de una pequeña versión de la gran variedad de intrusiones militares simuladas en un entorno de red, que tenían como objetivo evaluar el estudio y la investigación en la detección de intrusiones, conjuntamente con los datos capturados de la red de información donde se vaya a implementar el modelo.

3.- Selección limpieza y transformación se realiza la recolección de los datos que van a ser analizados, la depuración de acuerdo al objetivo del proceso, análisis de los tipos de datos y las variables que intervienen para la creación del archivo en formato .arff.

Algoritmos

En este paso se procede aplicar la tecnología de minería de datos, la misma que se define como "la extracción no trivial de información implícita, desconocida y potencialmente útil de los datos" [50]. Se aplican los algoritmos de clasificación Part y Random Forest para obtener los patrones para la creación de nuevas reglas.

Si los patrones encontrados no cumplen con los objetivos hará falta retroceder a uno de los pasaos anteriores.

COMPONENTE 2: DETECCIÓN DE INTRUSOS.

En este componente se integran las reglas generadas por el componente de minería de datos que serán guardadas en la base de datos del sistema detector de intrusos.

En el modelo propuesto, incluye dos agentes el uno que es de host y el otro que es el agente de red. Y por otro lado está la comunicación de agentes, IDS, reglas y el módulo de administración como se puede ver en la Figura III 17.

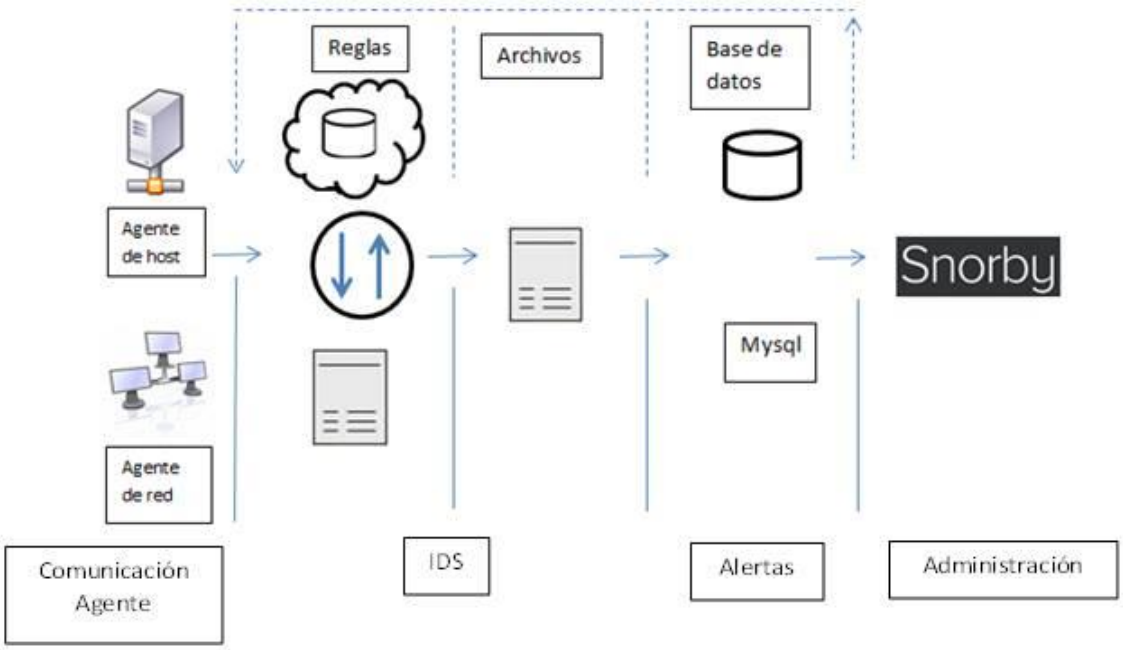


Figura III 17: Componente detección de intrusos
Fuente: Autoras

A continuación se detalla cada uno de los módulos utilizados en el componente de detección de intrusos.

Comunicación de los agentes: Es el encargado de la comunicación entre el agente host, el agente de red hacia el motor IDS para generar las respectivas alertas.

El modelo propuesto basado en la minería de datos va a constar de dos agentes los mismos que son:

- **Agente Host:** Sagan HIDS (Sistema de Detección de host Instrucción). Este agente es el encargado de detectar las intrusiones que se están llevando a cabo hacia el ordenador donde se encuentra configurado el motor de detección de intrusos IDS. Casi nunca se tiene que instalar el agente de host.
- **El agente de red:** en este caso Suricata es el encargado de obtener los paquetes de datos de todos los nodos de red, este agente recoge la información que pasa por la red compara con las reglas propias del Suricata más las que se agregan dependiendo de la tarea que se desee realizar.

Para la instalación del agente de red que en este modelo se hace uso de Suricata se requiere tener instalado el siguiente software.

1. Suricata, el motor IDS.
2. Apache2, el servidor web.
3. MySQL, el servidor de base de datos.
4. Barnyard2, el analizador que estudia el formato unified2 de Suricata y escribirlos en la base de datos MySQL.
5. Snorby, el frontend interfaz web para la gestión de alertas de IDS.
6. Ruby, al menos la versión 1.9.2 que se necesita para apoyar Snorby. [51]

"Los Sistemas de Prevención y Detención de Intrusos (IDS) facilitan un nivel agregado de seguridad en las redes de datos notificando vulnerabilidades que los firewall simplemente no pueden realizar". [41]

Para el análisis de paquetes el IDS Suricata funciona de la siguiente manera, como se describe en la Figura III 18.

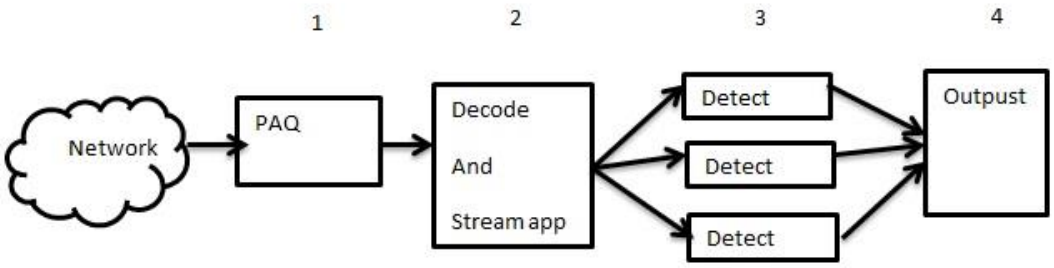


Figura III 18: Etapas del trabajo del IDS Suricata
Fuente: Autoras

Como se puede observar en la Figura III 18 se indica el procesamiento de un paquete y los módulos de Suricata que intervienen.

PAQ: Este es el encargado de la adquisición de paquetes.

Decode: Este se encarga de la decodificación de los paquetes.

Stream app. Layer "Realiza el seguimiento del flujo de reconstrucción". [41]

Detect: Se encarga de la comparación de firmas.

Output: Procesa todos los eventos y alertas [41].

Suricata es multihilos donde un hilo es como un proceso que se ejecuta en un ordenador, por lo que varios subprocesos están activos al mismo tiempo.

Un hilo-módulo es una parte de una funcionalidad. Un módulo es, por ejemplo, para la decodificación de un paquete, otro es el módulo detector y otro la salida de alertas. Un paquete puede ser procesado por más de un hilo. El paquete se pasará al siguiente hilo a través de una cola. Los paquetes serán procesados por un hilo a la vez, pero puede haber varios paquetes que se están procesando en un momento por el motor.

Un hilo puede tener uno o más módulos de rosca. Si tienen más módulos, sólo pueden estar activos en un tiempo.

“Detección de Protocolos Automático: El motor de suricata tiene palabras claves para algunos protocolos como: IP, TCP, UDP, ICMP, HTTP, TLS, FTP y SMB. Esto quiere decir que se puede detectar una ocurrencia dentro de un stream de datos, sin importar el puerto en donde ocurre. Esta característica es importante para el control y detección de malware”. [41]

En este modelo también se hace uso del motor de base de datos Mysql.

“Mysql es un sistema gestor de bases de datos muy conocido y ampliamente usado por su simplicidad y notable rendimiento”. [52]

Para realizar los reportes Suricata escribe eventos en formato unified2, por lo que los usuarios necesitaran Barnyard2 para luego poder analizar los acontecimientos de Unified2 dentro de una base de datos en este caso Mysql.

El IDS Suricata una vez que va analizando el tráfico que atraviesa por él, va guardando las alertas en la base de datos Mysql, el usuario puede configurar esta opción de acuerdo a sus necesidades, por lo general los días que se van a guardar las alertas en la base de datos son 3, pasado este tiempo las alertas se eliminaran automáticamente.

IDS: Es el encargado de escuchar el tráfico en la red, para detectar las actividades sospechosas, y de esta manera generar las alertas para el módulo de administración.

Almacén de Reglas: Se trata de la obtención de conocimiento para la creación de nuevas reglas, las cuales serán guardadas en la base de datos del IDS y se añadirán al funcionamiento del sistema, las que ayudara a cumplir con el objetivo.

Alertas: Estas son generadas por el IDS, luego de la comparación de las reglas con el flujo de datos de la red, las mismas que se puede visualizar a través de Snorby el mismo que es un front-end para la gestión de alertas generadas por los sensores como Suricata, Snort entre otros.

Administración: Este módulo es el encargado de informar al administrador de las respectivas alertas generadas en el IDS, y es él quien decide las acciones adicionales a fin de disminuir el retardo del sistema, aumentar la rapidez de reacción y evitar la devastación por BROADCASTIN eficientemente.

CAPÍTULO IV

APLICACIÓN DEL MODELO PROPUESTO

4.1 Introducción

En el siguiente capítulo se implantará el modelo propuesto para la detección de intrusos, el mismo que será aplicado en la red de datos de la FIE como en la prototipo creado para la pruebas. Esperando de esta manera que la FIE pueda mejorar la seguridad de la red de datos.

4.2 Escenario requerido para la implementación del modelo propuesto

El escenario mínimo que se requiere para la implementación del modelo propuesto se describe en la Figura IV 19.

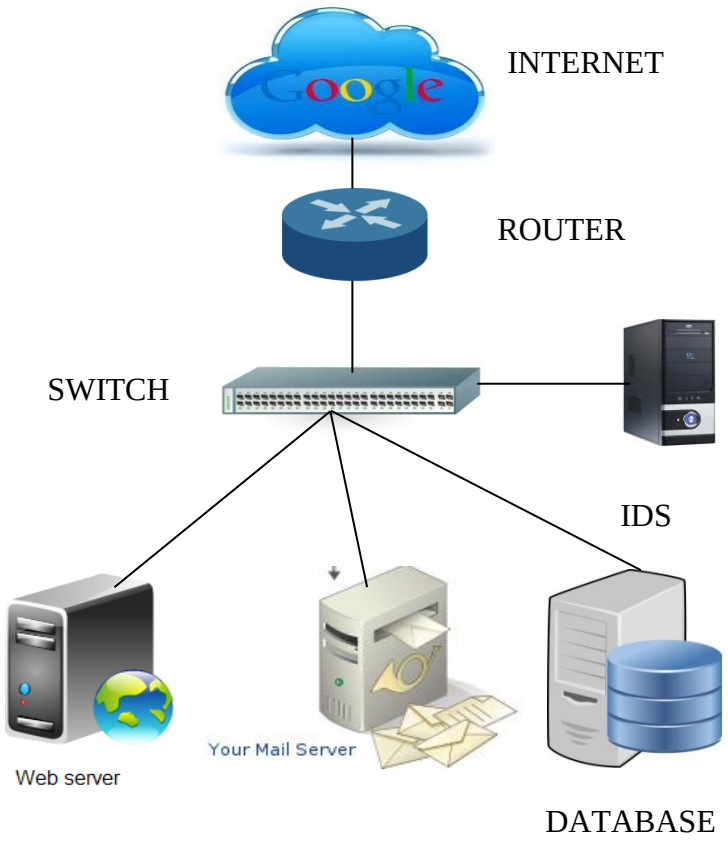


Figura IV 19: Escenario requerido para la aplicación del modelo
Fuente: Autoras

Para la implementación del modelo en la FIE fue necesario contar con un switch en el cual fue configurado un puerto spam y un puerto de administración para de esta manera poder monitorear todo el tráfico de la red de datos del edificio de la facultad.

4.3 Procedimiento para aplicar el modelo propuesto

El procedimiento que se realiza para la implementación del modelo se lo esquematiza en la Figura IV 20.

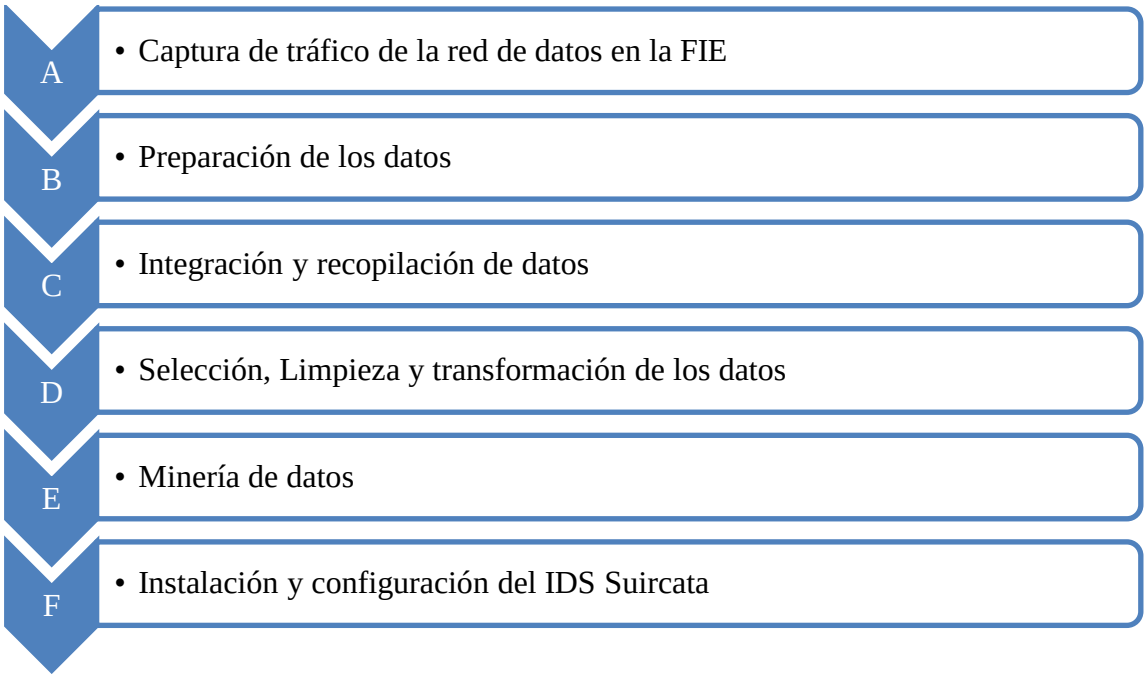


Figura IV 20: Procedimiento de la aplicación del modelo
Fuente: Autoras

A continuación se describe cada uno de los pasos a seguir en la implementación del modelo propuesto

A. Captura del tráfico

Para la captura de tráfico en la red de información de la FIE se utilizó de la herramienta TCPDUMP por lo cual a continuación se hará una breve descripción de su instalación.

La herramienta tcpdump permite realizar un análisis del tráfico que circula a través de nuestra red. Esta herramienta nos permitió capturar en tiempo real los paquetes que circulaban por la red de la FIE.

Tcpdump está disponible para el sistema operativo UNIX aunque también hay una adaptación para Windows llamado WinDump. Para nuestro caso se lo instaló en Centos 6.4 para lo cual se hizo uso de la siguiente línea de comando.

- Yum install tcpdump.

Una vez instalada la herramienta se procedió a la captura de tráfico en la FIE lo cual tuvo una duración de 4 semanas, los datos obtenidos nos facilitó el ingeniero encargado del departamento de redes de la ESPOCH.

B. Preparación de los datos

Para lograr este objetivo se hará uso de las fases del KDD el mismo que consiste en usar métodos de minería de datos (algoritmos) para extraer (identificar) el conocimiento de acuerdo a las especificaciones de ciertos parámetros haciendo uso de una base de datos junto con pre-procesamientos y pos-procesamientos. Para un mayor entendimiento a continuación se ilustra en la Figura IV 21.

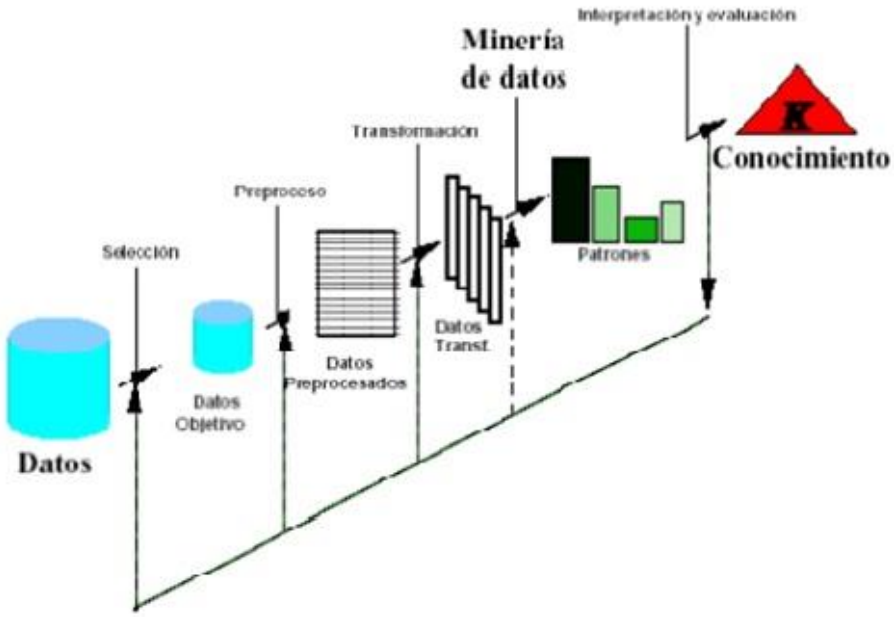


Figura IV 21: Proceso de KDD **[53]**
Fuente: Vallejo

El formato de los datos que se obtuvo dentro de la fuente no fue el idóneo por lo que no fue posible aplicar casi ningún algoritmo de minería de datos sobre los mismos mediante el pre procesado, se realiza un filtrado de los datos de modo que se borren los datos desconocidos, datos considerados como inválidos e incorrectos de acuerdo al objetivo de análisis, de lo cual se obtuvieron muestras en busca de reducir el tamaño de la información y una mayor velocidad de respuesta durante el proceso.

Para este trabajo se tomó los datos de dos fuentes como son la información capturada de la red de información de la FIE, la misma que se realizó con la herramienta TCPDUMP durante un tiempo de 4 semanas.

Al igual que la información de la base de datos de los militares, se debe tener en cuenta que la herramienta WEKA tiene que manejar grandes conjuntos de datos y que al usar la máquina virtual de java se dispone de una memoria limitada. Por este motivo hay que tomar un conjunto de datos que sea lo bastante grande como para poder hacer un análisis que se considere lo suficientemente bueno, pero que a la vez sea lo necesariamente pequeño para que WEKA pueda manejarlo con cierta soltura sin llegar a colapsar el equipo.

Luego de haber sido pre procesados los datos se sigue teniendo una cantidad enorme de datos. Por lo cual se procedió a buscar características que permitan reducir el tamaño de los mismos, eligiendo las variables que se encontraba con mayor frecuencia.

C. Integración y Recopilación de datos

Una vez obtenida ya la captura de tráfico de la red de información de la FIE se procede a sacar los datos que sirven e intervienen en la investigación junto con los datos de la

base de datos de los militares, obteniendo así un archivo de texto con un total de 31333 registros.

“Este conjunto de datos se tienen que pasar a formato WEKA. Los datos de entrada sobre los que operarán los algoritmos, deben estar codificados en un formato denominado, attribute-Relation File Format (extension "arff")”. [54]

El formato de un fichero arff sigue la estructura siguiente vea Tabla IV VI.

Tabla IV VI: Estructura del fichero arff

```
@relation NOMBRE_RELACION
@attribute at1 tipo
@attribute at2 {valor1, valor2, …}
…
@attribute atN {valor1, valor2, …}
@data
dato11,dato21...,datoN1
…
dato1M,dato2M...,datoNM
```

Fuente: Autoras

“Los atributos con los que Weka trabaja son de dos tipos: numéricos de tipo real o entero (real o integer), y simbólicos (donde se describe los valores posibles que obtener el atributo)”. [54]

En este caso, los atributos la mayoría son de tipo string como por ejemplo para el tipo de servicio, los distintos ataques, integer que hace referencia a los bytes recibidos, paquetes urgentes, números de accesos fallidos, números de acceso sospechoso entre otros, así también tenemos booleano para hacer referencia a al login que tuvo éxito, cuando intenta acceder como root.

También se han tenido que obtener los tipos de ataque que aparecen en los datos, para añadírselos a las cabeceras de ambos conjuntos para que sean compatibles entre sí. Los valores que pueden tomar estos atributos son finalmente los siguientes vea la Tabla IV VII.

Tabla IV VII: Posibles valores de los atributos simbólicos.

Atributo	Valores
protocol_type	tcp, udp, icmp
Service	http, mtp, smtp, finger, domain, domain_u, auth, telnet, ftp, eco_i, ntp_u, ecr_i, other, private, pop_3, ftp_data, rje, time, link, remote_job, gopher, ssh, name, whois, login, imap4, daytime, ctf, nntp, shell, IRC, nnsp, http_443, exec, printer, icmp, efs, courier, uucp, klogin, kshell, echo, discard, systat, supdup, iso_tsap, hostnames, csnet_ns, pop_2, sunrpc, uucp_path, netbios_ns, netbios_ssn, netbios_dgm, sql_net, vmnet, bgp, Z39_50, ldap, netstat, urh_i, X11, urp_i, pm_dump, tftp_u, tim_i, red_i
Flag	SF, S1, REJ, S2, S0, S3, RSTO, RSTR, RSTOS0, OTH, SH
Class (ataque)	back, buffer_overflow, ftp_write, guess_passwd, imap, ipsweep, land, loadmodule, multihop, neptune, nmap, normal, perl, phf, pod, portsweep, rootkit, satan, smurf, spy, teardrop, warezclient, warezmaster, snmpgetattack, named, xlock, xsnoop, sendmail, saint, apache2, udpstorm, xterm, mscan, processtable, httptunnel, worm, mailbomb, sqlattack, snmpguess, ps

Fuente: Autoras

D. Selección, Limpieza y Transformación de los datos

El conjunto seleccionado de datos junto con el tráfico capturado en la red de la FIE y los datos de la base de datos de los militares dan un total de 31333 registros, los mismos que intervendrán en la investigación.

En el archivo final cada conexión (registro) está etiquetada como "normal" o como ataque, con un tipo específico de ataque, que se podrán agrupar en:

- DOS: denegación de servicio es un ataque que procura evitar que la víctima pueda utilizar todo o una parte de su conexión de red
- R2L: acceso no autorizado a una máquina remota
- U2R: acceso no autorizado a los privilegios de superusuario.
- Probe: monitorización.

Con este archivo se procede a cargarlo a la herramienta weka con el cual se aplicara los diferentes algoritmos de la minería de datos.

Finalmente el archivo que fue subido a la herramienta quedo como se muestra en la Figura IV 22.

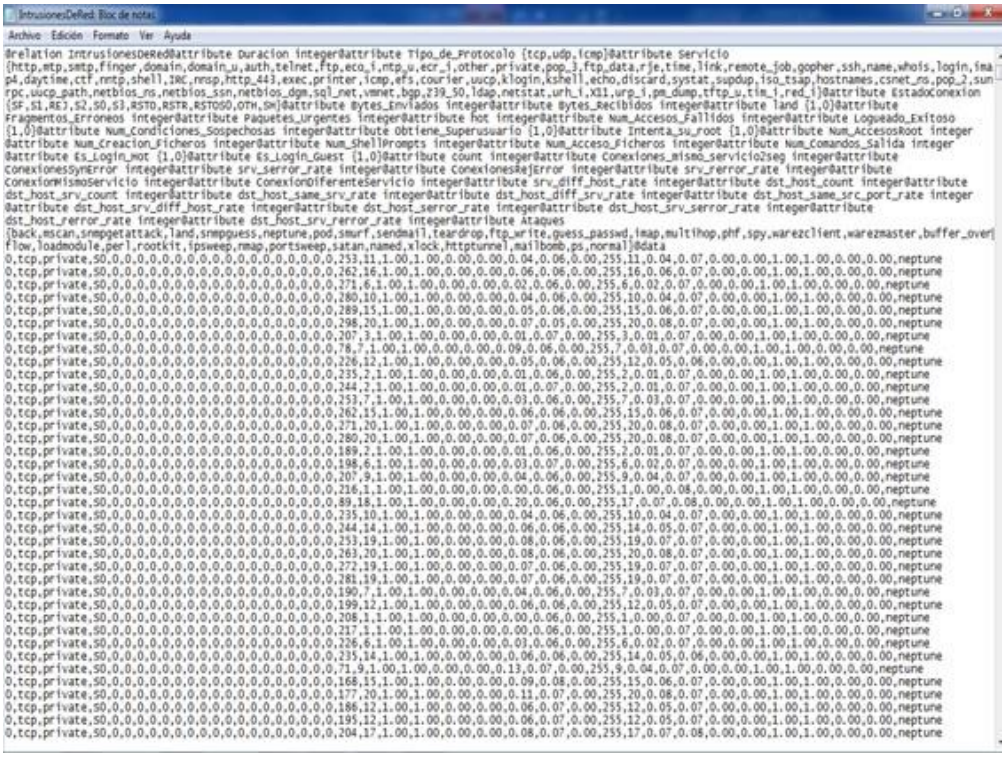


Figura IV 22: Archivo que acepta weka
Fuente: Autoras

E. Minería de datos

En el conjunto de datos que se llegó a obtener se procede a emplear los dos algoritmos seleccionados anteriormente, como son el algoritmo Part como el algoritmo Random forest.

A continuación se muestra los resultados obtenidos aplicando los dos algoritmos en la herramienta.

