Universidad de Antioquia

Reporte 3

Metricas - Parte 1

Author:
Henry Arcila

Supervisors:
Prof. Natalia Gaviria Prof.
Danny Múnera

23 de septiembre de 2018

Índice

Resumen

De a cuerdo al World internet usage and poblation statistics, aproximadamente un 54.4 % tienen acceso a internet [3]. Como el recurso por excelencia intercambiado a traves de internet es la información este debe ser protegido; sin embargo, dicha tarea es cada vez mas desafiante debido a la mayor facilidad, numero y sofisticación de los ataques actualmente existentes. Para hacer frente éstos se han creado diferentes sistemas de seguridad como firewalls, antivirus, IDS e IPS entre otros.

Un sistema de seguridad puede ser visto como una caja negra con unas entradas (trafico de red, logs, reportes de hardware, etc.), unas salidas (alarmas, reportes de red, logs) y un proceso cuya finalidad es actuar sobre las entradas, procesarlas y generar las salidas necesarias . Como punto de partida es necesario definir el sistema haciendo las restricciónes necesarias en cuanto a los mecanismos de ataque y defensa. Para el presente caso, el sistema de seguridad a tratar se restringirá a los sistemas de detección de intrusiones (IDS) y el ataque a explorar, será el ataque de denegación de servicios (DoS).

deficion de sistema de seguridad incompleta

1. Objetivos

1. Describir de manera consista el diagrama de bloques de un sistema de seguridad.
2. Hacer un estudio breve de entradas de trafico asociado con ataques de denegacion de servicio.
3. Hacer un inventario a partir de la literatura de algunas metricas del ataque.
4. Cosultar como obtener las metricas.

2. Introducción

En la figura 1 de muestra el diagrama de bloques de un sistema de seguridad simplificado que se divide en los siguientes componentes:

Puede que sea mejor modificar la figura (ver cuaderno)

1. **Preprocesamiento**: Componente que procesa los datos de entrada (datos de red sin procesar) para extraer sus principales caracteristicas con el objetivo de generar una representación equivalente pero mas reducida (datos o vectores caracteristicos, estadisticas) y apropiada para etapas de procesamiento posteriores.
2. **Alarma**: tal y como se muestra en la figura 1, este componente toma los datos caracteristicos y lanza alarmas de red (logs que reporta eventos, reportes de red, etc) con el fin de indicar a los administradores posibles problemas en la red. El papel de las alarmas no se limita meramente al de indicadores, tambien pueden ser empleadas como entradas adicionales a un componente de procesamiento posterior para posterior analisis.

3. **Procesamiento**: este componente lleva a cabo acciones de control (bloquear trafico, limitar ancho de banda, reconfigurar la red, aislar equipos infectados, lanzar indicadores de alarma, etc) con el fin de mitigar problemas en la red sin intervención humana.

posible renombramiento de este componente

Al momento de analizar y probar una propuesta de un sistema de seguridad, una de las limitaciones con las que se cuenta esta relacionada con la disponibilidad de datos de tráfico reales. Para tratar este problema, el presente documento explora diferentes alternativas (como data sets y generadores de tráfico) que, de acuerdo a la literatura pueden ofrecer una manera aceptable de imitar una fuente de tráfico real cuando se carece de esta. Posteriormente, se exploran metricas de analisis de tráfico tratando de hacer enfasis en las mas relevantes para los ataques de denegación de servicio. Finalmente, el documento culminará con una sección dedicada a las discuciones y conclusiones.

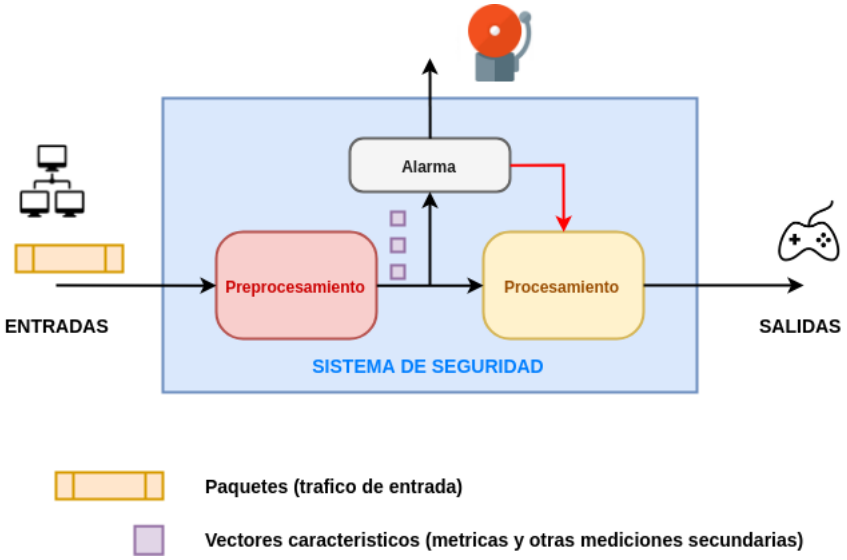

Figura 1: Sistema de seguridad simplificado

3. Entradas

De todos los tipos de entradas existentes (trafico de red, carga de memoria, logs, puertos abiertos, etc), solo unas cuantas son empleadas en un determinado sistema de seguridad. Las entradas utilizadas dependen del tipo de sistema implentado (antivirus, IDS, IPS, firewall, etc). Asi, por ejemplo, antivirus no empleará las mismas entradas que un IDS.