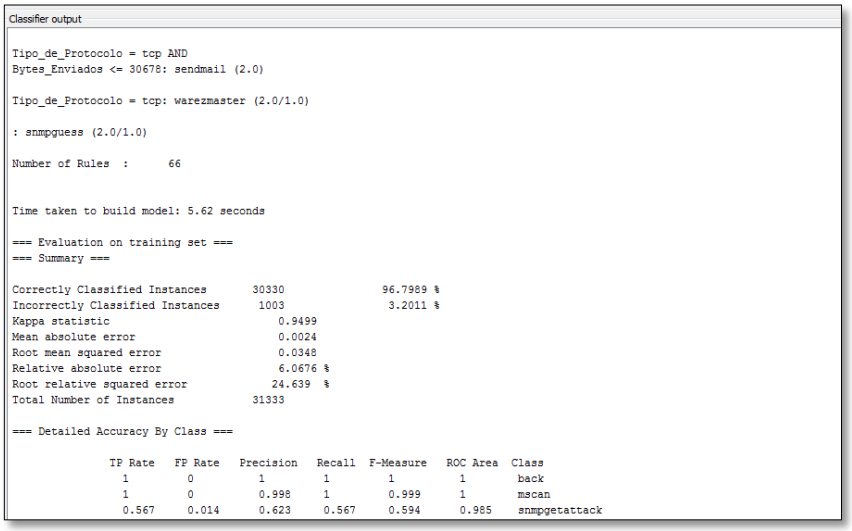
Algoritmos basados en reglas

"Algoritmos que aprenden modelos basados en reglas. Mediante los datos de entrenamiento, aprende una serie de reglas y con ellas predice un resultado u otro". [55]

Dependiendo del algoritmo que se seleccione, se ha conseguido mayor o menor eficacia.

PART

Algoritmo que crea una serie de reglas utilizando los atributos más significativos para cada tipo de ataque y los resultados se muestra en la Figura IV 23.



```
Classifier output

Tipo_de_Protocolo = tcp AND
Bytes_Enviados <= 30678: sendmail (2.0)

Tipo_de_Protocolo = tcp: warezmaster (2.0/1.0)

: snmpguess (2.0/1.0)

Number of Rules  :      66


Time taken to build model: 5.62 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances       30330               96.7989 %
Incorrectly Classified Instances      1003                3.2011 %
Kappa statistic                          0.9499
Mean absolute error                      0.0024
Root mean squared error                  0.0348
Relative absolute error                  6.0676 %
Root relative squared error             24.639  %
Total Number of Instances            31333

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0          1         1         1          1        back
                 1         0          0.998     1         0.999      1        mscan
                 0.567     0.014      0.623     0.567     0.594      0.985    snmpgetattack
```

Figura IV 23: Modelo Part
Fuente: Autoras

Tiempo necesario para construir el modelo: 5,62 segundos

Los casos clasificados correctamente 30330 96.7989 %

Los casos clasificados incorrectamente 1003 3.2011 %

Kappa estadística 0.9499

Error absoluto promedio de 0.0024

Raíz del error cuadrático 0.0348

Error absoluto relativo 6.0676 %

Relativas a la raíz error cuadrado 24.639 %

Número total de instancias 31333

Algoritmos basados en Arboles

“Estos algoritmos simbolizan a los conjuntos de decisiones las mismas que permiten generar reglas para la categorización de los datos”. [41]

Random Forest

“Se basan en el desarrollo de muchos árboles de clasificación. Para clasificar un objeto desde un vector de entrada, se pone dicho vector bajo cada uno de los árboles del bosque. Cada árbol genera una clasificación, el bosque escoge la clasificación teniendo en cuenta el árbol más votado sobre todos los del bosque”. [41] Y los resultados se puede observar en la Figura IV 24.

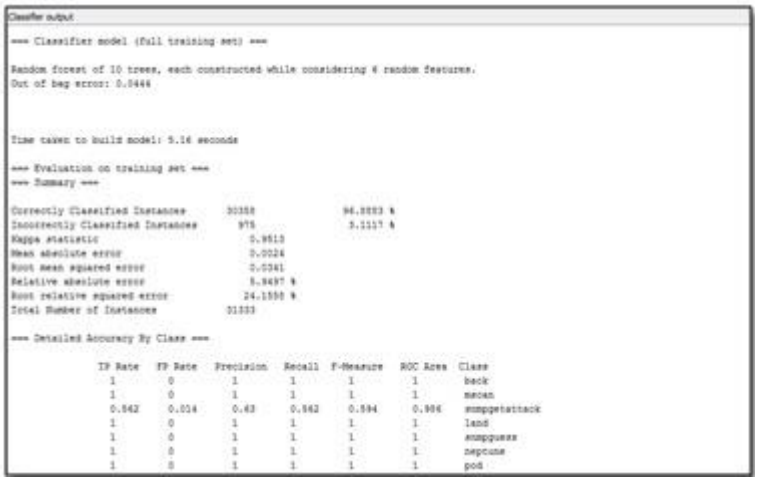


Figura IV 24: Modelo Random Forest:
Fuente: Autoras

Tiempo necesario para construir el modelo: 5,16 segundos

Los casos clasificados correctamente 30358 96.8883%

Las instancias incorrectamente clasificados 975 3.1117%

Kappa estadística 0.9513

Error absoluto promedio de 0.0024

Raíz del error cuadrático 0.0341

Error absoluto relativo 5.9497%

Relativas a la raíz error cuadrado 24,1558%

Número total de instancias 31333

F. Instalación y configuración del IDS

Para este proceso se recomienda realizar los siguientes pasos, como se puede observar en la Figura IV 25.

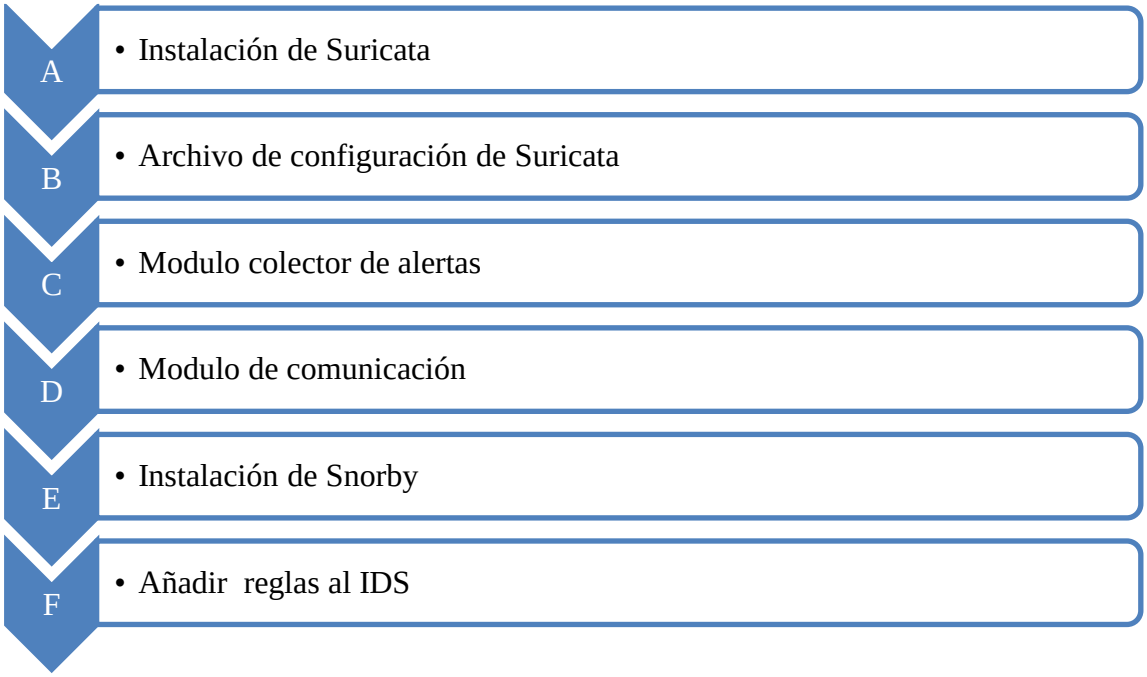


Figura IV 25: Procedimiento de la Instalación y configuración del IDS
Fuente: Autoras

Los pasos utilizados en el procedimiento se describen a continuación.

El motor de detección de intrusos que fue utilizada en la investigación es suricata debido a sus altos beneficios brindados en su desempeño como IDS mismas que son descritas en el capítulo II a continuación se verá la instalación y configuración del motor de detección.

A. Instalación Del Suricata

Para una correcta instalación es necesario verificar los requerimientos tanto de hardware y software.

Estos requerimientos tanto de hardware como de software van a variar dependiendo de la cantidad de tráfico que va a ser monitoreado y de la cantidad de firmas que serán subidas o cargadas. Si la cantidad de tráfico es numerosa se va a necesitar de un buen procesador con mayor cantidad de núcleos al igual que la memoria RAM[4].

Si al suricata se lo va a ejecutar solamente como IDS es necesario una sola tarjeta de red para capturar el tráfico, pero si se lo va a ejecutar como IPS como minino se necesita 2 tarjetas de red la una para escuchar el tráfico de red y la segunda para la administración, además porque el IPS es un dispositivo de capa 2.

Los requerimientos de software, sin importar el sistema operativo sobre la cual se vaya a instalar es necesario tener instaladas lo siguiente como se ilustra en la Figura IV 26.

[4] Un dispositivo de memoria de acceso aleatorio.

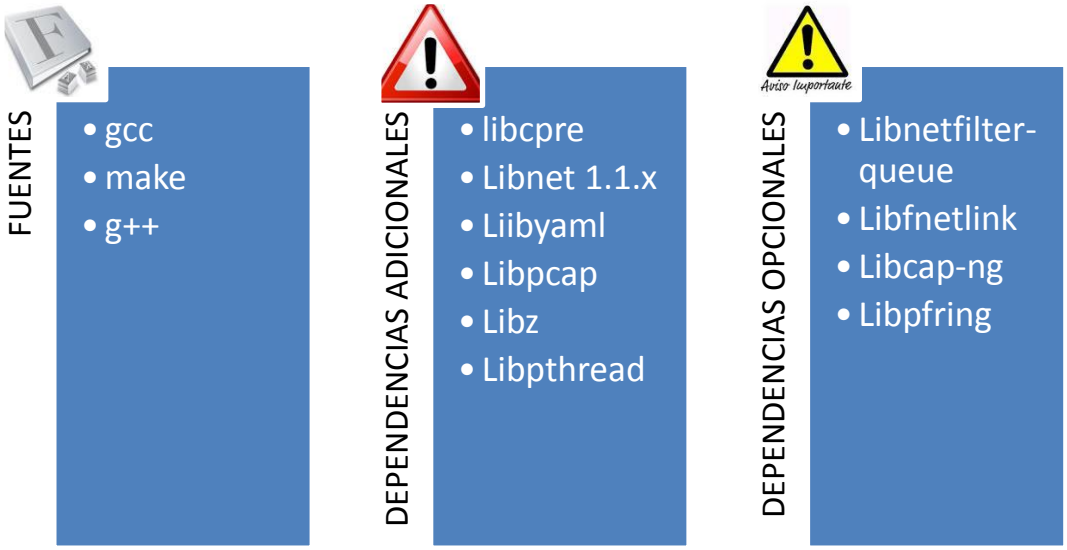


Figura IV 26: Requerimientos para la instalación del Suricata
Fuente: Autoras

La herramienta suricata puede ser instalado en varias plataformas como Linux (Centos, Ubuntu/Debian), Windows y Mac OS para esta investigación se instaló en linux Centos 6.4.

Instalamos todas las dependencias.

```
yum install git
yum install libpcap libnet gcc automake autoconf make
yum install bridge-utils iptables iproute2
yum install libyaml libnfnetlink libnetfilter_queue
#Libcap-ng no se encuentra en repositories de arch por lo que tiene que #ser descargada y compilada manualmente wget http://people.redhat.com/sgrubb/libcap-ng/libcap-ng-0.6.5.tar.gz
```

Descargamos las fuentes de Suricata mediante el comando # wget. http://www.openinfosecfoundation.org/download/suricata-1.0.5.tar.gz dependiendo si su sistema operativo es de 32 o 64 bits.

Compilamos los ficheros fuentes:

```
./autogen.sh
./configure –enable-nfqueue #enable-nfqueue para modo IPS
make
make install
```

La opción enable-nfqueue se lo configurará solo si se desea que Suricata corra en modo IPS caso contrario lo hará en modo IDS.

Por último se procede a descargar y descomprimir las reglas de emerging en el directorio respectivo de Suricata. Como segundo paso se procede a copiar las reglas a los archivos respectivos con los que trabajará el colector de alertas. (barnyard2).

```
cd /etc/suricata/
wget http://www.emergingthreats.net/rules/emerging.rules.tar.gz
tar -xzvf emerging.rules
cp rules/classification.config .
cp rules/reference.config .
cp rules/sid-msg.map .
cp rules/gen-msg.map .
```

B. Archivo de configuración del Suricata

Antes de la ejecución del motor Suricata se debe realizar las respectivas configuraciones de los parámetros.

El archivo de configuración de Suricata utiliza un formato YAML la estructura de este formato se denota en identificar los espacios en blanco, hay que tener cuidado con este archivo ya que las tabulaciones pueden ocasionar problemas, además procure que las opciones del mismo nivel baya bajo la misma lineación.

Las opciones más importantes de configuración del Suricata son:

Max-pending packet

En este se pone el número de paquetes a procesar simultáneamente el número va a depender del hardware que se dispone y aumenta el buffer para las colas NFQ, en este caso se tiene Max-pending-packets: 2000.

Outputs

Es aquel que define los archivos de salida que tendrá el sensor. Para esto es necesario que este habilitada **unified-alert** y deshabilitado el http-log como se puede observar a continuación.

Action order

Aquí se va a definir el orden en las que las reglas serán cargadas, pero serán procesadas en distinto orden es decir las firmas más importantes se procesaran primero.

unified2-alert:	http-log:
enabled: yes	enabled: no

El orden que en este caso es el siguiente:

Pass: Deja pasar el trafico
Drop: Elimina información invalida
Reject: Rechaza posibles ataques
Alert: Da aviso de los ataques

Detect-engine

Es el encargado de la configuración del perfil de optimización de memoria, mientras agrupa y procesa las reglas.

```
detect-engine:
profile: high
```

Logging

Es el encargado de permitir la habilitación de los registros del sensor y quedaría de la siguiente manera.

```
file:
enabled: no
filename: /var/log/suricata.log
```

Rule- files

Aquí se podrá visualizar todas aquellas reglas cargadas en memoria al inicializar el motor Suricata y por lo tanto contra las que se compararán los paquetes en busca de ataques.

```
default-rule-path: /etc/suricata/rules/
 rule-files:
 botcc.rules
 emerging-attack_response.rules
 emerging-dos.rules
 emerging-exploit.rules
 emerging-malware.rules
 emerging-p2p.rules
 emerging-scan.rules
 emerging-voip.rules
 emerging-web_client.rules
 emerging-web_server.rules
```

Registro de alertas basado en líneas

Todos los avisos o alertas son almacenados en un archivo de texto, una alerta posee descripción, el tiempo en que se efectuó y las direcciones ip que intervienen en el evento.

```
fast:
enabled: yes
filename: fast.log
append: yes/no
```

Salida de las alertas para utilizar con Barnyard2

Este archivo es el unified2.alert el mismo que es necesario para enviar los sucesos detectados que sea necesario guardarlas en una base de datos u otras aplicaciones.

```
unified2-alert:
enabled: yes
filename: unified2.alert
limit: 32
```

Registro de peticiones HTTP

Este archivo contiene información sobre el tráfico de la red, como por ejemplo el http request, el nombre del host, la URI y el usuario.

```
http-log:
enabled: yes
filename: http.log
append: yes/no
```

Salida de syslog

Esta es la opción más importante ya que este es el camino por donde se puede enviar las alertas y evetos syslog.

```
syslog:
enabled: no
facility: local5
level: Info
```

Motor de detección

El motor de detección en si está constituido por grupos internos de firmas de seguridad, aquí se cargan las firmas que serán comparadas con el tráfico, algunas de ellas no serán

utilizadas en ciertos caso por lo que las firmas serán divididas en grupos para tener un mejor rendimiento del motor.

La cantidad de grupos de firman determinaran el balance entre la memoria y el rendimiento, como también va a depender del hardware que se disponga es recomendable tener un buen hardware.

Afinidad del CPU

Esto es necesario para aquellos ordenadores que tienen más de un procesador en el en donde será instalado el motor Suricata o todos los procesadores a los diferentes hilos de ejecución debido a que suricata es una herramienta multihilos.

Variables para las Reglas

“Las reglas que Suricata usará como patrones para determinar si los paquetes son peligrosos o no se utilizan variables definidas.

Esto permite configurar las redes y direcciones IP a las que se desea asignar las reglas y a cuáles no. Debido a que Snort también utiliza la misma configuración Suricata es compatible con las reglas de Snort por lo que también las soporta.

Por lo que es muy importante modificar estas variables adaptándolas a la red que se encuentra instalado Suricata”. [41]

```
Vars:
Address-groups:
HOME_NET:"[192.168.0.0/16,10.0.0.0/8,172.16.0.0/12]"
EXTERNAL_NET: any
HTTP_SERVERS: "$HOME_NET"
SMTP_SERVERS: "$HOME_NET"
SQL_SERVERS: "$HOME_NET"
DNS_SERVERS: "$HOME_NET"
TELNET_SERVERS: "$HOME_NET"
```

Reglas y firmas de seguridad

"La herramienta de Suricata tiene sus propias reglas con las cuales se puede comparar los paquetes y discernir cuando hay ataques en la red.

Suricata depende de reglas externas para su funcionamiento, pero también es totalmente compatible con las reglas de Snort. Existe un set de reglas tanto pagadas como gratuitas la diferencia entre estas es que el set de reglas pagadas tienen las actualizaciones al día mientras que las gratuitas pueden tardar algunos meses en ser actualizadas. Para la solución se hizo uso del set de reglas gratuitas." [41]

C. Módulo Colector de alertas

"Este módulo es el encargado de coleccionar los archivos de salida estándar Unified2 de Suricata, donde se encuentran los eventos más conocidos como alertas los mismos que son registrados por el motor, dependiendo de la configuración realizada anteriormente.

Estos eventos serán guardados en una base de datos". [41]

Instalación del colector de alertas (Barnyard2)

Instalamos todas las dependencias

- Pacman-S libmysqclient

Se procede a la realización de la copia de las distintas fuentes de barnyard2, localizadas en /usr/local/src y compilarlas como se puede observar a continuación en la Figura IV 27.

```
./configure --bindir=/usr/bin --sysconfdir=/etc/barnyard2 --with-
mysql-libraries=/usr/lib  --with-mysql-includes=/usr/include/
make
make install
```

Figura IV 27: Archivo del barnyard2
Fuente: Autoras

"Agregar en el archivo **/etc/rc.local** el comando de ejecución de barnyard para que corra siempre al incio como se puede ver en la Figura IV 28". [41]

```
barnyard2 -c /etc/barnyard2/barnyard2.conf -d /var/log/suricata/ -f unified2.alert -w
barnyard2.waldo -D
```

Figura IV 28: Ejecución del barnyard
Fuente: Autoras

Configuración del colector de alertas (Barnyard2)

Editamos el archivo /etc/barnyard/barnyard2.conf como se puede ver a continuación, en la Figura IV 29.

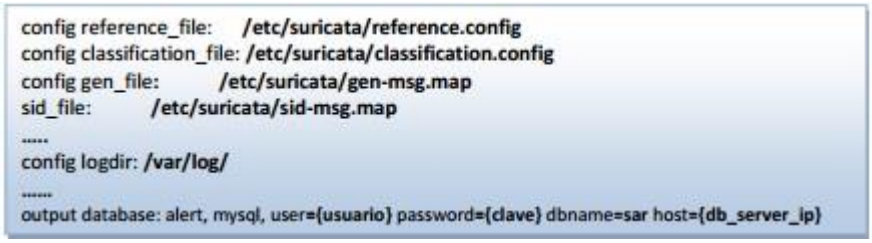


Figura IV 29: Archivo de configuración del barnyard2
Fuente: Autoras

"En la última línea hay que llenar los valores entre llaves con los correctos según la configuración de la base de datos a la que se conecta barnyard". [41]

D. Módulo de Comunicación

"Para la gestión de alertas se hará uso de la interfaz web Snorby. Snorby es un **front-end** para la gestión de alertas basado en sensores, su interface gráfica es muy sencilla con una visión amplia e intuitiva de la visualización de las alertas. Tiene mucha menos configuración y por tanto es más sencillo". [56]

"En resumen, con **Snorby** podemos tener una visión rápida de las alertas generadas por los distintos sensores y poco más". [56]

E. Instalación de Snorby

Primeramente se instala los paquetes y requisitos previos como se puede observar en la siguiente línea de comando.

```
yum install httpd-devel libyaml git ImageMagick ImageMagick-devel libxml2-devel libxslt-devel gcc-c + + rizo-devel httpd-devel abril-devel apr-util-devel readline-devel-y
```

Una vez instalado todos los paquetes necesarios se procede a la descarga e instalación del paquete.

Descargamos e instalamos Rubi mediante la siguiente línea de comando

```
cd / usr / local / src / snort
 wget            http://ftp.ruby-lang.org/pub/ruby/1.9/ruby-1.9.3-
p327.tar.gz
 tar xvzf ruby-1 *
 cd ruby-1 *
 . / Configure && make && make install
```

Instalamos la extensión openssl con la ayuda de la siguiente línea de comando

```
cd ext / openssl /
 ruby extconf.rb
 make && make install
```

Se instala también las dependencias de la gema.

```
gem install thor i18n bundler tzinfo builder
memcache-client rack ack-test erubis mail rack-
mount rails --no-rdoc --no-ri
gem install rake --version=0.9.2 --no-rdoc --no-ri
gem uninstall rake --version=0.9.2.2
```

Descargamos e instalamos wkhtmltopdf como se puede observar en la siguiente línea de comando.

```
cd / usr / local / src / snort
```

Se ha instalado todos los paquetes necesarios pues ahora configuramos Snorby

Configuración del Snorby

Descargamos y configuramos Snorby con la ayuda del siguiente comando.

```
cd / var / www / html /
git clone http://github.com/Snorby/snorby.git
cd / var / www / html / snorby / config /
cp database.yml.example database.yml
cp snorby_config.yml.example snorby_config.yml
chown-R apache: apache / var / www / html / snorby
```

También hay que establecer la contraseña de root para el gestor de base de datos mysql como se puede observar.

```
tesis mysqladmin password
```

De la misma forma se configura el nombre de usuario y la contraseña para Snorby para mysql.

```
vi database.yml
snorby: & snorby
adapter: mysql
username: root
password: tesis
host: localhost
Instale Snorby
cd /var/www/snorby
bundle install --deployment
rake snorby:setup
```

Configure Barnyard que es el encargado de la salida de las alertas a la base de datos mediante el mismo que se encuentra en el siguiente fichero.

```
vi / etc / snort / barnyard.conf
```

Hecho esto reiniciamos el servicio Barnyard. Luego instalamos el módulo de passenger para apache como se puede observar en la siguiente línea de comando.

```
gem install passenger
 cd / usr/local/lib/ruby/gems/1.9.1/gems/passenger-3.0.19/bin
 ./passenger-install-apache2-module
```

Por ultimo configuramos y reiniciamos apache en el archivo de configuración httpd.conf con la ayuda de la siguiente línea de comando vi / etc / httpd / conf / httpd.conf y el archivo de configuración es el siguiente.

```
# <VirtualHost *:80>
# ServerAdmin webmaster@dummy-host.example.com
# DocumentRoot / www / docs / dummy host.example.com
# ServerName dummy host.example.com
# ErrorLog logs / dummy host.example.com-error_log
Logs # CustomLog / maniquí-host.example.com-access_log común
# </ VirtualHost>
LoadModule passenger_module / usr/local/lib/ruby/gems/1.9.1/gems/passenger-
3.0.19/ext/apache2/mod_passenger.so
PassengerRoot / usr/local/lib/ruby/gems/1.9.1/gems/passenger-3.0.19
PassengerRuby / usr / local / bin / ruby


<VirtualHost *:80>
ServerAdmin admin@nachum234.com
ServerName snorby.nachum234.com
DocumentRoot / var / www / html / snorby / public

"/var/www/html/snorby/public"> <Directory
        AllowOverride todo
        Orden negar, permitir
        Dejar de todas
        Opciones-MultiViews
</ Directory>
</ VirtualHost>
```

Por ultimo reiniciamos el servicio apache con el siguiente comando.

```
service httpd restart
```

F. Añadir las reglas al IDS

"Debido a que las reglas de snort son totalmente compatibles con suricata y que el lenguaje usado por snort es flexible y potente, este mismo lenguaje sirve de guía para la escritura de las reglas". [56]

"Las reglas deben ser escritas en una sola línea, las reglas las podemos dividir en dos secciones lógicas como son: **cabecera de la regla y opciones**" [56]

- "La **cabecera** contiene la acción de la regla en sí, protocolo, IPs, máscaras de red, puertos origen y destino y destino del paquete o dirección de la operación". [56]
- "La sección **opciones** contiene los mensajes y la información necesaria para la decisión a tomar por parte de la alerta en forma de opciones" [56]

"Veamos ahora un ejemplo de regla Snort para alertar de un *escaneo nmap*del *tipo* TCP ping". [56]

```
alert tcp $EXTERNAL_NET any -> $HOME_NET any / (msg:"Escaneo ping con
nmap";flags:A;ack:0; / reference:arachnids,28;classtype:attempted-recon; sid:628;/ rev:1;)
```

Analicemos esta alerta:

Cabecera

Acción de la regla: alert.
Protocolo: tcp.
Direccion IP origen: $EXTERNAL_NET *(toda la red).*
Puerto IP origen: any *(cualquiera).*
Direccion IP destino: $HOME_NET *(toda nuestra red).*
Puerto IP destino: any *(cualquiera)* (59).
Dirección de la operación: -> *(puede ser ->, <-,).*

Opciones

Mensaje: msg.
Opciones: flags:A;ack:0; reference:arachnids..(1).

Un poco de teoría:

- **"flags:** *A* Establece el contenido de los flags o banderas TCP, en este caso ACK". [56]
- **"ack**: *0* Caso particular para valor ACK=0, es el valor que pone **nmap** para TCP ping scan". [56]
- **"reference:** *arachnids,28* Referencia un a un Advisory, alerta tipo Bugtrac, etc" [56]
- **"classtype:** attempted-recon Categoría de la alerta según unos niveles predefinidos y prioridades". [56]
- **"sid:** *628* Identificación única para esta regla snort según unos tramos determinados". [56]
- **"rev:** *1* Identificación de la revisión o versión de la regla". [56]

Basándonos en este y varios ejemplos se logró obtener las siguientes reglas.

- alert udp $HOME_NET any -> any 53 (msg:"SE ESTA USANDO RED DE micorsofts.com"; content:"|0a|micorsofts|03|com|00|"; nocase; fast_pattern:only; threshold: type limit, track by_src, count 1, seconds 300; reference:url,labs.alienvault.com; classtype:bad-unknown; sid:2016570; rev:2;)
- alert udp $HOME_NET any -> any 53 (msg:"ERROR EN EL DOMAIN hotmal1.com"; content:"|07|hotmal1|03|com|00|"; nocase; fast_pattern:only; threshold: type limit, track by_src, count 1, seconds 300; reference:url,labs.alienvault.com; classtype:bad-unknown; sid:2016571; rev:1;).
- alert udp any 53 -> $HOME_NET any (msg:"ALGUIEN QUE ESTA INTENTANDO ENTRAR - 46.149.18.14 blacklistthisdomain.com"; content:"|00 01 00 01|"; content:"|00 04 2e 95 12 0e|"; distance:4; within:6; classtype:trojan-activity; sid:2016591; rev:5;).
- alert udp $HOME_NET any -> any 53 (msg:"DNS SOSPECHOSO *.upas.su domain"; content:"|01 00 00 01 00 00 00 00 00 00|"; depth:10; offset:2; content:"|04|upas|02|su|00|"; fast_pattern; nocase; distance:0; classtype:bad-unknown; sid:2015550; rev:1;).
- alert udp $HOME_NET any -> $EXTERNAL_NET 53 (msg:"DNS SOSPECHOSO .nl.ai Domain"; content:"|01 00 00 01 00 00 00 00 00 00|"; depth:10; offset:2; content:"|02|nl|02|ai"; fast_pattern; nocase; distance:0; classtype:bad-unknown; sid:2013861; rev:1;).
- alert udp $HOME_NET any -> $EXTERNAL_NET 53 (msg:"DNS SOSPECHOSO .xe.cx Domain"; content:"|01 00 00 01 00 00 00 00 00 00|"; depth:10; offset:2;

content:"|02|xe|02|cx"; fast_pattern; nocase; distance:0; classtype:bad-unknown; sid:2013862; rev:1;).

- alert udp $HOME_NET any -> $EXTERNAL_NET 53 (msg:"DNS SOSPECHOSO .noip.cn Domain"; content:"|01 00 00 01 00 00 00 00 00 00|"; depth:10; offset:2; content:"|03|noip|02|cn|00|"; fast_pattern; nocase; distance:0; classtype:bad-unknown; sid:2013970; rev:1;).
- alert udp $HOME_NET any -> $EXTERNAL_NET 53 (msg:"DNS SOSPECHOSO .eu.tf Domain"; content:"|01 00 00 01 00 00 00 00 00 00|"; depth:10; offset:2; content:"|02|eu|02|tf"; fast_pattern; nocase; distance:0; classtype:bad-unknown; sid:2013848; rev:1;).
- alert udp $HOME_NET any -> $EXTERNAL_NET 53 (msg:"DNS SOSPECHOSO .int.tf Domain"; content:"|01 00 00 01 00 00 00 00 00 00|"; depth:10; offset:2; content:"|03|int|02|tf"; fast_pattern; nocase; distance:0; classtype:bad-unknown; sid:2013849; rev:1;).

CAPÍTULO V

ANÁLSIS DE RESULTADOS

5.1 Introducción

Este capítulo se lo ha dividido en dos partes: La primera se refiere al análisis de los datos obtenidos en la implementación del modelo propuesto en la red de datos de la FIE, la segunda parte está conformada por el desarrollo y el análisis de la aplicación del modelo propuesto en el prototipo planteado. También la comprobación de la hipótesis de este trabajo.

5.2 Análisis de resultados de la implementación del modelo propuesto en la red de datos de la FIE.

Para obtener los resultados en la red de datos de la facultad de Informática y Electrónica de la ESPOCH se utilizó dos puertos, un puerto spam y un puerto de administración los mismos que fueron configurados por el departamento de DTIC (Dirección de Tecnologías de Información y Comunicación).

Para realizar el análisis se tomó en cuenta el parámetro rendimiento en relación de un IDS tradicional y un IDS con minería de datos el cual se va a medir con la ayuda de las siguientes variables: números de ataques detectados, números de falsos positivos detectados, tiempo de duración para la detección de intrusos, tipos de intrusos detectados.

Las pruebas que se realizaron en la facultad durante un período de dos días, en los cuales el primero se hizo las pruebas con un IDS tradicional y el segundo con el IDS con reglas generadas con la técnica de minería de datos. Los datos que se obtuvieron se observan en la Tabla V VIII.

Tabla V VIII: Resultados obtenidos en la red de información de la ESPOCH

No. Ataques Detectados		
Horas	**IDS con reglas minería de datos**	**IDS Tradicional**
7:00-10:00	884319	678345
10:00-13:00	778134	593000
13:00-16:00	567352	453845
16:00-19:00	876924	524387
19:00-20:00	877645	814123

Fuente: Autoras

A continuación se muestra la representación de la Tabla V VIII, como se puede observar en la Figura V 30.

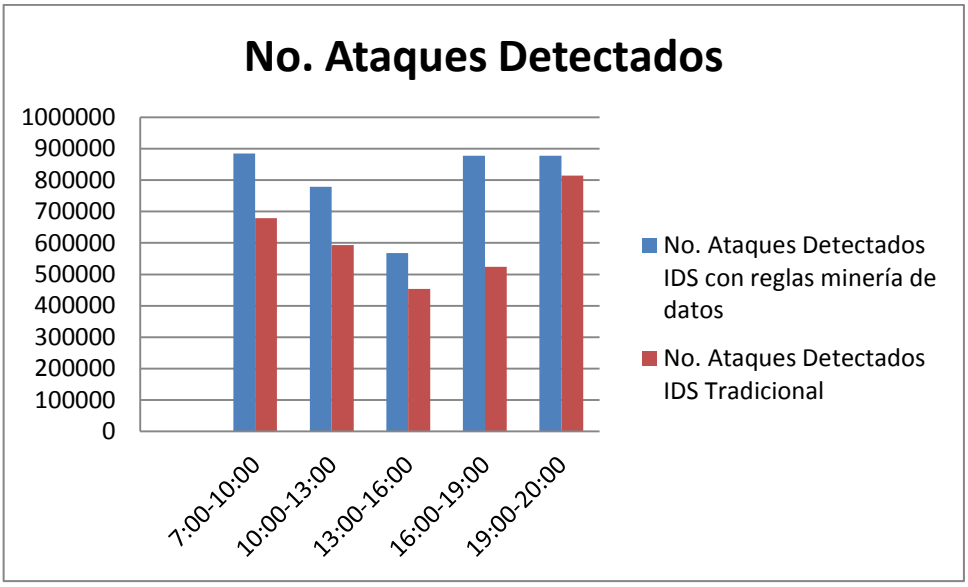


Figura V 30: Resultados obtenidos en la red de información de la FIE
Fuente: Autoras

Los tipos de ataques que se detectó en la red de datos de la FIE con un IDS con las reglas de minería de datos se los puede observar en la Tabla V IX.

Tabla V IX: Datos de la red de la FIE con el IDS con reglas de minería de datos

Tipo de ataques	Número de ataques				
	Hora 1	**Hora 2**	**Hora 3**	**Hora 4**	**Hora 5**
DOS: Denegación de servicios	226264	236608	172147	230656	237066
R2L: Acceso no autorizado a una maquina remota	206362	180651	120588	203209	211188
U2R: Acceso no autorizado a los privilegios de super usuario	215954	177143	141098	214881	202256
PROBE: Monitorización	235739	183732	133521	228178	227135
TOTALES	884319	778134	567354	876924	877645

Fuente: Autoras

La representación gráfica de los mismos se puede observar en la Figura V 31.

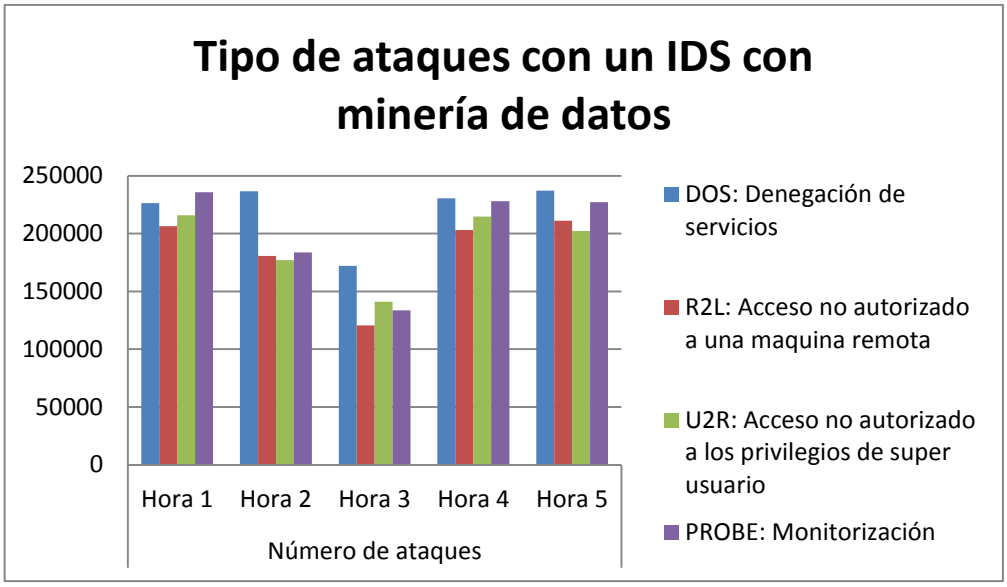


Figura V 31: Datos de la red de datos de la FIE con el IDS con reglas de la minería de datos
Fuente: Autoras

Los tipos de ataques que se detectó en la red de datos de la FIE con un IDS tradicional se los puede observar en la Tabla V X.

Tabla V X: Datos de la red de datos de la FIE con un IDS tradicional

Tipo de ataques	Número de ataques				
	Hora 1	Hora 2	Hora 3	Hora 4	Hora 5
DOS: Denegación de servicios	145180	169600	133592	130521	203803
R2L: Acceso no autorizado a una maquina remota	167366	152100	113592	130323	193626
U2R: Acceso no autorizado a los privilegios de super usuario	187977	126100	103386	121417	195892
PROBE: Monitorización	177822	145200	103275	141584	220802
TOTALES	678345	593000	453845	523845	814123

Fuente: Autoras

Representación gráfica de los tipos de intrusos detectados se los puede visualizar en la Figura V 32.

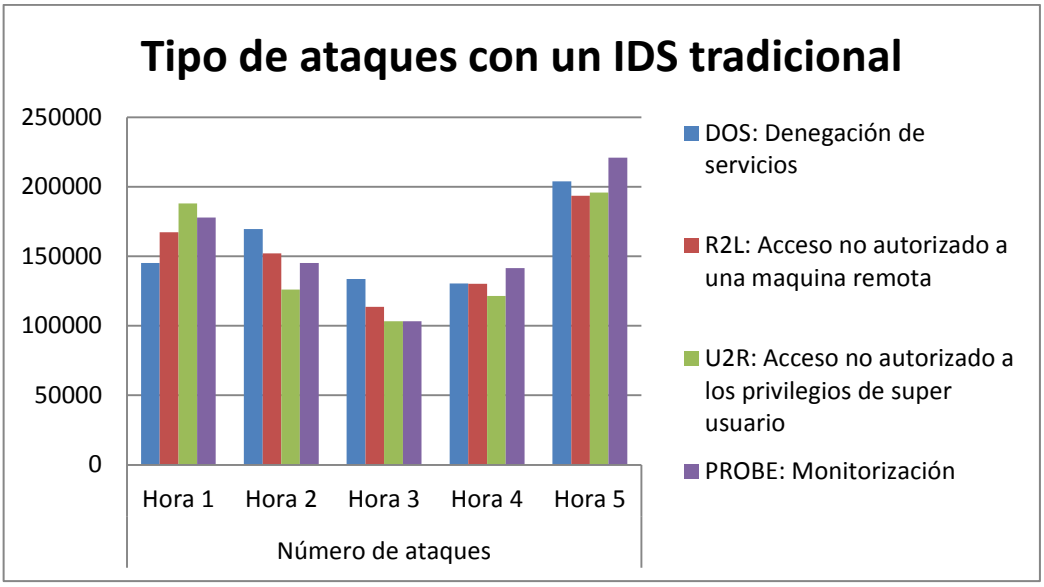


Figura V 32: Datos de la red de datos de la FIE con un IDS tradicional
Fuente: Autoras

5.2 Desarrollo y análisis de la aplicación del modelo propuesto en el prototipo planteado

Para la realización de la pruebas se armó un prototipo en cual se tiene configurado un servidor DNS y DHCP dinámico, servidor web, servidor FTP, servidor SSH y el servidor de correo como se puede observar un bosquejo del prototipo en la Figura V 33.

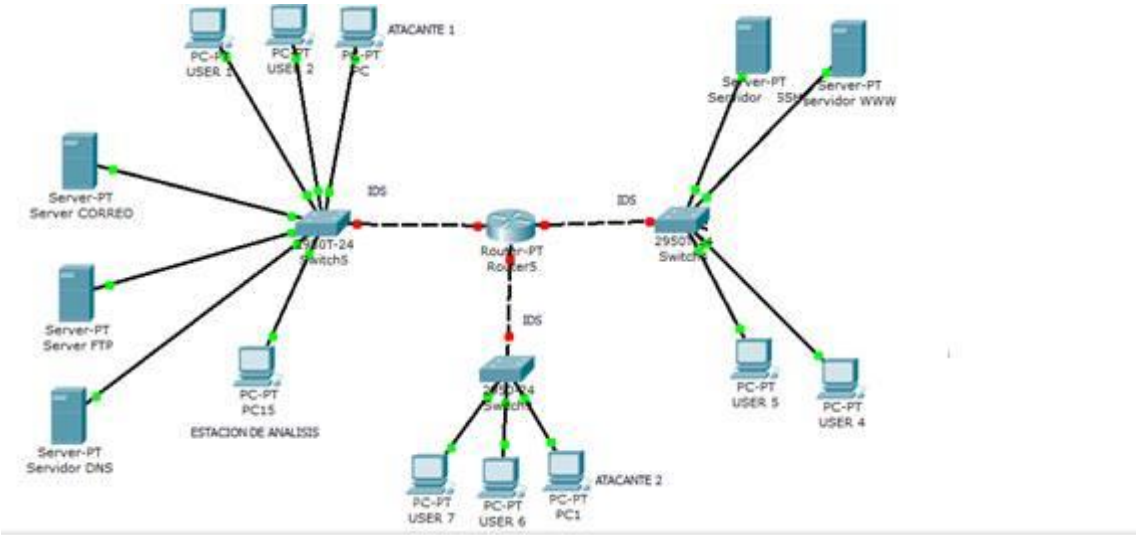


Figura V 33: Escenario del prototipo
Fuente: Autoras

Para la implementación del prototipo se tuvo que configurar los servidores DNS, FTP, HTTP y de Correo.

"Un servidor, es un ordenador o máquina informática que está al "servicio" de otras máquinas, ordenadores o personas llamadas clientes y que le suministran a estos, todo tipo de información". [57]

"Por tanto un servidor en informática será un ordenador u otro tipo de dispositivo que suministra una información requerida por unos clientes (que pueden ser personas, o también pueden ser otros dispositivos como ordenadores, móviles, impresoras, etc.)". [57]

Servidor DNS

"DNS son las siglas de Domain Name Service (Servicio de Nombre de Dominio), y sirve para traducir los nombres de dominio (por ejemplo www.turismotour.com) en direcciones IP y viceversa". [58]

"Para entender mejor que son los DNS vamos e explicarlo mediante un gráfico que representa como se compone la estructura de Internet para la petición de una página web". [58]

"Un servidor DNS nos va a servir para la resolución de nombres cuando queramos acceder a las web que tenemos alojadas en nuestra intranet" [58], como se puede ver en la Figura V 34.

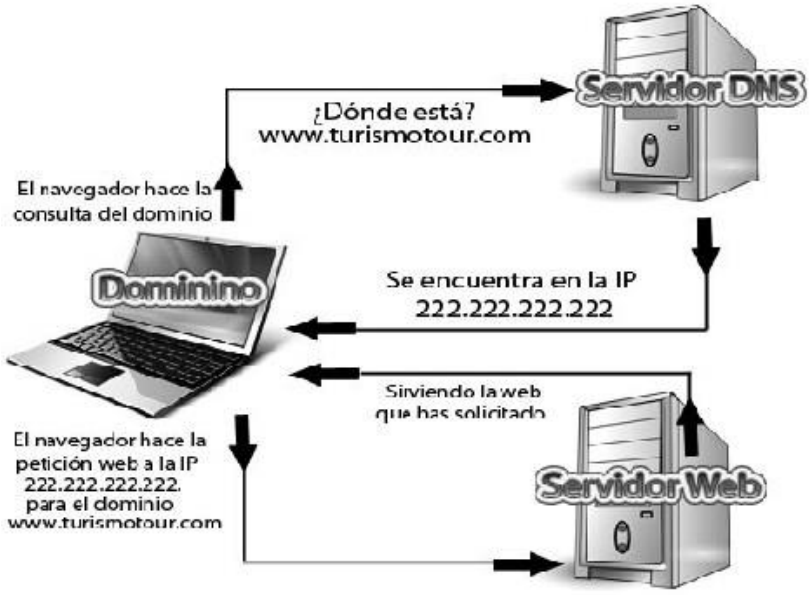


Figura V 34: Funcionamiento del servidor DNS
Fuente: http://www.turismotour.com/%C2%BFque-son-los-servidores-dns/

5.2.1 Configuración de los servidores DHCP y DNS dinámicos

Como primer paso se procede a verificar si el dhcp está instalado caso contrario se procede a instalar con el comando rpm –ihv dhcp como se puede ver Figura V 35.

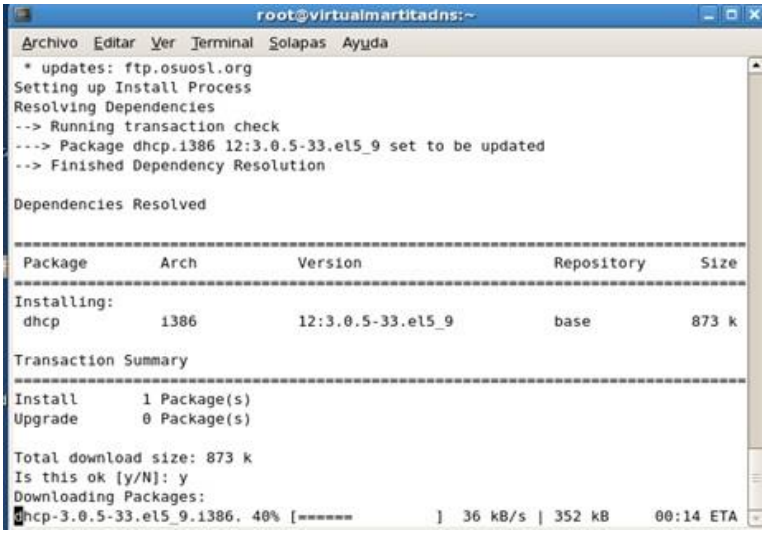


Figura V 35: Instalación del paquete DHCP
Fuente: Autoras

Una vez instalado el dhcp iniciamos el servicio con el comando service dhcpd start como se muestra en la Figura V 36.

Figura V 36: Inicialización del servicio DHCPD
Fuente: Autoras

La continuación de la configuración tanto del servidor DNS, DHCP y demás servidores se puede observar en la sección de ANEXOS 1: Implementación y configuración del prototipo.

5.2.2 Pruebas realizadas en el prototipo

Con el prototipo armado se procede a realizar las respectivas pruebas para lo cual se tuvo que realizar ataques, los mismos que se realizó con la ayuda de backtrack del cual a continuación de describe su instalación y uso.

"BackTrack es una distribución GNU/Linux en formato LiveCD pensada y diseñada para la auditoría de seguridad y relacionada con la seguridad informática en general. Actualmente tiene una gran popularidad y aceptación en la comunidad que se mueve en torno a la seguridad informática". [59]

"Se deriva de la unión de dos grandes distribuciones orientadas a la seguridad, el Auditor + WHAX. WHAX es la evolución del Whoppix (WhiteHat Knoppix), el cual pasó a basarse en la distribución Linux SLAX en lugar de Knoppix. La última versión de esta distribución, aún en estado Beta, cambió el sistema base, antes basado en Slax y ahora en Ubuntu". [59]

Instalar Backtrack 5

Procedemos a descargar Backtrack 5 desde la página principal http://www.backtrack-linux.org/

Para la instalación es necesario grabar lo que se ha descargado en un DVD o usb reiniciamos la máquina y aparecerá la Figura V 37 se selecciona la primera opción "Backtrack Text – Default Boot Text Mode".

Figura V 37: Pantalla inicial Backtrack 5
Fuente: Autoras

Al arrancar correctamente aparecerá la Figura V 38.

Figura V 38: Instalación
Fuente: Autoras

Una vez concluida la instalación aparecerá la Figura V 39 se digita “startx” y se presiona enter para continuar.

Figura V 39: Iniciando Backtrack 5
Fuente: Autoras

En la parte superior izquierda aparecerá el icono de Install BackTrack damos doble click y comenzara la instalación como se observa en la Figura V 40.

Figura V 40. Install Backtrack
Fuente: Autoras

Seleccionamos el idioma Figura V 41 y presionamos adelante.

Figura V 41: Selección del Idioma
Fuente: Autoras

Elegimos la zona horaria # como se muestra en la Figura V 42.

Figura V 42: Zona Horario
Fuente: Autoras

Para mayor información sobre la instalación de esta herramienta encontrará en la parte de ANEXOS 2: Instalación del backtrack.

ATAQUES REALIZADOS EN ELPROTOTIPO

Ataque Arp-Spoofing

"Es una técnica empleada para infiltrarse en una red y de esta forma relacionar la dirección MAC del atacante con la dirección IP de otro nodo (el nodo atacado)". [60]

"El atacante, puede entonces adoptar, entre reenviar el tráfico a la puerta de enlace predeterminada real (ataque pasivo o escucha), o cambiar los datos antes de reenviarlos (ataque activo) y de esta forma producir problemas al nodo atacado". [60]

"El principal objetivo de efectuar este tipo de ataque es interponerse entre una o varias máquinas con el fin de interceptar, modificar o capturar paquetes" [60] Backtrack cuenta con una variedad de herramientas entre ellas se encuentra Ettercap para realizar el ataque MitM (Man in the MIddle). Para usar la herramienta Ettercap de Backtrack 5 se debe abrir una terminal y digitamos ettercap –G como se puede ver en la Figura V 43.

Figura V 43: Iniciando Ettercap NG-0.7.3
Fuente: Autoras

Es necesario configurar un par de líneas del archivo de configuración Ettercap para el correcto funcionamiento se digita en consola nano /etc/etter.conf y se abrirá el archivo como se observa en la Figura V 44.

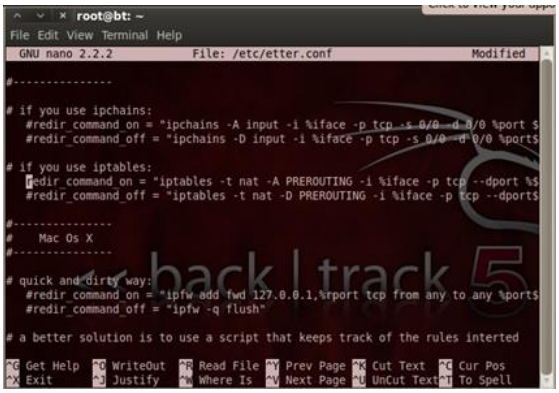


Figura V 44: Arcivo etter.conf
Fuente: Autoras

En el archivo se debe des comentar la línea que se muestra en la Figura V 45 salimos guardando los cambios del archivo.

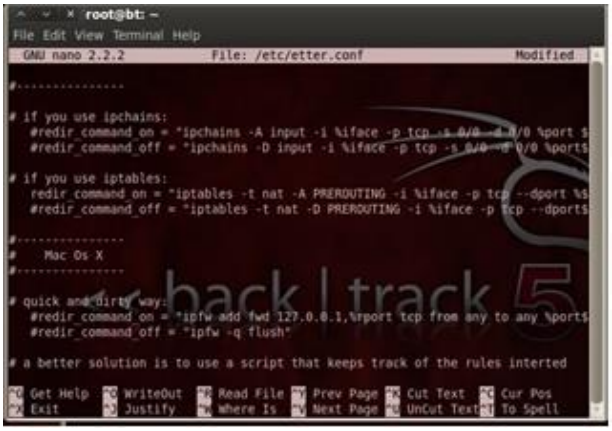


Figura V 45: Modificar archivo
Fuente: Autoras

Para acceder a la pantalla inicial de Ettercap se lo puede realizar de dos formas la primera abriendo una terminal y digitando Ettercap –G y se abrirá la interfaz gráfica de Ettercap la segunda forma es desde el menú de Backtrack que está ubicado en la parte superior izquierda y se abrirá la interfaz gráfica ver en la Figura V 46.

Figura V 46: Pantalla inicio Ettercap
Fuente: Autoras

Hacemos clic en la opción Sniff a continuación aparecerá un submenú en el cual seleccionamos Unifield Sniffing que permite seleccionar la interfaz de red ver en la Figura V 47.

Figura V 47: Selección de interfaz de red
Fuente: Autoras

A continuación el menú cambiara y mostrara más opciones se selecciona Hosts, aparecerá un submenú en el cual seleccionamos Scan for Hosts, realiza un escaneo de la red encontrando los equipos conectados. Seleccionamos host y en el sub menú seleccionamos Host List y se verá la lista de todos los equipos conectados a la red ver en la Figura V 48.

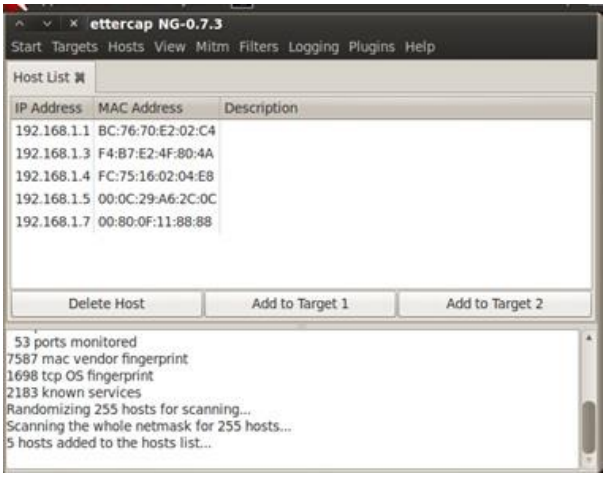


Figura V 48: Lista de equipos
Fuente: Autoras

Ahora se debe elegir el equipo que será la victima del ataque objetivo 1 (Target1) y como objetivo 2 (Target 2) seleccionaremos al Router.

Procedemos a realizar el ataque seleccionando del menú en la opción MitM a continuación del submenú seleccionamos Arp Poisoning aparece una pantalla en la que selecionamos Sniff Remote Connections y para continuar presionmas OK como se puede ver en la Figura V 49.

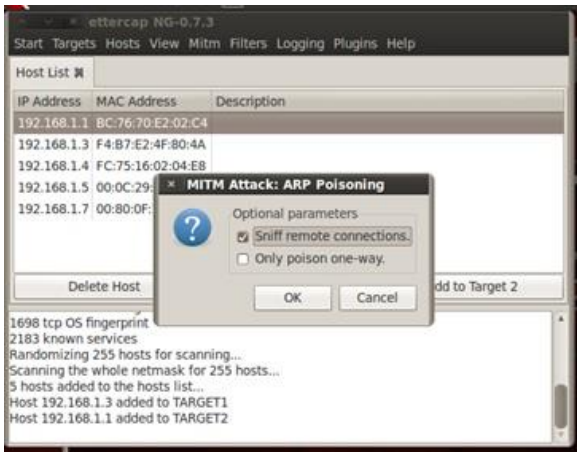


Figura V 49: Selección de la herramienta
Fuente: Autoras

Para que inicialice el ataque debemos dar clic en Start en esta en el menú y acontinuacion clic en Start sniffing y se ejecuta el ataque como se puede ver en la Figura V 50.

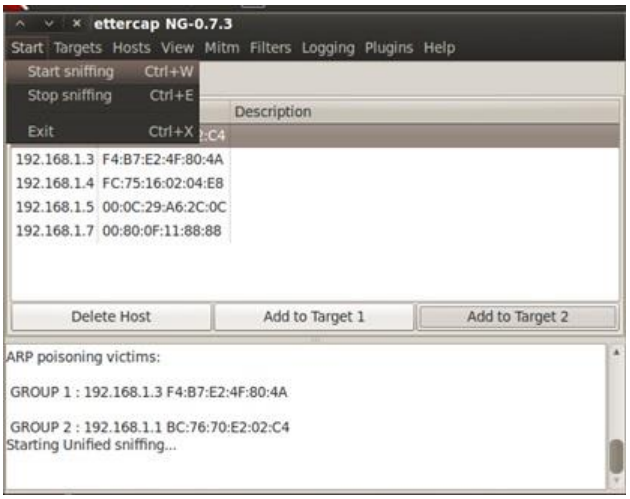


Figura V 50: Inicialización del ataque
Fuente: Autoras

Para observar todo el tráfico de la víctima debemos dar clic en View del menú y seleccionar Profiles y podremos observar los sitios que visita como se puede observar en la Figura V 51.

Figura V 51: Visualización del tráfico de la víctima
Fuente: Autoras

Para conocer con detalle el tráfico que la víctima se hará uso de Sstrip de bactrack para poder realizarlo se deberá detener el ataque MitM, para detener el ataque damos clic en Mitm del menú y luego se selecciona Stop mitm attack .

Arrancamos nuevamente ettercap –G en un terminal y abrimos otra terminal y ejecutamos el siguiente comando echo "1" >/proc/sys/net/ipv4/ip_forward para verificar ejecutamos cat/proc/sys/net/ipv4/ip_forward como se puede ver en la Figura V 52.

Figura V 52: Inicia backtrack
Fuente: Autoras

Del menú de herramientas de Backtrack selecionamos Informatio Gathering, Network Analysis, SSLAnalysis, SSLtrip como se puede observar en la Figura V 53.

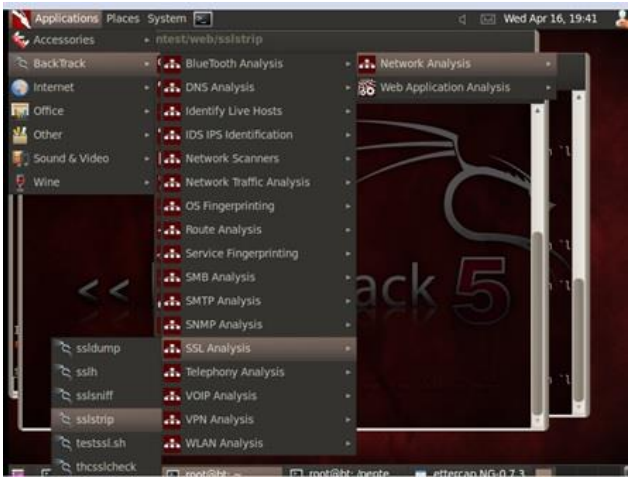


Figura V 53: Informatio Gathering
Fuente: Autoras

Se abrirá una consola en la que se ejecuta el siguiente comando sudo python ./setup.py install como se puede observar en la Figura V 54.

Figura V 54: Consola del backtrack
Fuente: Autoras

Se escanea la red utilizando Ettercap, seleccionamos la interfaz de red, listamos los host seleccionamos los equipos que van a ser atacados, clic en Mitm y seleccionamos ARP Poisnoning aparecerá una ventana seleccionamos la primera casilla igual procedimiento que para el ataque MitTM

Se abrirá una nueva consola en la que se ejecutara el siguiente comando iptables –t nat –A PREROUTING -p tcp-destination-port 80 –j REDIRECT --to-port 8080 como se puede observar en la imagen Figura V 55.

Figura V 55: Pantalla del Ettercap
Fuente: Autoras

En la consola de SslStrip que ya se tenía abierta ejecutamos el siguiente comando sudo sslstrip –l 8080 –w /root/Desktop/log.txt como se puede observar en la Figura V 56.

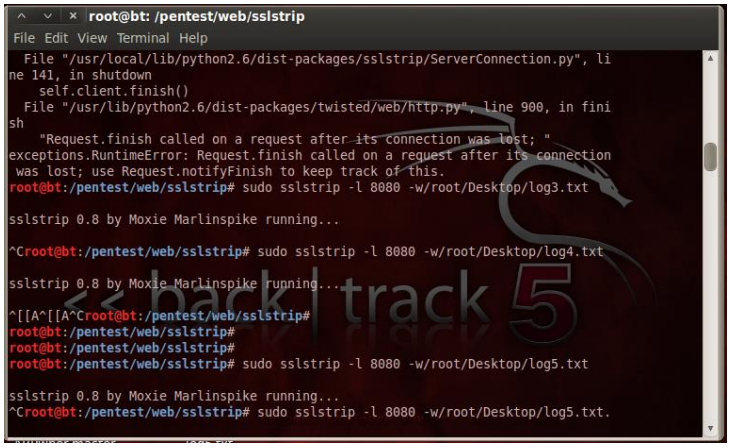


Figura V 56: Consola del SslStrip
Fuente: Autoras

Para ver los detalles de la captura lo realizamos abriendo e log que creamos como se puede observar en la Figura V 57.

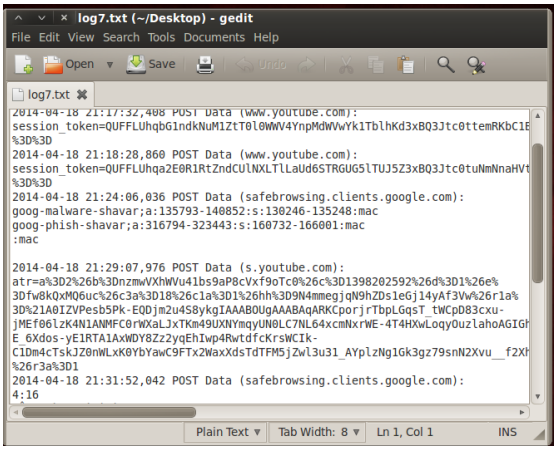


Figura V 57: Pantalla de Resultados
Fuente: Autoras

La herramienta ettercap permite ver las páginas que visita la victima seleccionamos View del menú principal y del menú la opción Profiles como se puede ver en la Figura V 58.

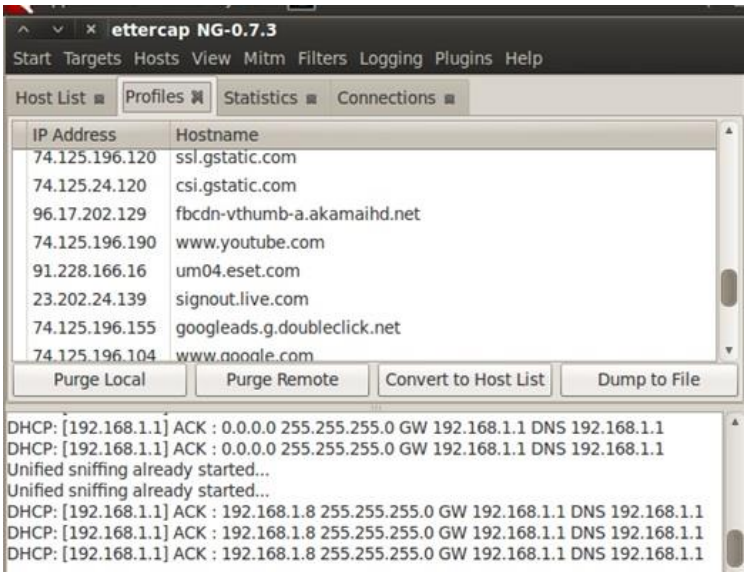


Figura V 58: Herramienta ettercap
Fuente: Autoras

La herramienta ettercap permite ver el estado de las conexiones de la víctima seleccionamos View del menú principal y del menú la opción Connections como se puede ver en la Figura V 59.

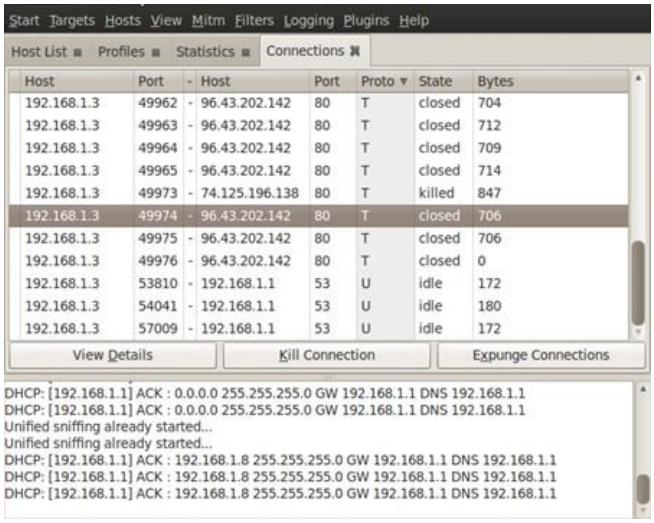


Figura V 59: Conexiones de la víctima
Fuente: Autoras

Hping es una herramienta de comandos que permite analizar paquetes TCP/IP, tiene muchas utilidades: escaneo de red, testing de firewalls, redes y tiene la capacidad de provocar un SYN Flood Attack (DDos).

Para instalar la herramienta se debe abrir una ventana y ejecutar el siguiente comando sudo apt-get install hping3 como se puede ver en la Figura V 60.

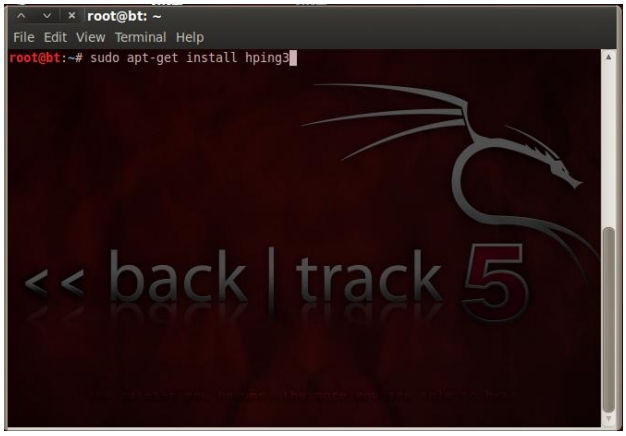


Figura V 60: Herramienta Hping
Fuente: Autoras

Para realizar el ataque ejecutamos el siguiente comando hping3 –p 80 –S –flood ip_victima como se puede ver en la Figura V 61.

Figura V 61: Ejecución del comando hping3 –p 80 –S –flood ip_victima
Fuente: Autoras

Dónde: -p 80 es e puerta que se ha elegido para realizar el ataque, -S activa el flag Synflood le indica a hping que envie los paquetes a la máxima velocidad posible, ip_victima es la ip que se va a realizar el ataque como se puede ver en la Figura V 62.

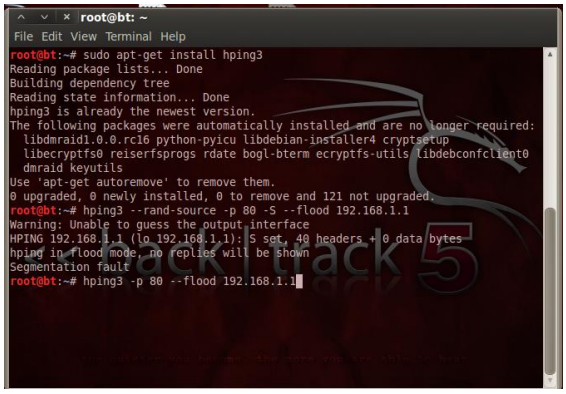


Figura V 62: Ejecución del ataque
Fuente: Autoras

CLONACION DE DNS

Para realizar el ataque de clonación de DNS se deberá en el Backtrack abrir una consola y ejecutar el siguiente comando dc /pentest/exploits/set/ y a continuación el comando./set y se abrirá la aplican SET para realizar ataques de tipo ingeniería social como se puede observar en la Figura V 63.

Figura V 63: Clonación del DNS
Fuente: Autoras

De la lista de opciones que se presenta seleccionamos la primera opción Social-Engneering Attacks, a continuación se seleccionamos Website Attack Vectors como se puede observar en la Figura V 64.

Figura V 64: Website Attack Vectors
Fuente: Autoras

En el siguiente menú de Multi-Attack selecionamos la tercera opción Credential Harvester Attack Method como se puede ver en la Figura V 65.

Figura V 65: Credential Harvester Attack Method
Fuente: Autoras

A continuación seleccionamos Site Cloner que significa clonar una página en este caso se clonara la página de Facebook como se puede ver en la Figura V 1.

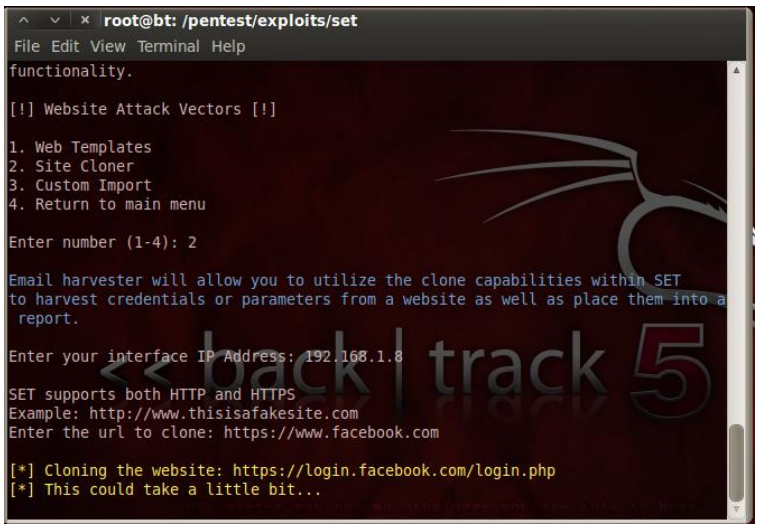


Figura V 1 Site Cloner
Fuente: Autoras

Hasta aquí se tiene el envenenamiento por DNS, la victima tendría que poner la dirección IP en el navegador, y a continuación aparecería l mensaje de login de Facebook. Para que el usuario no ingrese la IP en el navegador, se ocupara una herramienta llamada ettercap , esta herramienta se utilizara de forma que el usuario

pueda ingresar el nombre de la página y se re direccione a la página clonada de esta manera se envenena automáticamente a continuación aparecerá la página de login

Para esto ingresamos a la ubicación /usr/local/share/ettercap/etter.dns y se agregó los siguientes parámetros:

Dominio.com A IP

*.dominio A IP

www.dominio.com como se puede observar en la Figura V 67.

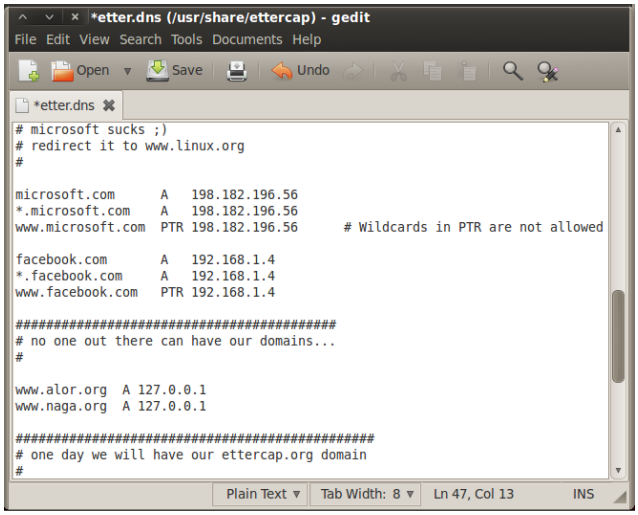


Figura V 67: Etter.dns
Fuente: Autoras

Para continuar realizamos el envenenamiento ejecutando el siguiente comando ettercap –T –q –i eth0 –P dns_spoof –M arp // // como se puede ver en la Figura V 68.

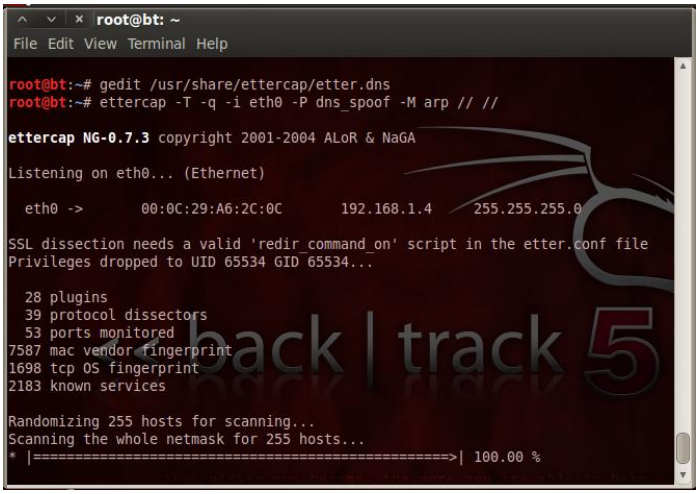


Figura V 68: Ejecución del ataque
Fuente: Autoras

Del lado de la víctima ingresamos en el navegador https://www.facebook.com y aparecerá la página de login, la victima ingresara la cuenta y contraseña como se puede ver en la Figura V 69.

Figura V 69: Captura de datos confidenciales
Fuente: Autoras

Captura la cuenta y la contraseña de la cuenta de Facebook revisando en backtrack la consola que se utilizó las herramientas de ingeniería social imagen podemos observar que se ha capturado la información como se puede ver en la Figura V 70.

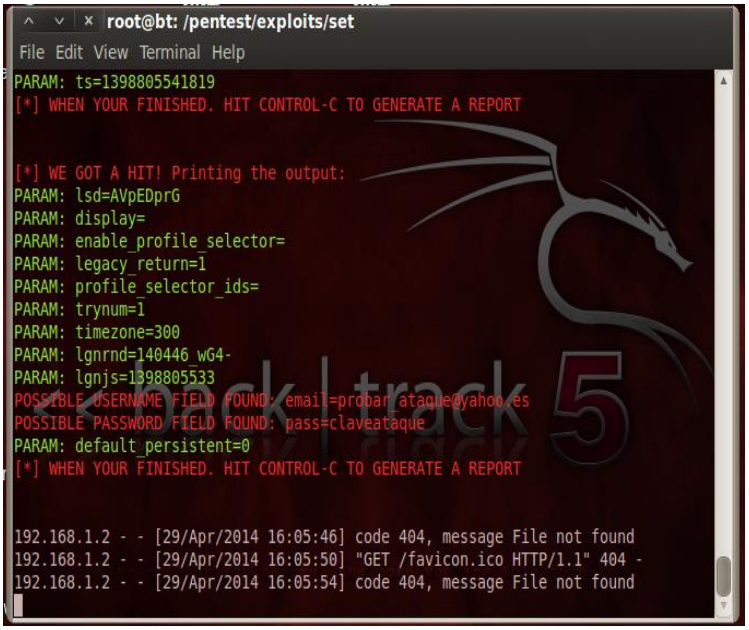


Figura V 70: Datos capturados
Fuente: Autoras

Una vez que se realizaron todos los ataques mencionados anteriormente dentro del prototipo implementado los datos que se obtuvieron son los siguientes.

5.5 Análisis de resultados

5.5.1 Comparación de resultados

Tanto en el prototipo como en la red de información de la FIE facultad de la ESPOCH las pruebas se realizaron con un IDS tradicional y un IDS en el cual se cargó las reglas obtenidas con la minería de datos y los resultados que arrojaron son los que se muestra a continuación. En el prototipo las pruebas se las realizaron durante cinco días, una semana en diferentes horarios.

Resultados del prototipo

Datos obtenidos en un IDS tradicional, sin cargar las reglas como se puede observar en la Tabla V XI.

Tabla V XI: Datos de un IDS tradicional

Días	Ataques detectados	Falsos positivos	Tiempo
Día1 6:00-10:00	2680	120	0.001
Día 2 10:00-14:00	4870	132	0.001
Día 3 14:00-18:00	7538	143	0.001
Día 4 18:00-22:00	5243	87	0.001
Día 5 22:00-6:00	6141	75	0.001

Fuente: Autoras

Resultados obtenidos representados gráficamente en la Figura V 71.

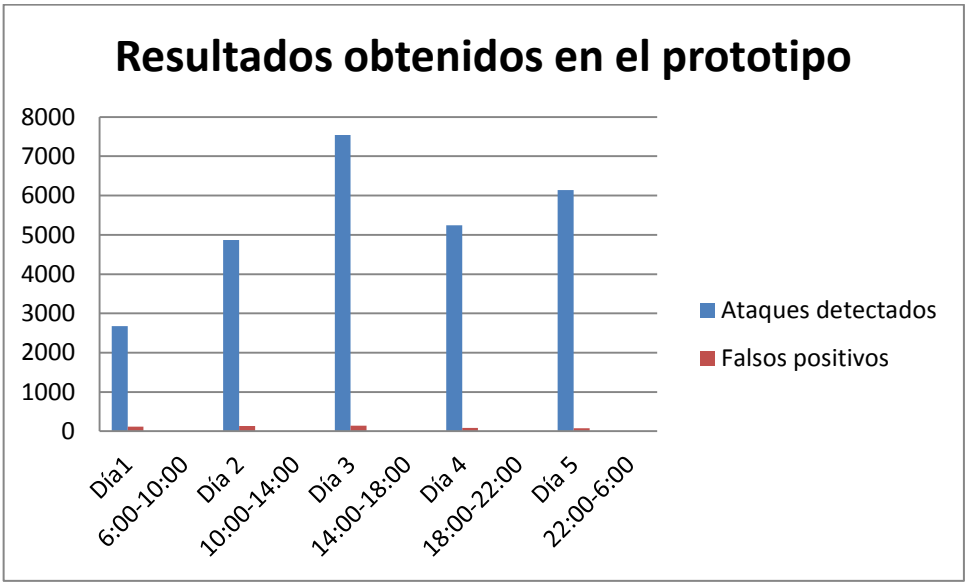


Figura V 71: Datos de un IDS tradicional
Fuente: Autoras

Tipos de ataques detectados en el prototipo con un IDS tradicional

En la Tabla V XII se realiza la clasificación de los tipos de ataques detectados en las pruebas realizadas.

Tabla V XII: Tipos de ataques

Tipo de ataques	Número de ataques				
	Día 1	Día 2	Día 3	Día 4	Día 5
DOS: Denegación de servicios	688	1253	1863	1541	1548
R2L: Acceso no autorizado a una maquina remota	537	1158	2185	1089	1440
U2R: Acceso no autorizado a los privilegios de súper usuario	578	1084	1515	1377	1369
PROBE: Monitorización	757	1243	1832	1149	1709
TOTALES	2560	4738	7395	5156	6066

Fuente: Autoras

Datos de los tipos de intrusos detectados representados gráficamente se los puede ver en la Figura V 72.

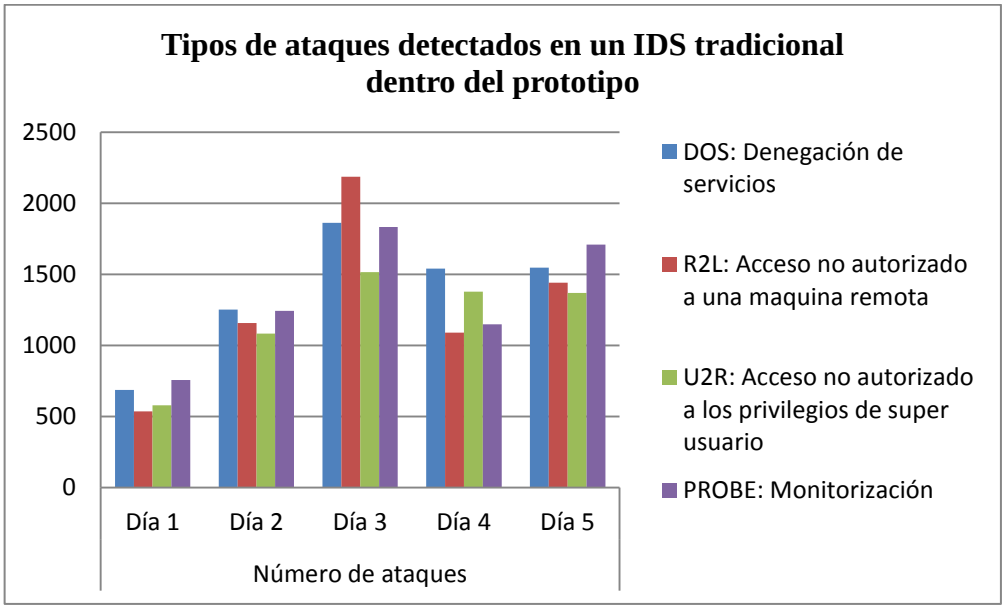


Figura V 72: Tipos de ataques
Fuente: Autoras

Datos obtenidos en un IDS con reglas de minería de datos

Los datos que se muestran a continuación son datos de un IDS en el cual se cargaron las reglas que se obtuvo de la minería de datos como se puede observar en la Tabla V XIII.

Tabla V XIII: Datos obtenidos en el IDS con reglas de la minería de datos

Días	No. Ataques detectados	No. Falsos positivos	Tiempo()
Día1 6:00-10:00	5040	20	0.003
Día 2 10:00-14:00	8320	30	0.003
Día 3 14:00-18:00	10410	50	0.003
Día 4 18:00-22:00	6380	25	0.003
Día 5 22:00-6:00	9520	18	0.003

Fuente: Autoras

La representación gráfica de los datos se puede observar en la Figura V 73.

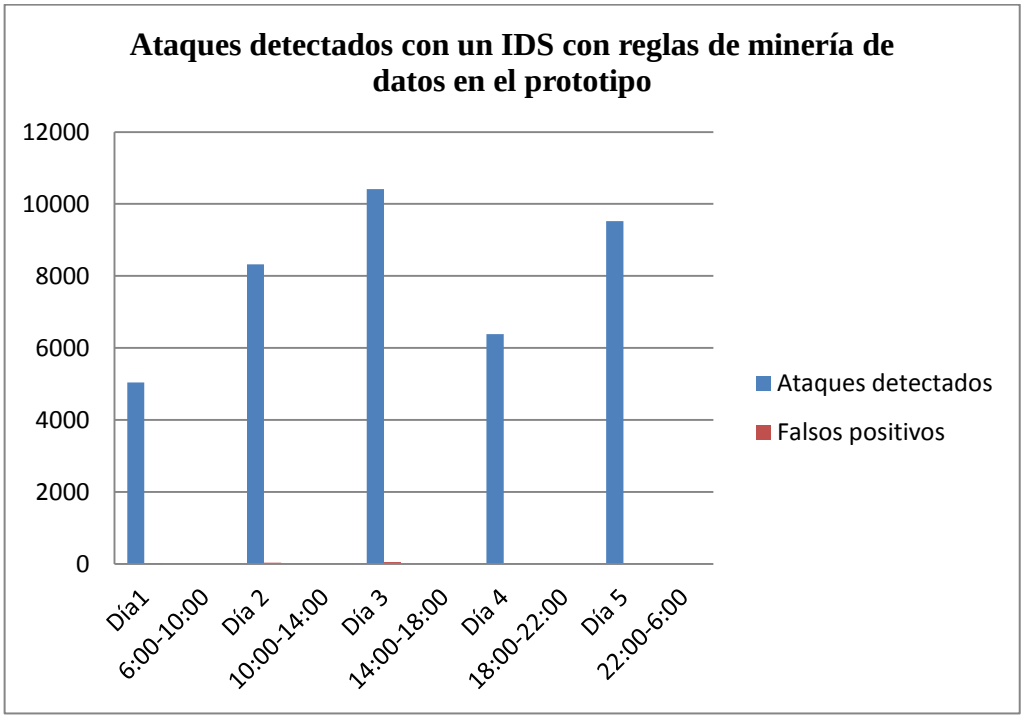


Figura V 73: Datos del IDS con reglas
Fuente: Autoras

Los tipos de intrusos detectados se puede observar en la Tabla V XIV.

Tabla V XIV: Tipos de intrusos detectados con reglas

Tipo de ataques	Número de ataques				
	Día 1	Día 2	Día 3	Día 4	Día 5
DOS: Denegación de servicios	1527	2188	2827	1689	3672
R2L: Acceso no autorizado a una maquina remota	1232	2073	2352	1279	2376
U2R: Acceso no autorizado a los privilegios de super usuario	1120	1709	2440	1798	2436
PROBE: Monitorización	1341	2320	2741	1589	1018
TOTALES	5220	8290	10360	6355	9502

Fuente: Autoras

Representación gráfica de los tipos de ataques se puede observar en la Figura V 74.

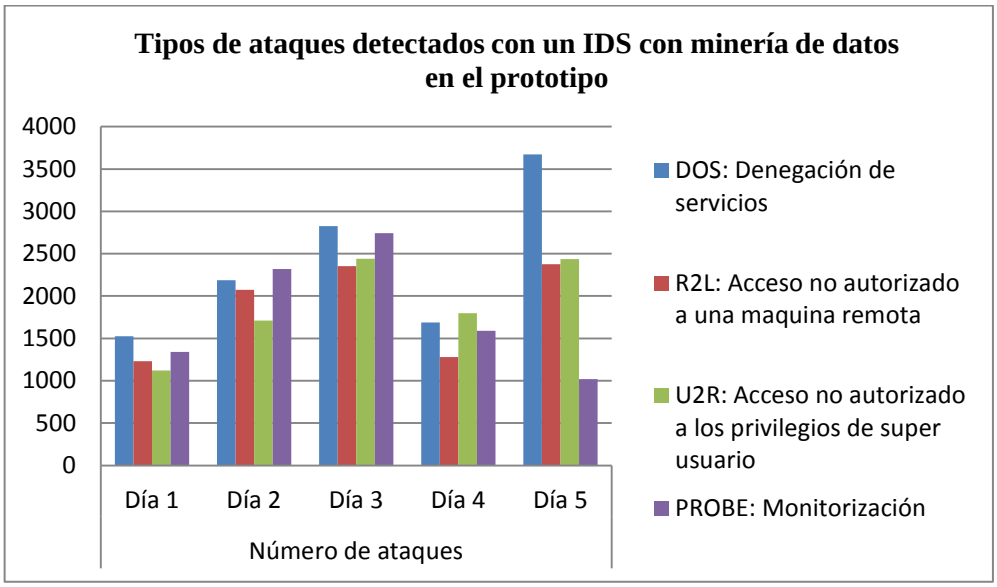


Figura V 74: Tipos de intrusos detectados con reglas
Fuente: Autoras

Como se puede observar los resultados en el prototipo el IDS tradicional captura menos ataques, mientras que el IDS al que se ha cargado las reglas propias del IDS y las reglas obtenidas con la ayuda de la técnica de minería de datos el rendimiento es mejor.

5.6 Comprobación de la hipótesis

5.6.1 Planteamiento de la hipótesis

Hi: El modelo basado en técnicas de Data Mining permitirá la detección del tráfico anómalo de la red de datos de la Facultad de Informática y Electrónica.

Ho: El modelo basado en técnicas de Data Mining no permitirá la detección del tráfico anómalo de la red de datos de la Facultad de Informática y Electrónica.

5.6.2 Determinación de las variables

En la hipótesis planteada para la comprobación de los resultados de este trabajo de investigación las variables encontradas son las siguientes.

Variable Independiente

- Modelo basado en técnica de data mining

Variable Dependiente

- Detección del tráfico anómalo

5.6.3 Operacionalización conceptual de las variables

Tabla V XV: Operacionalización conceptual

Variable	Tipo	Concepto
Modelo basado en las técnica de data mining	Independiente	Es una tecnología que permite la explotación, consistente en la búsqueda de información valiosa (conocimiento) en grandes volúmenes de datos.
Detección del tráfico anómalo	Dependiente	Es el número o la cantidad de anomalías o intrusiones se logre observar al momento de aplicar el modelo.

Fuente: Autoras

5.6.4 Operacionalización metodológica de las variables

Tabla V XVI: Operacionalización metodológica

Variable	Categoría	Indicadores	Técnicas	Fuente de Verificación
Modelo basado en las técnica de data mining	Actividad de Investigación	• IDS tradicional • IDS con reglas de Minería de datos	Revisión de documentos	• Internet • Tutoriales
Detección del tráfico anómalo	Actividad de Investigación	Rendimiento en relación con : • Números de ataques detectados • Números de falsos positivos detectados • Tiempo de duración para la detección de intrusos	Observación directa	• Prototipo implementado • Cuarto de servidores de la facultad FIE de la ESPOCH

Fuente: Autoras

5.6.5 Comprobación de la Hipótesis

Para la comprobación de la hipótesis se ha utilizado el método estadístico T de Student con dos colas este método se aplica por existir un número de muestras menor a 30 y se realiza comparación de dos conjuntos de datos.

A continuación se va a realizar la comprobación de hipótesis con los datos obtenidos de las pruebas en la FIE.

Análisis de datos en función de los números de ataques detectados en la red de datos de la FIE.

Para realizar las pruebas en la red de datos de la FIE se ha implementado un IDS tradicional y un IDS con reglas con minería de datos como se muestran en la siguiente Tabla V XVII, se aplica los formulas del método seleccionado

Tabla V XVII: Datos de la red de datos de la FIE

Horas	No. Ataques Detectados	
	IDS con reglas de minería de datos	IDS tradicional
6:00-10:00	884319	678345
10:00-14:00	778134	593000
14:00-18:00	567352	453845
18:00-22:00	876924	524387
22:00-6:00	877645	814123

Fuente: Autoras

Formula general

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

En donde

n=5

m=5

Factor de nivel de confianza α= 5%=0.05

Donde $\bar{X}$ es la media o el promedio de los números de ataques detectados y se lo obtiene sumando todos los datos y dividiéndole para el número de observaciones que en este caso son los 5 días como se puede ver en la fórmula.

$\bar{X}$=IDS Minería de datos

$$\bar{X} = \frac{1}{n}\sum_{i-1}^{n} X_i$$

$$\bar{X} = \frac{1}{5}\,(884319 + 778134 + 567352 + 876924 + 877645)$$

$$\bar{\bar{X}} = 796874{,}8$$

Covarianzas muéstrales correspondientes

$$S_1 = \frac{1}{n}\sum_{i-1}^{n} (X_{1-}\bar{X})^2$$

$S_{1-1} = \frac{1}{4}(884319 - 796874{,}8)^2$ =1911622028

$S_{1-2} = \frac{1}{4}(778134 - 796874{,}8)^2$= 87804396,16

$S_{1-3} = \frac{1}{4}(567352 - 796874{,}8)^2$= 210719675380.16

$S_{1-4} = \frac{1}{4}(876924 - 796874{,}8)^2$= 1601968605

$S_{1-5} = \frac{1}{4}(877645 - 796874{,}8)^2$= 1630956302

$S_1 = (1911622028 + 87804396{,}16 + 210719675380.16 + 1601968605 +$

$1630956302)$= 18402530262

Donde S_1 es la covarianza muestral y se obtiene con la sumatoria de la observación menos la media elevada al cuadrado y multiplicado por el número de observaciones menos uno como se ve en la formula.

Y S_T es la sumatoria total de las S_{1-1} .

$\bar{Y}$ =IDS-Tradicional

$$\bar{Y} = \frac{1}{m}\sum_{i-1}^{n} Y_i$$

$$\bar{Y} = \frac{1}{5}\,(678345 + 593000 + 453845 + 524387 + 814123)$$

$$\bar{\bar{Y}} = 612740$$

Covarianzas muéstrales correspondientes

$$S_2 = \frac{1}{m-1}\sum_{i-1}^{n}(Y_{1-}\bar{Y})^2$$

$S_{2-1} = \frac{1}{4}(678345 - 612740)^2$ =1076004006

$S_{2-2} = \frac{1}{4}(593000 - 612740)^2$=97416900

$S_{2-3} = \frac{1}{4}(453845 - 612740)^2$=6311905256

$S_{2-4} = \frac{1}{4}(524387 - 612740)^2$=1951563152

$S_{2-5} = \frac{1}{4}(814123 - 612740)^2$=10138778172

$S_2 = (1076004006 + 97416900 + 6311905256 + 1951563152 +$

$10138778172)$= 19575667487

Aplicando la formula General

$$t = \frac{796874.8 - 612740}{\sqrt{\frac{(5-1)(18402530262)^2 + (5-1)(19575667487)^2}{5+5-2}}\sqrt{\frac{1}{5}+\frac{1}{5}}}$$

$$t = 3.74$$

GL=n+m-2

GL=5+5-2

GL=8

Donde GL son los grados de libertad que se obtiene de la suma del número de muestras en los dos casos.

Valor de la tabla de t de Student $t_{Critico} = 2.77$ por lo que Si ($t_{obtenida}$) > ($t_{critico}$) se dice que se rechaza la H_0 $3.74 > 2.77$ Se rechaza H_0 y se acepta la H_i

Hi: El modelo basado en técnicas de Data Mining permitirá la detección del tráfico anómalo de la red de datos de la Facultad de Informática y Electrónica.

Análisis de datos en función de los falsos positivos que fueron detectados en la red de datos de la FIE.

Al igual que para el número de ataques detectados se realizó con un IDS tradicional y un IDS con reglas con minería de datos para el número de falsos positivos se aplicó la misma lógica, datos que a continuación se los puede visualizar en la Tabla V XVIII.

Tabla V XVIII: Número de falsos positivos detectados en la red de datos de la FIE

Días	**No. Falsos Positivos**	
	IDS con reglas de minería de datos	**IDS Tradicional**
Día1 7:00-10:00	2054	1280
Día 2 10:00-14:00	2345	1352
Día 3 14:00-18:00	1678	1413
Día 4 18:00-22:00	1528	817
Día 5 22:00-6:00	2178	795

Fuente: Autoras

Al igual que el parámetro número de ataques detectados se hace también el cálculo respectivo con el segundo parámetro que son los falsos positivos, para lo cual se aplica las mismas fórmulas.

Formula general

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

En donde

n=5

m=5

Factor de nivel de confianza α= 5%=0.05

$\bar{X}$=IDS Minería de datos

$$\bar{X} = \frac{1}{n} \sum_{i-1}^{n} X_i$$

$$\bar{X} = \frac{1}{5} (2054 + 2345 + 1678 + 1528 + 2178)$$

$$\bar{\bar{X}} = 1956.6$$

Covarianzas muéstrales correspondientes

$$S_1 = \frac{1}{n} \sum_{i-1}^{n} (X_{1-} \bar{X})^2$$

$S_{1-1} = \frac{1}{4}(2054 - 1956.6)^2$ =2371,69

$S_{1-2} = \frac{1}{4}(2345 - 1956.6)^2$=37713,64

$S_{1-3} = \frac{1}{4}(1678 - 1956.6)^2$=19404,49

$S_{1-4} = \frac{1}{4}(1528 - 1956.6)^2$=45924,49

$S_{1-5} = \frac{1}{4}(2178 - 1956.6)^2$=12254,49

$S_1 = (2371,69 + 37713,64 + 19404,49 + 45924,49 + 12254,49)$= 117668,8

$\bar{Y}$ =IDS-Tradicional

$$\bar{Y} = \frac{1}{m}\sum_{i-1}^{n} Y_i$$

$$\bar{Y} = \frac{1}{5}\ (1280 + 1352 + 1413 + 817 + 795)$$

$$\bar{\bar{Y}} = 1131.4$$

Covarianzas muéstrales correspondientes

$$S_2 = \frac{1}{m-1}\sum_{i-1}^{n}(Y_{1-}\bar{Y})^2$$

$S_{2-1}\ = \frac{1}{4}(1280 - 1131.4)^2$ =5520,49

$S_{2\ -2} = \frac{1}{4}(1352 - 1131.4)^2$=12166,09

$S_{2-3}\ = \frac{1}{4}(1413 - 1131.4)^2$=19824,64

$S_{2-4}\ = \frac{1}{4}(817 - 1131.4)^2$=24711,84

$S_{2-5}\ = \frac{1}{4}(795 - 1131.4)^2$=28291,24

$S_2 = (18.49 + 106.09 + 249.64 + 148.84 + 331.)$= 90514,3

Aplicando la formula General

$$t = \frac{1956.6 - 1131.4}{\sqrt{\frac{(5-1)(117668.8)^2 + (5-1)(90514.3)^2}{5+5-2}}\sqrt{\frac{1}{5}+\frac{1}{5}}}$$

$$t = 4.51$$

GL=n+m-2

GL=5+5-2

GL=8

Valor de la tabla de t de Student $t_{Critico} = 2.77$ y si $(t_{obtenida}) > (t_{critico})$ se dice que se rechaza la H_0 $4.51 > 2.77$ Se rechaza H_0 y se acepta la H_i donde

Hi: El modelo basado en técnicas de Data Mining permitirá la detección del tráfico anómalo de la red de datos de la Facultad de Informática y Electrónica.

Análisis de datos en función del tiempo en la red de datos de la FIE

El parámetro tiempo se lo analizo implementando un IDS tradicional y un IDS con reglas de minería de datos, los datos obtenidos de las pruebas se observan en la Tabla V XIX.

Tabla V XIX: Tiempo en relación del IDS tradicional y un IDS con reglas de minería de datos

Horas	Tiempo	
	IDS con reglas minería de datos	IDS Tradicional
7:00-10:00	0.008	0.002
10:00-13:00	0.008	0.002
13:00-16:00	0.008	0.002
16:00-19:00	0.008	0.002
19:00-20:00	0.008	0.002
Promedio	0.008	0.002

Fuente: Autoras

A continuación se puede observar en la Figura V 75 la representación gráfica de la Tabla V XIX.

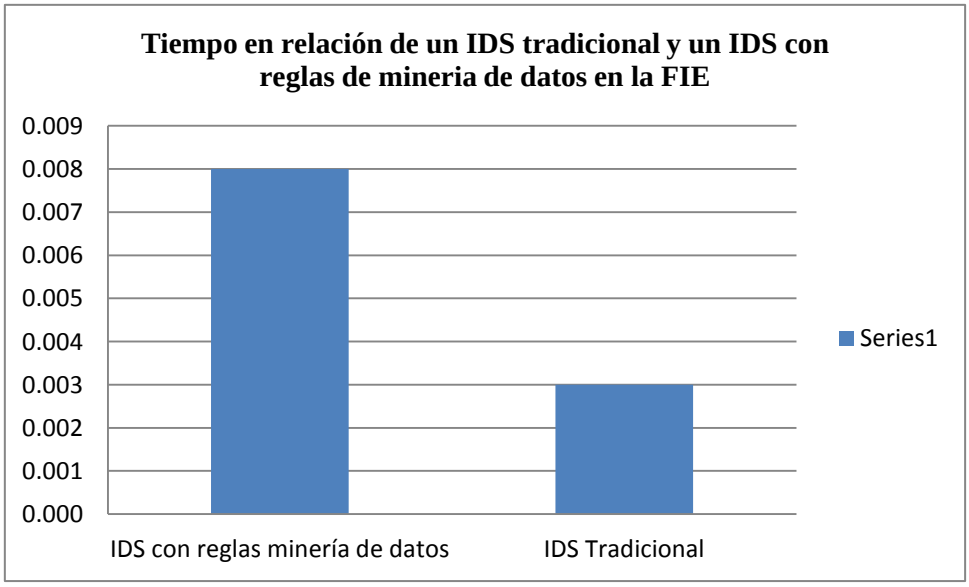


Figura V 75: Tiempo en relación del IDS tradicional y un IDS con reglas de minería de datos
Fuente: Autoras

Como se puede observar tanto en la tabla como en la gráfica el IDS tradicional lleva menos tiempo en detectar los ataques, mientras que el IDS con reglas de la minería de datos lleva un tiempo de 0.008 milisegundos.

Promedio del número de ataques en la red datos de la FIE

A continuación se puede visualizar la Tabla V XX de promedio del número de ataques detectados en la red de datos de la FIE, con un IDS tradicional y un IDS con reglas de la minería de datos.

Tabla V XX: Promedio de los números de ataques en la red de la FIE

Totales	Promedio No. Ataques Detectados FIE	
	IDS con reglas de minería de datos	IDS tradicional
Total No. detectados	3984374	3063700
Promedio	796874.8	612740

Fuente: Autoras

La representación del promedio de número de ataques detectados en la red de la FIE se lo puede observar en la Figura V 76.

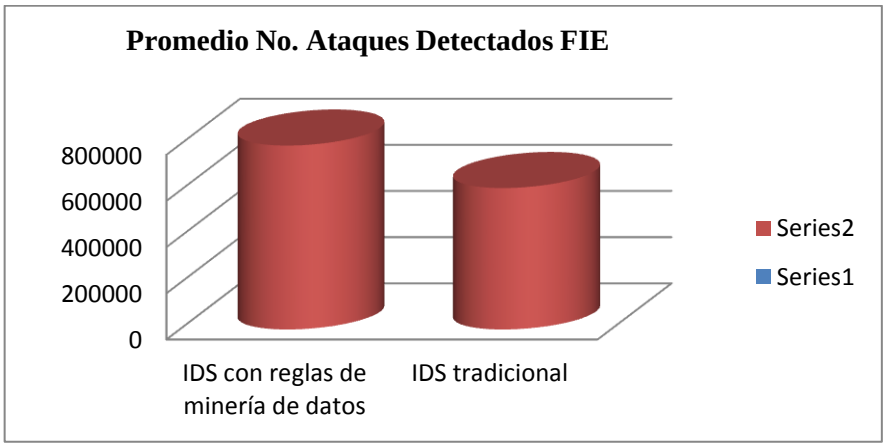


Figura V 76: Promedio de los números de ataques en la red de la FIE

Fuente: Autoras

La gráfica permite observar que un IDS con reglas de minería de datos captura mayor número de ataques en relación a un IDS tradicional.

Comprobación de la hipótesis para los datos obtenidos en las pruebas realizadas en el prototipo.

Las pruebas en el prototipo se realizaron al igual que en la red de datos de la FIE con una IDS tradicional y un IDS con reglas con minería de datos como se ve en la Tabla V XXI .

Tabla V XXI: Datos de las pruebas realizadas en el prototipo

Días	**No. Ataques Detectados**	
	IDS Con Reglas de minería de datos	**IDS Tradicional**
Día1 6:00-10:00	5040	2680
Día 2 10:00-14:00	8320	4870
Día 3 14:00-18:00	10410	7538
Día 4 18:00-22:00	6380	5243
Día 5 22:00-6:00	9520	6141

Fuente: Autoras

Formula general del método T de Student.

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

En donde

n=5 Número de muestras del IDS tradicional

m=5 Número de muestras del IDS con reglas

Factor de nivel de confianza α= 5%=0.05

Donde n y m son los números de observaciones realizadas como se dijo anteriormente en el prototipo las pruebas se las efectuó durante 5 días, tanto en el IDS tradicional como en el IDS con reglas.

$\bar{X}$=IDS-Con reglas

$$\bar{X} = \frac{1}{n}\sum_{i-1}^{n} X_i$$

$$\bar{X} = \frac{1}{5}\ (5040 + 8320 + 10410 + 6380 + 9520)$$

$$\bar{\bar{X}} = 7934$$

Donde $\bar{X}$ es la media o el promedio de los números de ataques detectados y se lo obtiene sumando todos los datos y dividiéndole para el número de observaciones que en este caso son los 5 días como se puede ver en la fórmula.

Covarianzas muéstrales correspondientes

$$S_1 = \frac{1}{n-1}\sum_{i-1}^{n}(X_{1-}\bar{X})^2$$

$S_{1-1} = \frac{1}{4}(5040 - 7934)^2$ =2093809

$S_{1-2} = \frac{1}{4}(8320 - 7934)^2$=37249

$S_{1-3} = \frac{1}{4}(10410 - 7934)^2$=1532644

$S_{1-4} = \frac{1}{4}(6380 - 7934)^2$=603729

$S_{1-5} = \frac{1}{4}(9520 - 7934)^2$=628849

$S_T = (2093809 + 37249 + 1532644 + 603729 + 628849)$=4896280

Donde S_1 es la covarianza muestral y se obtiene con la sumatoria de la observación menos la media elevada al cuadrado y multiplicado por el número de observaciones menos uno como se ve en la formula.

Y S_T es la sumatoria total de las S_{1-1} .

$\bar{Y}$ =IDS-Tradicional

$$\bar{Y} = \frac{1}{m}\sum_{i-1}^{n} Y_i$$

$$\bar{Y} = \frac{1}{5}\ (2680 + 4870 + 7538 + 5243 + 6141)$$

$$\bar{\bar{Y}} = 5294.4$$

Covarianzas muéstrales correspondientes

$$S_2 = \frac{1}{m-1}\sum_{i-1}^{n}(Y_{1-}\bar{Y})^2$$

$S_{2-1}\ = \frac{1}{4}(2680 - 5294.4)^2$ =170877184

$S_{2\ -2} = \frac{1}{4}(4870 - 5294.4)^2$=45028.84

$S_{2-3}\ = \frac{1}{4}(7538 - 5294.4)^2$=1258435.24

$S_{2-4}\ = \frac{1}{4}(5243 - 5294.4)^2$=660.49

$S_{2-5}\ = \frac{1}{4}(6141 - 5294.4)^2$=179182.89

$S_2 = (170877184 + 45028.84 + 1258435.24 + 660.49 + 179182.89)$=3192079.3

Formula General

$$t = \frac{7934 - 5494.4}{\sqrt{\frac{(5-1)(4896280)^2 + (5-1)(3192079.3)^2}{5+5-2}}\sqrt{\frac{1}{5}+\frac{1}{5}}}$$

$$t = 6.23$$

GL=n+m-2

GL=5+5-2

GL=8

Donde GL son los grados de libertas que se obtiene de la suma del número de muestras en los dos casos.

Valor de la tabla de t de Student $t_{Critico} = 2.77$ tabla que se puede observar en Anexos y su explicación se encuentra a continuación

Si ($t_{obtenida}$) > ($t_{critico}$) se dice que se rechaza la H_0, donde $6.23 > 2.77$ por lo que se rechaza H_0 y se acepta la H_i

Hi: El modelo basado en técnicas de Data Mining permitirá la detección del tráfico anómalo de la red de datos de la Facultad de Informática y Electrónica.

Análisis de datos en función de los falsos positivos que fueron detectados en el prototipo.

En la

Tabla V XXII se puede observar los falsos positivos detectados tanto en el IDS tradicional como en el IDS con reglas.

Tabla V XXII: Datos de la pruebas realizadas en el prototipo

DIAS	**No. falsos positivos**	
	IDS con reglas de minería de datos	**IDS Tradicional**
Día1 6:00-10:00	20	120
Día 2 10:00-14:00	30	132
Día 3 14:00-18:00	50	143
Día 4 18:00-22:00	25	87
Día 5 22:00-6:00	18	75

Fuente: Autoras

Al igual que el parámetro número de ataques detectados se hace también el cálculo respectivo con el segundo parámetro que son los falsos positivos, para lo cual se aplica las mismas fórmulas.

Formula general

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

En donde

n=5

m=5

Factor de nivel de confianza α= 5%=0.05

$\bar{X}$=IDS Minería de datos

$$\bar{X} = \frac{1}{n}\sum_{i-1}^{n} X_i$$

$$\bar{X} = \frac{1}{5}\ (20 + 30 + 50 + 25 + 18)$$

$$\bar{\bar{X}} = 28.26$$

Covarianzas muéstrales correspondientes

$$S_1 = \frac{1}{n}\sum_{i-1}^{n}(X_{1-}\bar{X})^2$$

$S_{1-1} = \frac{1}{4}(20 - 28.26)^2$ =17.0569

$S_{1-2} = \frac{1}{4}(30 - 28.26)^2$=0.7569

$S_{1-3} = \frac{1}{4}(50 - 28.26)^2$=118.1569

$S_{1-4} = \frac{1}{4}(25 - 28.26)^2$=2.6569

$S_{1-5} = \frac{1}{4}(18 - 28.26)^2$=26.3169

$S_1 = (S_{1-1} + S_{1-2} + S_{1-3} + S_{1-4} + S_{1-5})$=164,8

$\bar{Y}$ =IDS-Tradicional

$$\bar{Y} = \frac{1}{m}\sum_{i-1}^{n} Y_i$$

$$\bar{Y} = \frac{1}{5}\ (120 + 132 + 143 + 87 + 75)$$

$$\bar{\bar{Y}} = 111.4$$

Covarianzas muéstrales correspondientes

$$S_2 = \frac{1}{m-1}\sum_{i-1}^{n}(Y_{1-}\bar{Y})^2$$

$S_{2-1} = \frac{1}{4}(120 - 111.4)^2$ =18.49

$S_{2-2} = \frac{1}{4}(132 - 111.4)^2$=106.09

$S_{2-3} = \frac{1}{4}(143 - 111.4)^2$=249.64

$S_{2-4} = \frac{1}{4}(87 - 111.4)^2$=148.84

$S_{2-5} = \frac{1}{4}(75 - 111.4)^2$=331.24

$S_2 = (18.49 + 106.09 + 249.64 + 148.84 + 331.)$= 854,3

Aplicando la formula General

$$t = \frac{111.4 - 28.6}{\sqrt{\frac{(5-1)(854.3)^2 + (5-1)(164.8)^2}{5+5-2}}\sqrt{\frac{1}{5} + \frac{1}{5}}}$$

$$t = 8.57$$

GL=n+m-2

GL=5+5-2

GL=8

Valor de la tabla de t de Student $t_{Critico} = 2.77$

Si ($t_{obtenida}$) > ($t_{critico}$) se dice que se rechaza la H_0, y el resultado es el siguiente $8.57 > 2.77$ por lo que se rechaza H_0 y se acepta la H_i

Hi: El modelo basado en técnicas de Data Mining permitirá la detección del tráfico anómalo de la red de datos de la Facultad de Informática y Electrónica.

Análisis de datos en función del tiempo de las pruebas realizadas en el prototipo.

El parámetro tiempo se analizó con la implementación del IDS tradicional y el IDS con las reglas de minería de datos y los datos obtenidos se observa en Tabla V XXIII.

Tabla V XXIII: Tiempo en relación del IDS tradicional y un IDS con reglas de minería de datos

Horas	Tiempo	
	IDS con reglas minería de datos	IDS Tradicional
Día 1	0.003	0.001
Día 2	0.003	0.001
Día 3	0.003	0.001
Día 4	0.003	0.001
Día 5	0.003	0.001
Promedio	0.003	0.001

Fuente: Autoras

A continuación se puede observar en la Figura V 77 la representación gráfica de la Tabla V XXIII.

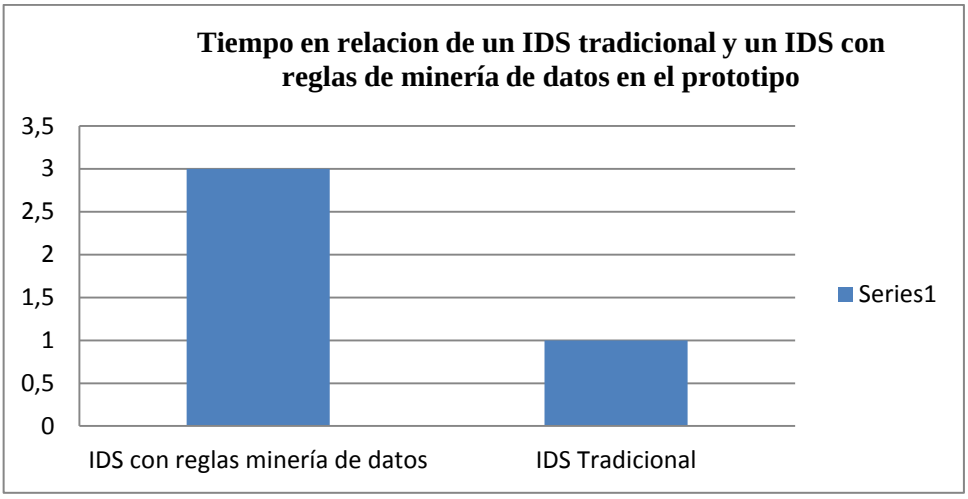


Figura V 77: Tiempo en relación del IDS tradicional y un IDS con reglas de minería de datos
Fuente: Autoras

Como se puede observar tanto en la tabla como en la gráfica el IDS tradicional lleva menos tiempo en detectar los ataques, mientras que el IDS con reglas de la minería de datos lleva un tiempo de 0.003 milisegundos.

Promedio del número de ataques en la red datos del prototipo

A continuación se ilustra el promedio del número de ataques tanto en el IDS tradicional como en el IDS con reglas con minería de datos como se puede ver en la Tabla V XXIV, dentro del prototipo.

Tabla V XXIV: Promedio de número de ataques dentro del prototipo

Totales	**Promedio No. Ataques Detectados**	
	IDS con reglas de minería de datos	**IDS tradicional**
Total No. detectados	45232	26472
Promedio	9046,4	5294,4

Fuente: Autoras

La Figura V 78 muestra el promedio entre un IDS tradicional y un IDS con reglas de minería de datos en relación al número de ataques detectados dentro del prototipo.

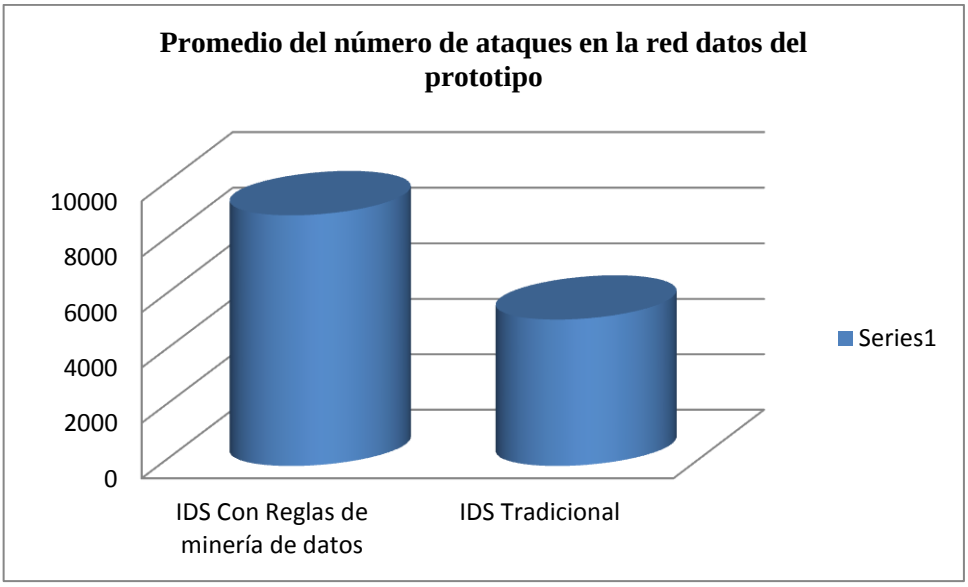


Figura V 78: Promedio de número de ataques entre un IDS tradicional y un IDS con minería de datos
Fuente: Autoras

Al igual que en la red de datos de la FIE en el prototipo, la gráfica permite observar que un IDS con reglas de minería de datos captura mayor número de ataques en relación a un IDS tradicional.

CONCLUSIONES

1. La minería de datos es un técnica muy eficiente que ayuda a encontrar patrones y relaciones ocultas con una orientación clara y potentes tecnologías para hacer búsquedas rápidas en grandes cantidades de datos y así sacar la información útil y necesaria.

2. La minería de datos es una técnica eficiente para la detección de anomalías, pero para esto se necesita un conjunto de datos suficientes para un correcto análisis.

3. La minería de datos es un proceso que ayuda a buscar y encontrar soluciones a problemas reales como lo es los ataques a las redes de información.

4. La disponibilidad de herramientas de minería de datos como weka permite automatizar gran parte de la tarea de encontrar los patrones de comportamiento ocultos en los datos.

5. El algoritmo que más reglas permitió generar es el PART para la detección de anomalías

6. La minería de datos es una técnica que permite convertir los datos en información y esta a su vez en conocimiento para la correcta toma de decisiones.

7. Los sistemas de detección de intrusos utilizan diferentes estrategias de análisis, la presente investigación se centra en aplicar el modelo implementado el mismo que se basa en las técnicas de minería de datos.

8. Es importante configurar el sistema de detección de intrusos y su integración junto con otras herramientas, para mejor su rendimiento y visualizar de mejor manera los resultados obtenidos.

RECOMENDACIONES

1. Es recomendable realizar un pre-proceso al conjunto de datos, antes de derivarlos al modelo de análisis sea cual sea el fin ya que el pre-proceso puede dar paso a transformaciones, reducciones o combinaciones de los datos.

2. Es necesario tener bien claro los objetivos perseguidos por el análisis, así como también la calidad y cantidad de datos disponibles para poder aplicar la técnica adecuada de minería de datos.

3. No todos los datos son adecuados para aplicar la minería de datos, por cual es recomendable que la búsqueda de patrones debe centrarse en aquello que sea de gran impacto y significativo dependiendo de los resultados que se esté buscando.

4. Para la implementación de este trabajo es necesario contar con un requerimiento de hardware debido a que las herramientas que se está utilizando depende de esto para un buen rendimiento.

5. Es recomendable activar solo los servicios necesarios y requeridos dentro del IDS debido a los recursos que este utiliza para su funcionamiento.

RESUMEN

Se desarrolló un modelo de detección de intrusos basado en técnicas de minería de datos para detección de ataques en redes de datos en la Facultad de Informática y Electrónica (FIE) de la Escuela Superior Politécnica de Chimborazo.

La investigación se realizó mediante método inductivo, aplicando la observación como técnica para analizar el funcionamiento del modelo desarrollado. Se utilizó herramienta de minería de datos weka y el motor detector de intrusos Suricata. Para validar el modelo se contó con un puerto spam y un puerto de administración, que fueron configurados en el switch principal para monitorear el tráfico de red en la facultad, sirviéndonos de una computadora como servidor IDS.

El modelo consiste en aplicar técnicas de minería de datos a un conjunto de datos previamente preparados para extraer conocimiento y generar reglas de detección de intrusos, las mismas que serán cargadas en la base de datos del IDS el cual analiza el tráfico de la red de datos y genera las alertas. De la implantación del modelo se obtuvo como resultado que el 86.10% de ataques fueron detectados en menos de 0.008 segundos que es equivalente a muy bueno generando alertas para la correcta administración y seguridad de la red de datos. Se concluye que el modelo desarrollado permite la detección oportuna de los ataques a una red de datos.

Se recomienda a los investigadores que para implementar el modelo de detección de intrusos se cuente con equipos de gran capacidad de almacenamiento y procesamiento, de acuerdo a la cantidad de datos a ser monitoreados.

SUMARY

It was developed an Instrusion Detection model based in data mining techniques to datect attacks in data networks in the Faculty of Informatics and Electronics (FIE) from Escuela Superior Politecnica de Chimborazo (Higher Education).

The research was carried out through the inductive method, applying the observation as a technique in order to analyze the functioning of the model develop the data mining was carried out with the help of weka software. And the Suricata which is an intrusion prevention engine, which were configured in the main switch for monitoring the network traffic in the faculty, with a computer as IDS server.

The model consists in the application of data mining techniques previously prepared to extract the knowledge and generate Intrusion Detection rules, which will be uploaded in the IDS data base which analyze the data network traffic and generate the warnings.

From the model implementation the result was obtained 86.10% of attacks were detected en less than 0.008 seconds which is equivalent to very good generating warnings for the correct administration and data network security.

It is concluded that the model developed allows the timely detection of attacks to data network.

It is recommended to the researchers implementing the Intrusion Detection model with large storage capacity and processing equipment, according to the data quantity to be monitored

GLOSARIO

Algoritmos: “Un Algoritmo, se puede definir como una secuencia de instrucciones que representan un modelo de solución para determinado tipo de problemas.”

Anomalías: “Cambio o desviación respecto de lo que es normal”.

Clasificadores Bayesianos: “En teoría de la probabilidad y minería de datos, un clasificador Bayesiano ingenuo es un clasificador probabilístico fundamentado en el teorema de Bayes y algunas hipótesis simplificadoras adicionales”.

Clustering: “Los algoritmos de segmentación (también conocidos como algoritmos de agrupamiento o, en inglés, clustering) pertenecen al grupo de métodos de mineria de datos (data mining) definido como no supervisados”.

Conocimiento: “El conocimiento es un conjunto de información almacenada mediante la experiencia o el aprendizaje (a posteriori), o a través de la introspección (a priori). En el sentido más amplio del término, se trata de la posesión de múltiples datos interrelacionados que, al ser tomados por sí solos, poseen un menor valor cualitativo

Decisión stump: “Como bien dice su nombre se trata de árboles de decisión de un solo nivel. Funcionan de forma aceptable en problemas de dos clases. No obstante, para problemas de más de dos clases es muy difícil encontrar tasas de error inferiores a 0.5”.

Falsos positivos: "En Informática es un error por el cual un software de antivirus informa que un archivo o área de sistema está infectada, cuando en realidad el objeto está limpio de virus; Esto también ocurre en los navegadores de internet cuando se descarga un archivo".

Falso negativo: En Informática es un error por el cual el software antivirus falla en detectar un archivo o área del sistema que está realmente infectada.

Intrusión: "Introducción en una propiedad, lugar, asunto o actividad sin tener derecho o autorización para ello".

Modelado: "Un modelo es por tanto una representación parcial o simplificada de la realidad que recoge aquellos aspectos de relevancia para las intenciones del modelador, y de la que se pretende extraer conclusiones de tipo predictivo. Se modela para comprender mejor o explicar mejor un proceso o unas observaciones".

Modelo: "Un modelo es una representación de un objeto, sistema o idea, de forma diferente al de la entidad misma. El propósito de los modelos es ayudarnos a explicar, entender o mejorar un sistema. Un modelo de un objeto puede ser una réplica exacta de éste o una abstracción de las propiedades dominantes del objeto
Patrones: tipo de tema de sucesos u objetos recurrentes".

Predicción: "Del latín praedictĭo, una predicción es una expresión que anticipa aquello que, supuestamente, va a suceder. Se puede predecir algo a partir de conocimientos científicos, relevaciones de algún tipo, hipótesis o indicios".

Prevención: "Del latín praeventio, prevención es la acción y efecto de prevenir (preparar con antelación lo necesario para un fin, anticiparse a una dificultad, prever un daño".

Redes neuronales: "Las redes neuronales son sistemas ideados como abstracciones de las estructuras neurobiológicas (cerebros) encontradas en la naturaleza y tienen la característica de ser sistemas desordenados capaces de guardar información".

Reglas: "Es el conjunto de medidas preventivas y reactivas de las organizaciones y de los sistemas tecnológicos que permiten resguardar y proteger la información buscando mantener la confidencialidad, la disponibilidad e integridad de la misma".

Técnicas: "La palabra técnica proviene de téchne, un vocablo de raíz griega que se ha traducido al español como "arte" o "ciencia". Esta noción sirve para describir a un tipo de acciones regidas por normas o un cierto protocolo que tiene el propósito de arribar a un resultado específico".

Virus: "Es aquel que se encarga de alterar el funcionamiento normal de un ordenador sin que el usuario lo consienta".

Weka: "(Waikato Environment for Knowledge Analysis - Entorno para Análisis del Conocimiento de la Universidad de Waikato) es una plataforma de software para aprendizaje automático y minería de datos escrito en Java y

desarrollado en la Universidad de Waikato. Weka es un software libre distribuido bajo licencia GNU-GPL".

Data mining: "Es la tecnología que permite la extracción de conocimiento de un conjunto de datos"

Ciber delincuencia: "Es la delincuencia o el cometimiento de delitos con la ayuda de la tecnología como ordenadores y redes de comunicación"

Ataques Informáticos: "Consiste en intentar hacer daño la información de un sistema"

ANEXOS

1: IMPLEMENTACIÓN Y CONFIGURACIÓN DEL PROTOTIPO

Luego procedemos a configurar el servicio para lo cual tenemos un fichero de configuración en el siguiente destino: /usr/share/doc/dhcp-3.0.5/dhcpd.conf.sample.

Lo que haremos con este fichero será copiarlo al directorio /etc y posteriormente lo personalizaremos para nuestros ajustes como se muestra en la Figura A 79

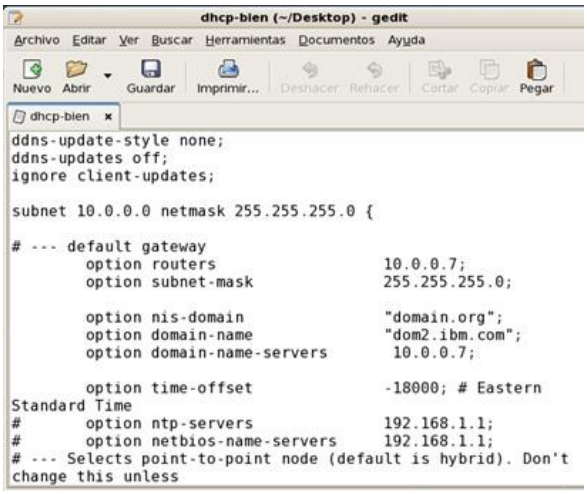


Figura A 79: Configuración del dhcp
Fuente: Autoras

Una vez que ha quedado configurado tenemos que probar y arrancar el servicio y comprobar que con la configuración establecida arranca como se puede ver en la Figura A 80.

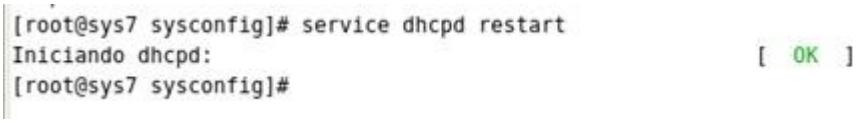


Figura A 80: Inicialización del servicio dhcp luego de la configuración
Fuente: Autoras

Ahora se pasará a comprobar que funcione. Para ello lo primero que tenemos que hacer es asegurarnos de que el servidor dhcp establecido anteriormente sea el que este Proporcionando la direcciones, una vez configurado eso y viendo que las direcciones este dentro del rango establecido se procederá a la configuración del servidor DNS.

Como primer paso agregamos las dos zonas en el fichero named.conf, el mismo que se encuentra en vi /var/named/chroot/etc/named.conf como se puede ver en la Figura A 81.

Figura A 81: Configuración del named.conf
Fuente: Autoras

Luego que agregamos las zonas procedemos crear las mismas para lo cual hacemos uso del siguiente comando vi /var/named/chroot/var/named/named.dom2.ibm.com como se puede ver en la Figura A 82.

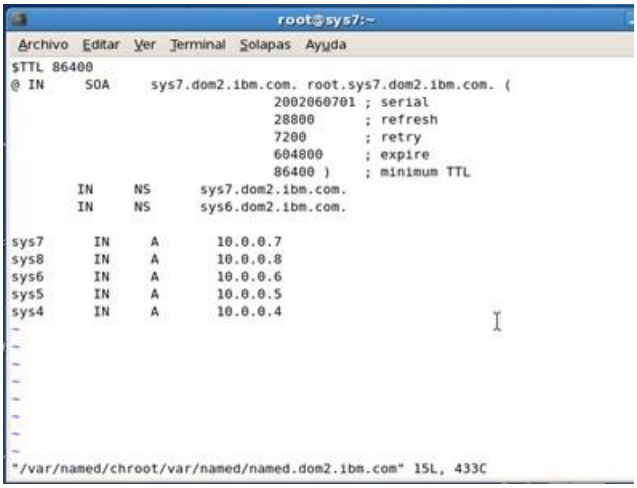


Figura A 82: Creación de la zona named.dom2.ibm.com
Fuente: Autoras

De la misma forma creamos el archivo named.10.0.0 para lo cual se hace uso del siguiente comando vi /var/named/chroot/var/named/named.10.0.0 cómo se puede observar en la Figura A 83.

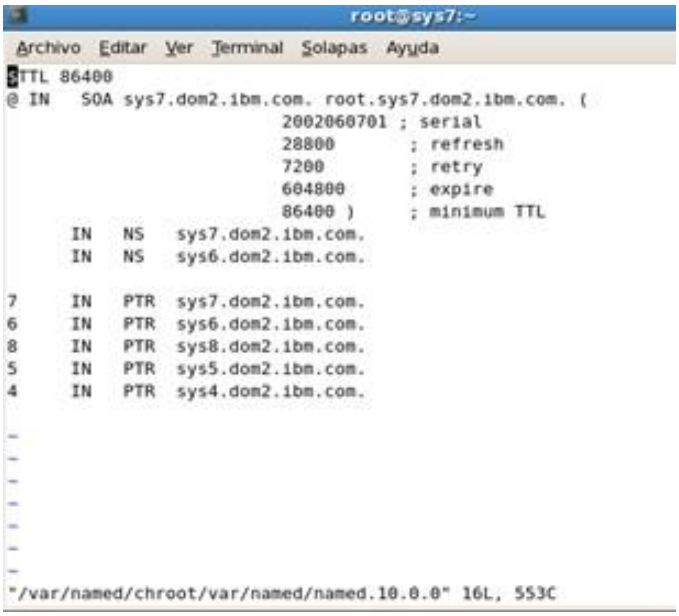


Figura A 83. Creación de la zona named.10.0.0
Fuente: Autoras

También se procede a crear el archivo localhost.zone en el siguiente directorio vi /var/named/chroot/var/named/localhost.zone como se puede ver en la Figura A 84.

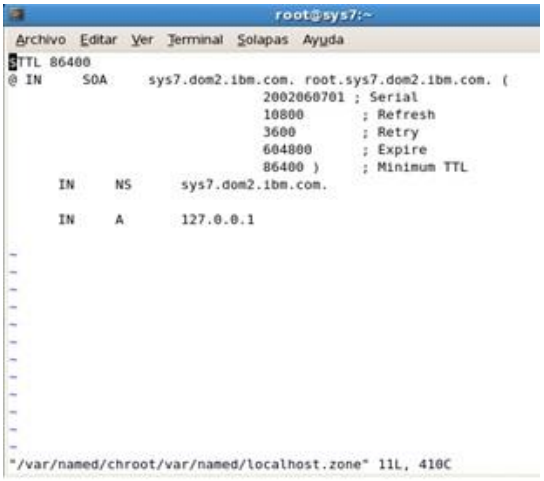


Figura A 84: Configuración del localhost.zone
Fuente: Autoras

De la misma forma creamos el archivo 127.0.0.zone en localhost haciendo uso del siguiente comando vi /var/named/chroot/var/named/127.0.0.zone como se puede ver en la Figura A 85.

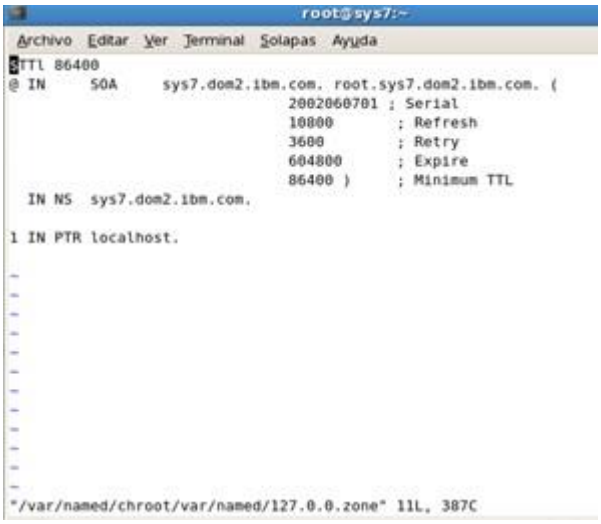


Figura A 85: Creación del archivo 127.0.0.zone en el localhost
Fuente: Autoras

Hecho esto se procede a la configuración del archivo hosts haciendo uso del siguiente comando vi /etc/hosts como se puede ver en la Figura A 16.

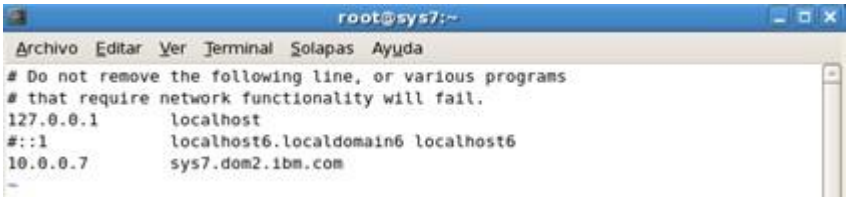


Figura A 16: Configuración del archivo Hosts
Fuente: Autoras

También se configuró el archivo network con el siguiente comando vi /etc/sysconfig/network como se puede observar en la Figura A 87.

Figura A 87: Configuración del archivo network
Fuente: Autoras

De la misma manera el archivo resolv.conf haciendo uso del siguiente comando vi /etc/resolv.conf como se puede ver en la Figura A 88.

Figura A 88: Configuración del archivo resolv.conf
Fuente: Autorast

Una vez hecho esto se procede a la inicialización del named como se puede observar en la Figura A 89.

Figura A 89: Inicialización del maned
Fuente: Autoras

CONFIGURACIÓN DEL SERVIDOR FTP

Un servidor FTP es el que va a permitir compartir datos entre diferentes servidores u ordenadores. El servicio FTP se da en la capa de aplicación del TCP/IP utilizando el puerto 20 y 21. Para una mejor descripción del servicio FTP se muestra la Figura A 90.

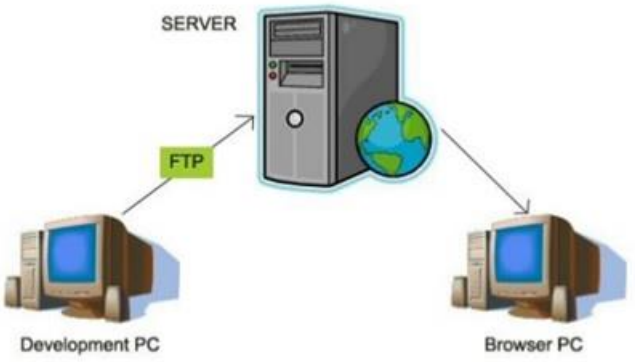


Figura A 90: Funcionamiento del servicio FTP
Fuente: Autoras

Instalamos el servicio FTP digitando el comando yum install vsftpd en la consola como se puede observar en la Figura A 91.

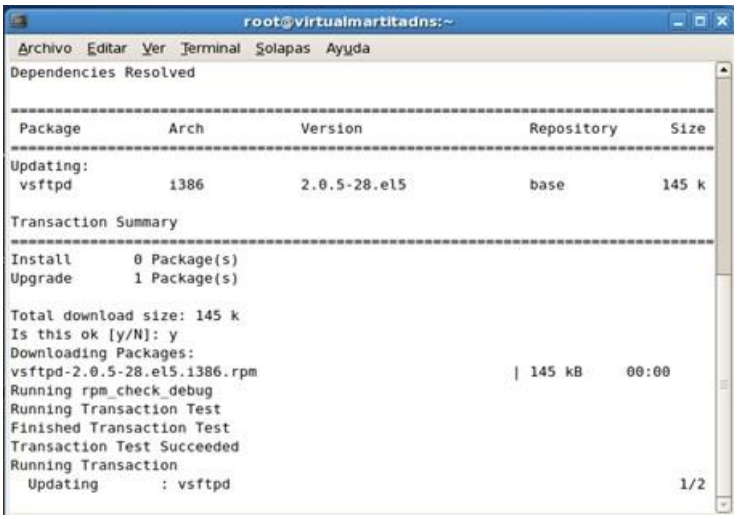


Figura A 91: Instalación del paquete FTP
Fuente: Autoras

Una vez instalado el servicio creamos un fichero en /var/ftp con la ayuda del comando mkdir y dentro de ese archivo se creará otro archivo como se muestra en la Figura A 92.

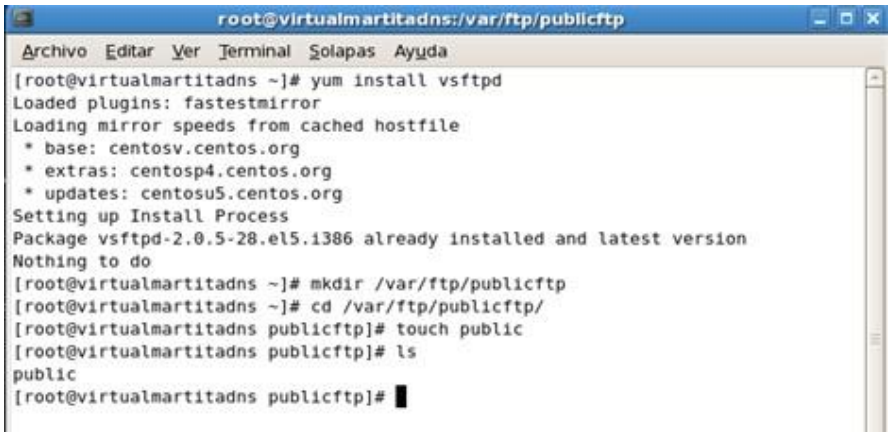


Figura A 92: Creación del fichero
Fuente: Autoras

Una vez hecho esto procedemos a configurar el archivo vsftpd.conf el mismo que se encuentra en /etc/vsftpd/ con se muestra en la Figura A 93.

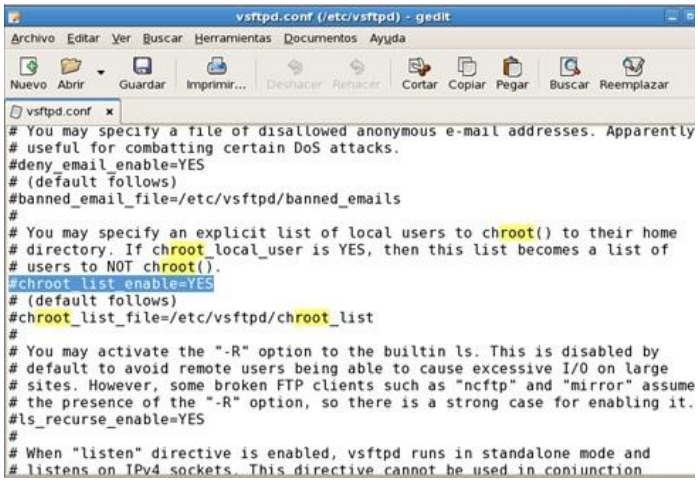


Figura A 93: Configuración del archivo vsftpd.conf
Fuente: Autoras

Al final agregamos las tres líneas de comando como se puede observar en la Figura A 94.

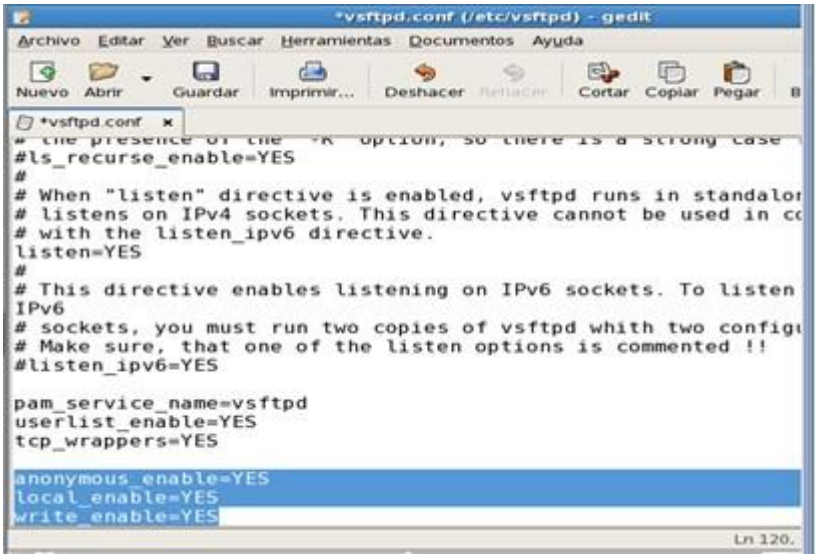


Figura A 94: Configuración del archivo vsftpd.conf
Fuente: Autoras

Lugo verificamos el funcionamiento del servicio con la ayuda del comando service vsftpd start como se muestra en la Figura A 95.

Figura A 95: Inicialización del servicio
Fuente: Autoras

Antes de verificar el funcionamiento del FTP en la web, primero se debe verificar que dirección IP tiene lo cual hacemos con la ayuda del comando ifconfig como se puede ver en la Figura A 96.

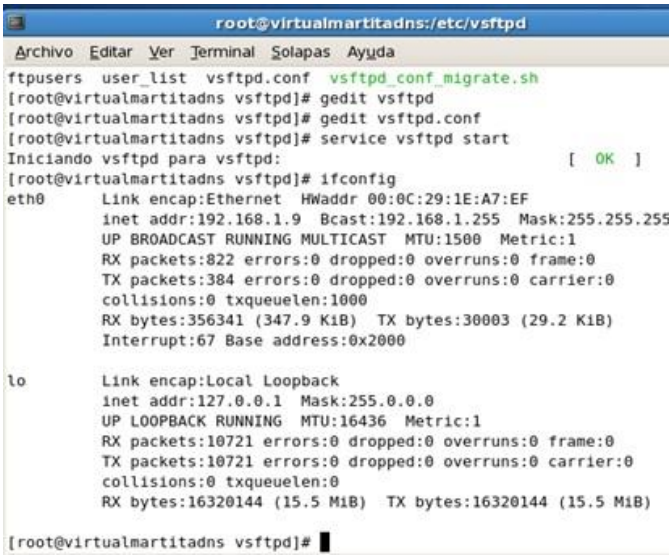


Figura A 96: verificación de la dirección IP
Fuente: Autoras

Una vez que ya se conoce la dirección IP se procede a colocarla en uno de los navegadores que se tenga disponible como se observa en la Figura A 97.

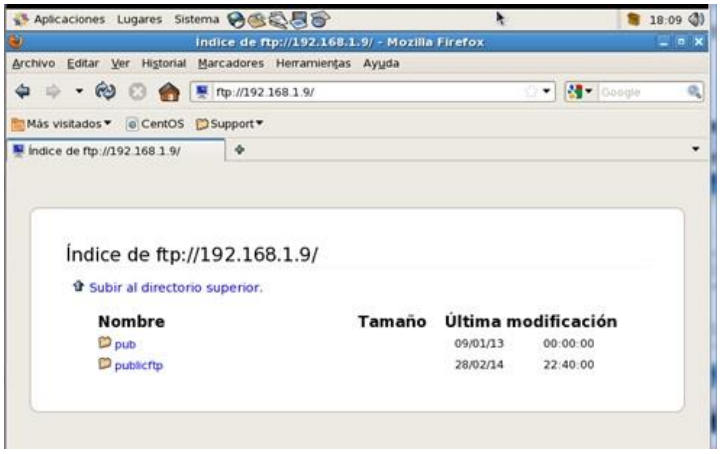


Figura A 97: Utilizando el servicio FTP
Fuente: Autoras

AUTENTIFICACION DEL FTP

Una vez configurado y comprobado que el servicio esté funcionando entramos al directorio vsftpd y procedemos a crear usuarios con su respectiva contraseña como se ve en la Figura A 98.

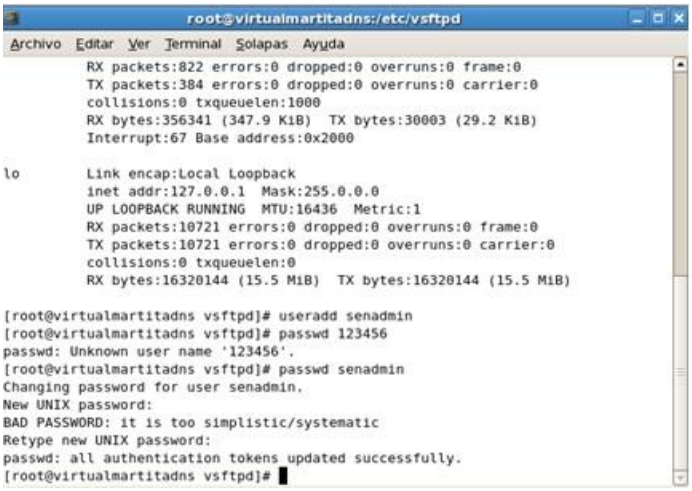


Figura A 98: Creación de usuarios
Fuente: Autoras

Para poder utilizar el servicio deshabilitamos el selinux con el comando setenforce 0 y el firewall como se puede ver en la Figura A 99.

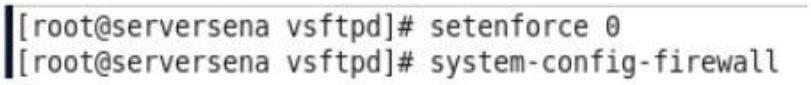


Figura A 99: Deshabilitación del selinux
Fuente: Autoras

Una vez hecho esto verificamos que el usuario y la contraseña hayan quedado correctamente para lo cual nos vamos a un navegador y entramos con el usuario seguido de la dirección IP que se asignó.

INSTALACION Y CONFIGURACION DEL SERVIDOR WEB

El servidor web es el que va a estar a la espera de peticiones realizando un trabajo como se puede ver en la Figura A 100.

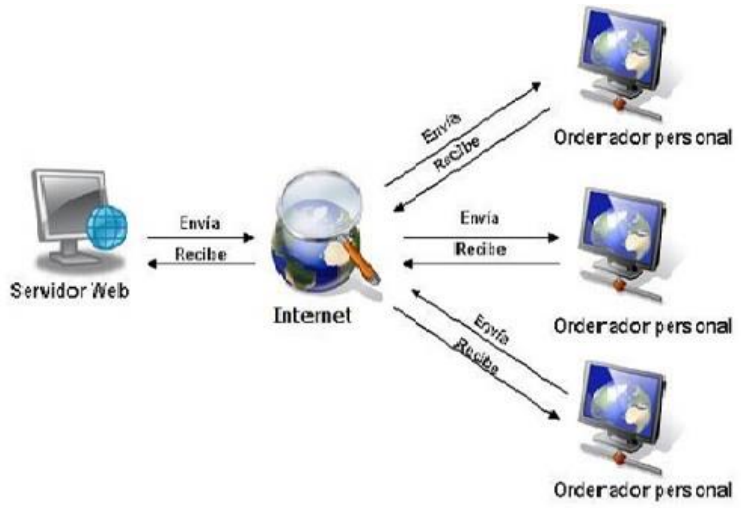


Figura A 100: Funcionamiento del servidor web
Fuente: Autoras

Para la instalación de este servidor lo primero que se debe hacer es descargarnos el paquete con la ayuda del comando yum install httpd como se puede observar en la Figura A 101.

```
root@virtualmartitadns:~
Archivo  Editar  Ver  Terminal  Solapas  Ayuda
Setting up Install Process
Resolving Dependencies
--> Running transaction check
--> Processing Dependency: httpd = 2.2.3-45.el5.centos for package: mod_ssl
--> Processing Dependency: httpd = 2.2.3-45.el5.centos for package: httpd-manual
---> Package httpd.i386 0:2.2.3-83.el5.centos set to be updated
--> Running transaction check
---> Package httpd-manual.i386 0:2.2.3-83.el5.centos set to be updated
---> Package mod_ssl.i386 1:2.2.3-83.el5.centos set to be updated
--> Finished Dependency Resolution

Dependencies Resolved

================================================================================
 Package                Arch           Version                    Repository      Size
================================================================================
Updating:
 httpd                  i386           2.2.3-83.el5.centos        updates        1.2 M
Updating for dependencies:
 httpd-manual           i386           2.2.3-83.el5.centos        updates        819 k
 mod_ssl                i386           1:2.2.3-83.el5.centos      updates         97 k

Transaction Summary
================================================================================
Install       0 Package(s)
Upgrade       3 Package(s)

Total download size: 2.1 M
Is this ok [y/N]:
```

Figura A 101: Instalación del paquete de correo
Fuente: Autoras

Como segundo paso comprobamos que la descarga haya sido exitosa con la ayuda del comando rpm –q httpd hecho esto procedemos a editar el archivo http.conf el mismo que se encuentra en el siguiente directorio /etc/httpd/conf/http.conf dentro de este archivo buscamos NameVirtualHosts y lo descomentamos como se observa en la Figura A 102.

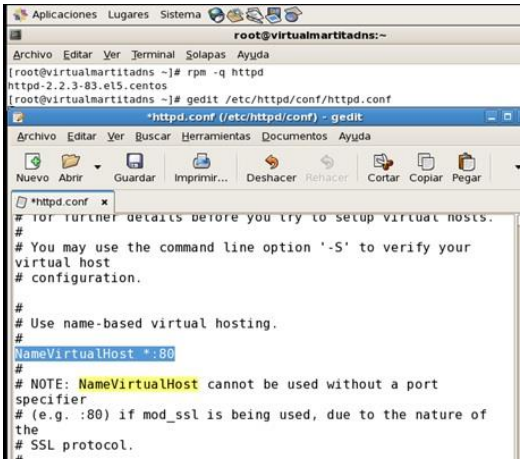


Figura A 102: Configuración del archivo
Fuente: Autoras

Luego se procede a crear los siguientes ficheros y directorios para modificar el servidor web que se va a montar como se puede ver en la Figura A 103.

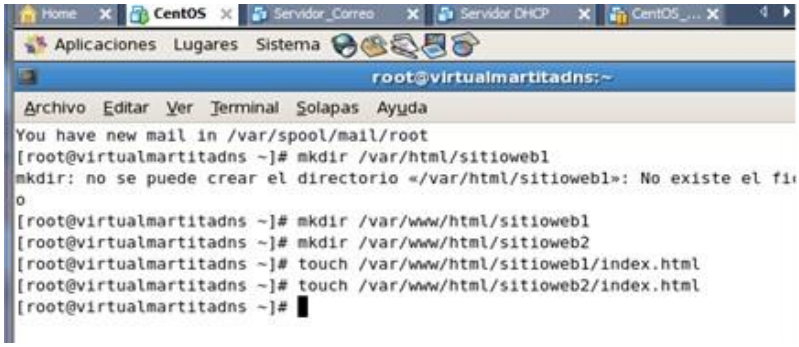


Figura A 103: Creación de ficheros
Fuente: Autoras

Junto con la ayuda de texto procedemos a editar las páginas web que se creó anteriormente como se puede observar en la Figura A 104.

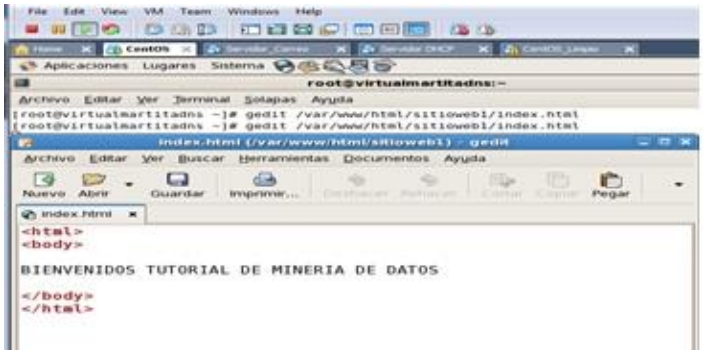


Figura A 104: Creación de páginas web
Fuente: Autoras

Lo mismo de hace con el sitio dos como se puede observar en la Figura A 105.

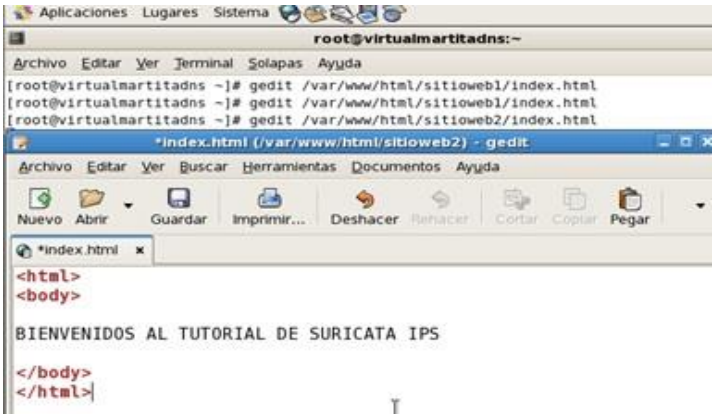


Figura A 105: Creación de la página web 2
Fuente: Autoras

Luego creamos y editamos un fichero el mismo que se encuentra en el directorio /etc/httpd/conf.d/tesis.conf como se ve en la Figura A 106.

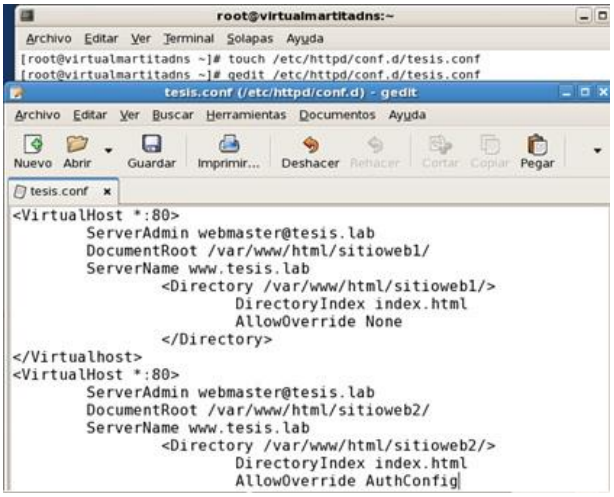


Figura A 106: Creación y configuración de directorios
Fuente: Autoras

Por último se configura el archivo zone.dl el mismo que se encuentra en el siguiente directorio /var/nemd/zone.dl luego iniciamos el servicio con la ayuda del comando service httpd start como se puede observar en la Figura A 107.

Figura A 107: Configuración del archivo zone.dl
Fuente: Autoras

Al final verificamos como se puede observar en la Figura A 107.

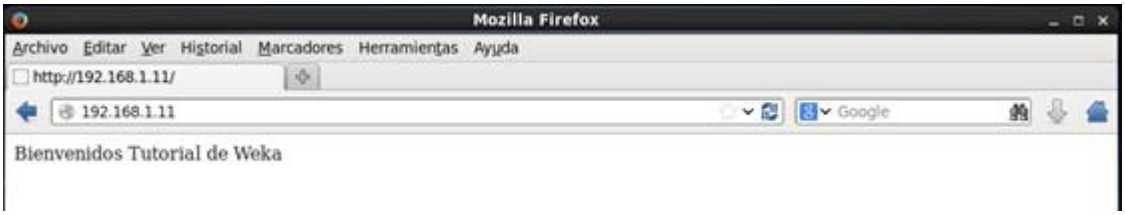


Figura A 108: Verificación del servicio http
Fuente: Autoras

INSTALACION Y CONFIGURACION DEL SERVIDOR SSH

Este servicio es muy similar al servicio de telnet con la diferencia que el servicio de SSH que la comunicación entre el servidor y el cliente viajan cifrados desde el primer momento brindando así mayor seguridad a la información.

Para su instalación ingresamos como root para poder realizar la instalación del servicio con la ayuda del comando yum install openssh-server como se puede observar en la Figura A 109.

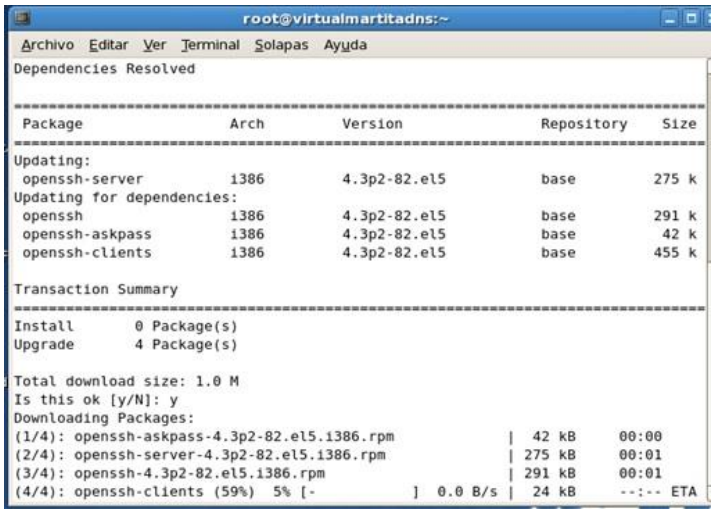


Figura A 109: Instalación del paquete ssh
Fuente: Autoras

Una vez que se ha terminado la instalación del servicio, se procede a levantar el servicio con la ayuda del comando chkconfig sshd on hecho esto lo reiniciamos mediante el comando service sshd restart como se puede observar en la Figura A 110.

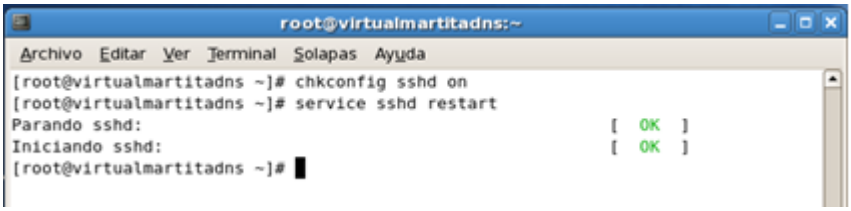


Figura A 110: Inicialización del servicio ssh
Fuente: Autoras

CONFIGURACION DEL SERVIDOR DE CORREO

El servidor de correo a utilizar será el Sendmail es uno de los agentes más populares de transporte de correo.

Para la instalación del servicio de correo se instaló el postfix mediante el comando yum install posstfix como se ve en la Figura A 111.

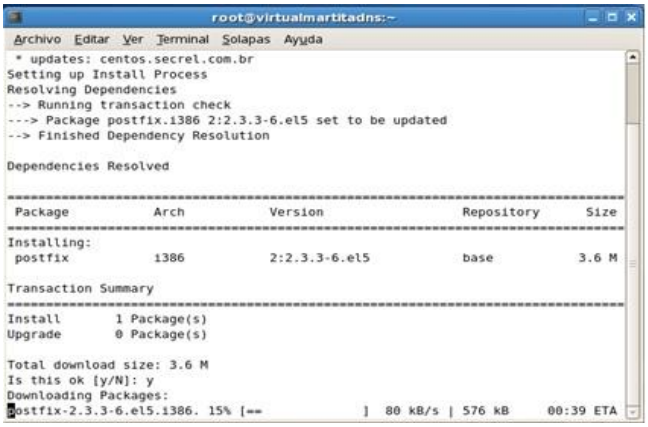


Figura A 111: Instalación del paquete postfix
Fuente: Autoras

Hecho esto se procede a configurar el archivo main.cf el mismo que se encuentra en el siguiente directorio /etc/postfix/main.cf como se puede ver en la Figura A 112.

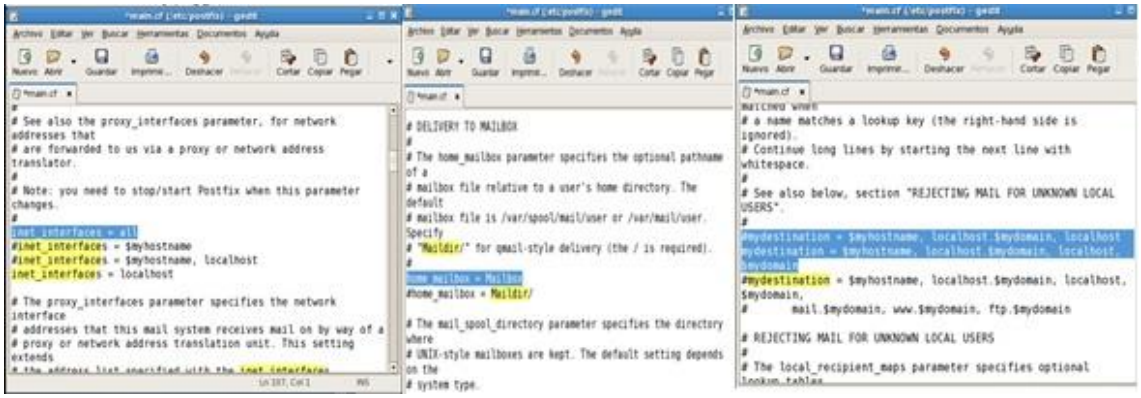


Figura A 112: Configuración del archivo main.cf
Fuente: Autoras

Luego se estos cambios realizados se procede a iniciar el servicio mediante el comando service postfix start como se observa en la Figura A 113.

Figura A 113: Inicialización del servicio postfix
Fuente: Autoras

Hecho esto procedemos a crear usuarios con su respectiva contraseña con la ayuda del comando useradd nombre del usuario y passwd nombre usuario y su respectiva contraseña, hecho esto se procede a la instalación de un servidor IMAP y POP3 que en este caso será DOVECOT como se puede ver en la Figura A 114.

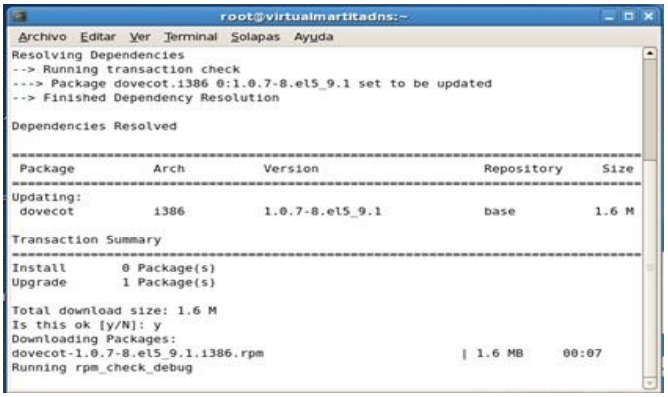


Figura A 114: Instalación del paquete Dovecot
Fuente: Autoras

Una vez que se tenga instalado el servidor se procede a su respectiva configuración el archivo dovecot.conf el mismo que se encuentra en el siguiente directorio /etc/dovecot.conf como se puede observar en la Figura A 115.

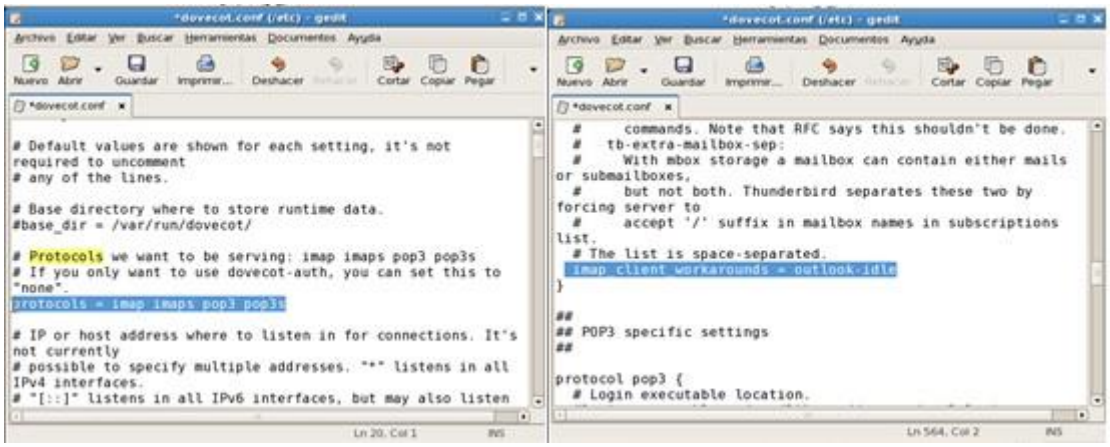


Figura A 115: Configuración del archivo dovecot.conf
Fuente: Autoras

Terminado las configuraciones respectivas se procede a iniciar el servicio para lo cual digitamos el comando service dovecot start.

Hecho esto se procede a la instalación del servidor http mediante el comando yum install hhtpd y de una vez a iniciar el servicio con service httpd start como se puede observar en la Figura A 116.

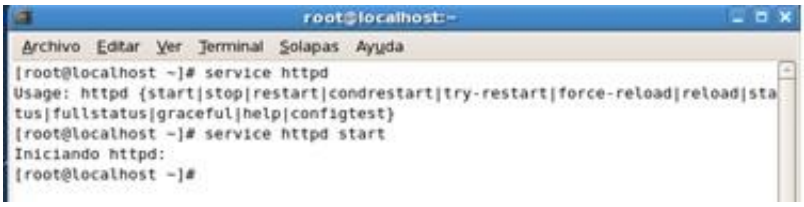


Figura A 116: Instalación del servidor http
Fuente: Autoras

Por ultimo instalamos la aplicación webmail el mismo que nos ayudará a revisar nuestros correos, en nuestro caso se instaló Squirrelmail como se puede ver en la Figura A 117.

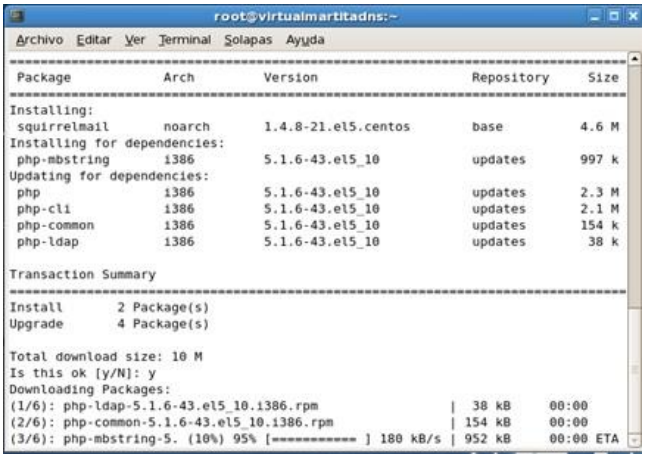


Figura A 117: Instalación de la aplicación webmail
Fuente: Autoras

Al igual que los demás servicios una vez instalado procedemos a configurarlo para lo cual buscamos el archivo de configuración que se encuentra en el siguiente directorio /usr/share/squirrelmail/config/conf.pl como se puede ver en la Figura A 118.

Figura A 118: Configuración de paquetes
Fuente: Autoras

Una vez configurado iniciamos todos los servicios como el http, postfix, dovecot y procedemos a usarlo como se puede ver en la Figura A 119.

Figura A 119: Inicialización del servicio
Fuente: Autoras

ANEXOS 2: Instalación del backtrack

Seleccionamos el tipo de teclado Figura A 120 y presionamos adelante.

Figura A 120: Opción de teclado
Fuente: Autoras

Aparecerá una pantalla detallando la ubicación en donde se instalara Figura A 121 presionamos adelante.

Figura A 121: Espacio de Instalación
Fuente: Autoras

Aparecerá una pantalla con el detalle de la instalación presionamos instalar como se puede observar en la Figura A 122.

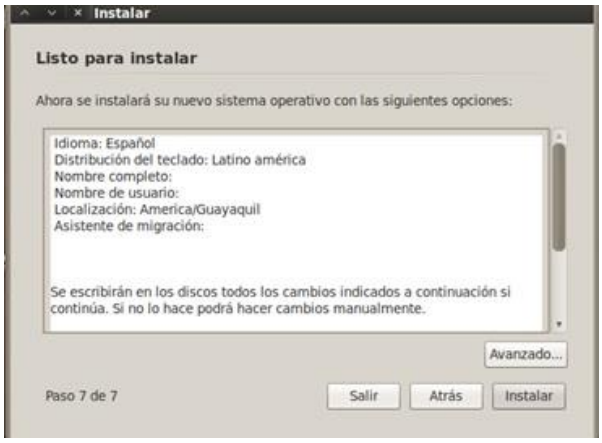
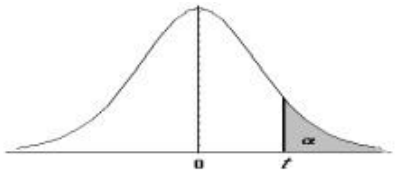


Figura A 122: Instalación
Fuente: Autoras

TABLA T DE STUDET

Tabla de la t de Student.

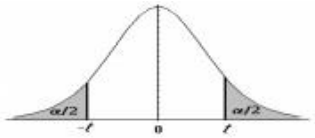
Contiene los valores t tales que $p[T > t] = \alpha$,
donde n son los grados de libertad.



$n \setminus \alpha$	0,30	0,25	0,20	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	0,7265	1,0000	1,3764	3,0777	6,3137	12,7062	31,8210	63,6559	127,3213	318,3088	636,6192
2	0,6172	0,8165	1,0607	1,8856	2,9200	4,3027	6,9645	9,9250	14,0890	22,3271	31,5991
3	0,5844	0,7649	0,9785	1,6377	2,3534	3,1824	4,5407	5,8408	7,4533	10,2145	12,9240
4	0,5686	0,7407	0,9410	1,5332	2,1318	2,7765	3,7469	4,6041	5,5976	7,1732	8,6103
5	0,5594	0,7267	0,9195	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	5,8934	6,8688
6	0,5534	0,7176	0,9057	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2076	5,9588
7	0,5491	0,7111	0,8960	1,4149	1,8946	2,3646	2,9979	3,4995	4,0293	4,7853	5,4079
8	0,5459	0,7064	0,8889	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	4,5008	5,0413
9	0,5435	0,7027	0,8834	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,2968	4,7809
10	0,5415	0,6998	0,8791	1,3722	1,8125	2,2281	2,7638	3,1693	3,5814	4,1437	4,5869
11	0,5399	0,6974	0,8755	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0247	4,4370
12	0,5386	0,6955	0,8726	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	3,9296	4,3178
13	0,5375	0,6938	0,8702	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	3,8520	4,2208
14	0,5366	0,6924	0,8681	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	3,7874	4,1405
15	0,5357	0,6912	0,8662	1,3406	1,7531	2,1315	2,6025	2,9467	3,2860	3,7328	4,0728
16	0,5350	0,6901	0,8647	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6862	4,0150
17	0,5344	0,6892	0,8633	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,6458	3,9651
18	0,5338	0,6884	0,8620	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,6105	3,9216
19	0,5333	0,6876	0,8610	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,5794	3,8834
20	0,5329	0,6870	0,8600	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,5518	3,8495
21	0,5325	0,6864	0,8591	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,5272	3,8193
22	0,5321	0,6858	0,8583	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,5050	3,7921
23	0,5317	0,6853	0,8575	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,4850	3,7676
24	0,5314	0,6848	0,8569	1,3178	1,7109	2,0639	2,4922	2,7970	3,0905	3,4668	3,7454
25	0,5312	0,6844	0,8562	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,4502	3,7251
26	0,5309	0,6840	0,8557	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,4350	3,7066
27	0,5306	0,6837	0,8551	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,4210	3,6896
28	0,5304	0,6834	0,8546	1,3125	1,7011	2,0484	2,4671	2,7633	3,0469	3,4082	3,6739
29	0,5302	0,6830	0,8542	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,3962	3,6594
30	0,5300	0,6828	0,8538	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,3852	3,6460
40	0,5286	0,6807	0,8507	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,3069	3,5510
80	0,5265	0,6776	0,8461	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,1953	3,4163
120	0,5258	0,6765	0,8446	1,2886	1,6576	1,9799	2,3578	2,6174	2,8599	3,1595	3,3735
∞	0,5244	0, 6745	0,8416	1,2816	1,6449	1,9600	2,3263	2,5758	2,8070	3,0902	3,2905

Tabla de la t de Student.

Contiene los valores t tales que $p[|T| > t] = \alpha$,
donde n son los grados de libertad.



n \ α	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,1584	0,3249	0,5095	1,0000	1,9626	3,0777	6,3137	12,7062	31,8210	63,6559	636,5776
2	0,1421	0,2887	0,4447	0,8165	1,3862	1,8856	2,9200	4,3027	6,9645	9,9250	31,5998
3	0,1366	0,2767	0,4242	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8408	12,9244
4	0,1338	0,2707	0,4142	0,7407	1,1896	1,5332	2,1318	2,7765	3,7469	4,6041	8,6101
5	0,1322	0,2672	0,4082	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8685
6	0,1311	0,2648	0,4043	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9587
7	0,1303	0,2632	0,4015	0,7111	1,1192	1,4149	1,8946	2,3646	2,9979	3,4995	5,4081
8	0,1297	0,2619	0,3995	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0414
9	0,1293	0,2610	0,3979	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,1289	0,2602	0,3966	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5868
11	0,1286	0,2596	0,3956	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4369
12	0,1283	0,2590	0,3947	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178
13	0,1281	0,2586	0,3940	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	4,2209
14	0,1280	0,2582	0,3933	0,6924	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	4,1403
15	0,1278	0,2579	0,3928	0,6912	1,0735	1,3406	1,7531	2,1315	2,6025	2,9467	4,0728
16	0,1277	0,2576	0,3923	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	4,0149
17	0,1276	0,2573	0,3919	0,6892	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,9651
18	0,1274	0,2571	0,3915	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,9217
19	0,1274	0,2569	0,3912	0,6876	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,8833
20	0,1273	0,2567	0,3909	0,6870	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,8496
21	0,1272	0,2566	0,3906	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,8193
22	0,1271	0,2564	0,3904	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,7922
23	0,1271	0,2563	0,3902	0,6853	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,7676
24	0,1270	0,2562	0,3900	0,6848	1,0593	1,3178	1,7109	2,0639	2,4922	2,7970	3,7454
25	0,1269	0,2561	0,3898	0,6844	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,7251
26	0,1269	0,2560	0,3896	0,6840	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,7067
27	0,1268	0,2559	0,3894	0,6837	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,6895
28	0,1268	0,2558	0,3893	0,6834	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,6739
29	0,1268	0,2557	0,3892	0,6830	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,6595
30	0,1267	0,2556	0,3890	0,6828	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,6460
40	0,1265	0,2550	0,3881	0,6807	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	3,5510
80	0,1261	0,2542	0,3867	0,6776	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	3,4164
120	0,1259	0,2539	0,3862	0,6765	1,0409	1,2886	1,6576	1,9799	2,3578	2,6174	3,3734
∞	0,126	0,253	0,385	0,674	1,036	1,282	1,645	1,96	2,326	2,576	3,291

BIBLIOGRAFÍA

[1] IMPORTANCIA DE LOS DATOS

http://web.usbmed.edu.co/usbmed/fing/v3n1/v3n1a6.pdf.ç

2012-04-25

[2] IMPORTANCIA DE LOS DATOS

http://dspace.espoch.edu.ec/bitstream/123456789/1933/1/38T00286.pdf

2012-03-23

[3] IMPORTANCIA DE LOS DATOS

http://sicodinet.unileon.es/dpi2011-0105/

2011-05-01

[4] EVOLUCIÓN DEL INTERNET

http://openaccess.uoc.edu/webapps/o2/bitstream/10609/8167/1/dcampoym_TFM_0611.pdf

2014-06-23

[5] ENCUESTAS INEC

http://www.hoy.com.ec/noticias-ecuador/las-cifras-del-uso-de-internet-crecen-en-el-pais-535452.html

2012-02-20

[6] ANTECEDENTES DE UN IPS

http://www.dspace.espol.edu.ec/bitstream/123456789/19502/2/tesina_seminario0.6.pdf

2013-08-12

[7] ANTECEDENTES DE UNA AMENAZA

http://www.cisco.com/web/ES/solutions/es/internet_security/index.html

2014-04-17

[8] ANTECEDENTES DE LA MINERÍA DE DATOS

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosYany2008.pdf

2014-06-19

[9] ANTECEDENTES DE LA MINERÍA DE DATOS

http://admondelconocimiento-edgarguardado.wikispaces.com/3.2+Metodos+estadisticos+aprendizaje+mineria+de+datos

2014-02-20

[10] PREVENCIÓN DE MOVIMIENTOS

http://www.adminso.es/recursos/Proyectos/PFC/PFC_marisa.pdf

2014-06-30

[11] ANTECEDENTES DE LA MINERÍA DE DATOS

http://www.it.uc3m.es/jvillena/irc/practicas/06-07/22.pdf

2012-05-12

[12]DEFINICIÓN DE MINERÍA DE DATOS SEGÚN OOCITES

http://www.oocities.org/es/mineria.datos/definicion_tecnicas_mineria_datos.pdf

2014-06.30

[13] DEFINICIÓN DE MINERÍA DE DATOS SEGÚN MICROSOFT

http://msdn.microsoft.com/es-es/library/ms174949.aspx

2014-05-07

[14] DEFINICIÓN DE MINERÍA DE DATOS SEGÚN MICROSOFT

http://msdn.microsoft.com/es-es/library/ms174949(v=sql.90).aspx

2014-06-19

[15] DEFINICIÓN DE MINERÍA DE DATOS SEGÚN SINNEXUS

http://www.sinnexus.com/business_intelligence/datamining.aspx

2012-05-25

[16] PROCESO DE DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS

http://www.dspace.espol.edu.ec/bitstream/123456789/19502/2/tesina_seminario0.6.pdf

2014-07-02

[17] OBJETIVO DEL KKD

http://www.dataprix.com/13-descubrimiento-conocimiento-bases-datos-kdd

2011-06-20

[18] PROCESO DE EXTRACCIÓN DE CONOCIMIENTO

http://www.webmining.cl/2011/01/proceso-de-extraccion-de-conocimiento/

2011-01

[19] TÉCNICAS DE MODELADO DE MINERÍA DE DATOS

http://santiagozapatakdd.files.wordpress.com/2011/03/curso-kdd-full-cap-3.pdf

2011-03

[20] TÉCNICAS SEGÚN VALLEJO Y TENELANDA

http://web.usbmed.edu.co/usbmed/fing/v3n1/v3n1a6.pdf

2014-02-07

[21] TÉCNICAS SEGÚN TIRNAUCA

http://personales.unican.es/tirnaucac/Slides/121030%20Teoria.pdf

2014-02-07

[22] TÉCNICAS SEGÚN MICROSOFT

http://msdn.microsoft.com/es-ec/library/ms175595.aspx

2014-02-07

[23] ALGORITMOS DE WEKA

http://www.it.uc3m.es/jvillena/irc/practicas/07-08/IntrusionesDeRed.pdf

2014-02-07

[24] ALGORITMOS DE WEKA

http://mydatamine.com/tag/induction-algorithms/

2014-02-07

[25] ALGORTIMO BAYESNET

http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/BayesNet.html

2014-02-07

[26] ALGORTIMO NAIVE BAYES

https://www.princeton.edu/~achaney/tmve/wiki100k/docs/Naive_Bayes_classifier.html

2014-02-07

[27] ANTECEDENTES DE LAS HERRAMIENTAS DE MINERÍA DE DATOS

http://blog.jmacoe.com/gestion_ti/base_de_datos/5-mejores-software-mineria-datos-codigo-libre-abierto/

2014-06-19

[28] HERRAMIENTA ORANGE

http://www.softpedia.es/programa-Orange-209910.html.

2014-06-19

[29] HERRAMIENTA DE MINERÍA DE DATOS

http://fraterneo.blogspot.com/2010/11/5-programas-libres-para-data-mining.html

2014-07-02

[30] INTRODUCCIÓN DE ATAQUES A RED DE INFORMACIÓN

http://es.kioskea.net/contents/622-introduccion-a-la-seguridad-informatica

2014-06-21

[31] ANTECEDENTES A LOS ATAQUES A LA RED DE INFORMACIÓN

http://cxo-community.com/articulos/estadisticas/81-seguridad-informatica/5875--informe-de-seguridad-anual-2014-de-cisco-crecieron-los-ataques-avanzados-y-el-trafico-malicioso.html

2010-02-18

[32] DEFINICIONES DE ATAQUES INFORMÁTICOS

http://www.ecured.cu/index.php/Ataque_inform%C3%A1tico

2014-06-20

[33] DEFINICIÓN DE ATAQUES SEGÚN EVILFINGERS

https://www.evilfingers.com/publications/white_AR/01_Ataques_informaticos.pdf

2013-09-29

[34] TIPOS DE ATAQUES A LAS REDES DE INFORMACIÓN

http://searchdatacenter.techtarget.com/es/definicion/Ataque-de-denegacion-de-servicio

2014-06-23

[35] ATAQUE R2L, U2R Y MONITORIZACIÓN

http://www.uelbosque.edu.co/sites/default/files/publicaciones/revistas/revista_tecnologia/volumen11_numero1/aplicacion_programacion_multiobjetivo11-1.pdf

2012-01

[36] ATAQUE SHOULDER SURFING

http://www.rediris.es/cert/doc/unixsec/node8.html

2014-06-30

[37] ATAQUE DECOY, SCANNING Y SNOOPING

http://www.segu-info.com.ar/ataques/ataques_monitorizacion.htm

2014-06-30

[38] DESCRIPCIÓN DE LOS ATAQUES

http://www.alegsa.com.ar/Dic/ataque%20informatico.php

2013-07-18

[39] INDICIOS DE UNA INTRUSIÓN

http://zystrax.wordpress.com/tag/auditoria-de-seguridad/

2014-06-20

[40] SISTEMAS DE DETECCIÓN DE INTRUSOS

http://es.kioskea.net/contents/162-sistema-de-deteccion-de-intrusiones-ids

2014-06-30

[41] SISTEMAS DE PREVENCIÓN DE INTRUSOS, MODOS DE DETECCIÓN Y DIFERENCIAS ENTRE IPS E IDS Y MOTOR DE DETECCIÓN SURICATA

http://www.dspace.espol.edu.ec/bitstream/123456789/19502/2/tesina_seminario0.6.pdf

2014-06-30

[42] MOTOR DE DETECCIÓN SNORT

http://www.buenastareas.com/ensayos/Seguridad-De-La-Red-Snort/44040655.html

2014-07-02

[43] MOTOR DE DETECCIÓN SNORT

https://seguridadinformaticaufps.wikispaces.com/file/view/honeypot%20-%20nids.pdf

2014-07-01

[44] TRABAJO DE LIU WENJUN

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5565749

2014-05-12

[45] HERRAMIENTA WEKA

http://www.uaeh.edu.mx/scige/boletin/huejutla/n2/a1.html

2014-07-01

[46] CARACTERISTICAS WEKA

http://cor-mineriadedatos.blogspot.com/2011/06/weka.html

2011-06

[47] CARACTERISTICAS WEKA

http://mineriadedatos.wikispaces.com/WEKA

2014-05-22

[48] PREPARACIÓN DEL ARCHIVO ADECUADO PARA LA HERRAMIENTA WEKA

http://www.metaemotion.com/diego.garcia.morate/download/weka.pdf

2014-07-01

[49] ESTADO DE LAS REDES INFORMATICAS

http://www.dtic.ua.es/grupoM/recursos/articulos/JDARE-09-B.pdf

2014-07-01

[50] ALGORITMOS

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SDataMining.pdf

2014-06-12

[51] REQUERIMIENTOS PARA LA INSTALACIÓN DEL SURICATA

http://cyruslab.net/2012/10/18/building-an-ids-part-1-installing-pre-requisites-and-snorby/

2014-05-26

[52] BASE DE DATOS MYSQL

http://ocw.uoc.edu/computer-science-technology-and-multimedia/bases-de-datos/bases-de-datos/P06_M2109_02151.pdf

2104-05-26

[53] PROCESO KDD

http://www.repositoriodigital.ipn.mx/bitstream/handle/123456789/11011/2011%20XVII%20STA-29.pdf?sequence=1

2014-07-02

[54] FORMATO DEL ARCHIVO ARFF

http://inteligenciartificialudec.wikispaces.com/file/view/TRABAJO%20DE%20INT%20ARTFICIAL%20WEKA.pdf

2014-07-01

[55] ALGORITMOS BASADOS EN REGLAS

http://www.dsic.upv.es/~cferri/weka/].

2014-06-06

[56] SNORBY

https://seguridadyredes.wordpress.com/2010/12/01/snort-snorby-un-front-end-para-analisis-y-gestian-de-alertas-para-snort/

2010-12-01

[57] DEFINICIÓN DE SERVIDORES

http://www.aprenderaprogramar.com/index.php?option=com_attachments&task=download&id=487

2014-07-02

[58] DEFINICIÓN DE UN SERVIDOR DNS

http://www.turismotour.com/%C2%BFque-son-los-servidores-dns/

2014-07-02

[59] DEFINICIÓN DE BACKTRACK

http://www.backtrack-linux.org/forums/showthread.php?t=24468

2014-07-02

[60] ATAQUE ARP-SPOOFING

https://www.evilfingers.com/publications/white_AR/01_Ataques_informaticos.pdf

2014-06-22