Teniendo en cuenta lo anterior, el primer paso es definir el tipo de sistema a implementar, que, para el caso es el IDS. Un IDS, es un sistema cuya finalidad es evaluar el

tráfico de red en busca de amenazas y lanzar alarmas en caso de detección de un patrón de tráfico anormal.

Una vez definido el sistema de seguridad, el siguiente paso es determinar de todas las entradas existentes cuales utilizar, siendo la entrada para el caso el **trafico de red**. Este se clasifica de la siguiente manera:

- **Trafico real**: En este caso el tráfico es generado por maquinas reales o virtuales conectados a la red.
- **Trafico sintético**: En este caso el tráfico es generado por una aplicación que simula el comportamiento del trafico generado por una maquina real.

Finalmente, como un mismo tipo de entrada puede estar asociada a muchos tipos de ataques, es necesario definir con claridad el ataque en el que se hará énfasis, siendo para el caso, el ataque de Denegación de servicio (DoS) el elegido.

En conclusión y resumiendo lo anterior, la defición de las entradas a emplear en un sistema de seguridad se reduce a los siguientes tres pasos básicos:

1. Definir el sistema de seguridad a emplear.
2. Definir de acuerdo al paso uno, las entradas que el sistema empleará.
3. Definir el ataque que se analizará.

Con estos tres items definidos representados por la triada [**herramienta, tipo de entrada, tipo de ataque**] que para el caso es [**IDS, tráfico de datos, DoS**], se tiene la información suficiente para empezar a definir de manera más específica la fuente que se empleará como entrada en el sistema.

De acuerdo a algunas fuentes de literatura consultadas [6, 8] la generación de tráfico de entrada asociado con ataques de denegación de servicio simples o distribuidos (DoS o DDoS) puede realizarse empleando diferentes tipos de fuentes, las cuales se pueden agrupar en los siguientes tres tipos:

- Datasets
- Herramientas de geneneracion de ataques de denegación de servicio.
- Generadores de tráfico.

En las siguientes secciones se explicará con un poco mas de detalle cada una de estas.

3.1. Fuentes de generación de ataques de degación de servicio

En la figura 2, se muestran las posibles fuentes que pueden ser seleccionadas para la generacion de tráfico de red previamente mencionadas. Tal y como se muestra en dicha fígura, inicialmente se define el tipo de fuente que se va a emplear en el experimento para la generación del tráfico de entrada que se aplicará al sistema de seguridad ya definido.

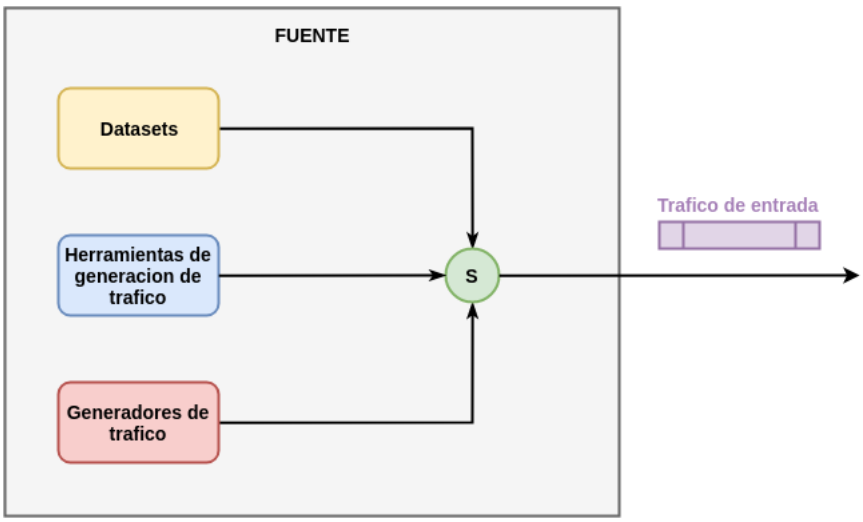

Figura 2: Sistema de seguridad simplificado

Una vez hecho lo anterior, se aplica este tráfico de entrada a dicho sistema con el objetivo de probarlo, evaluarlo y dado el caso (si se emplean tecnicas de machine learning) entrenarlo.

A continuación se aborda con mas detalle cada una de las fuentes mostradas en la figura 2.

3.1.1. Datasets

Un dataset se define como una coleccion de datos (items) distretos y relacionados con diferentes significados según el escenario y que fueron utilizados para alguna clase de experimento o analisis [7].

En la red existe diferentes fuentes de las cuales se pueden obtener datasets de manera libre [5, 4, 1]. En este caso, como la disciplina de interes se centra en datos asociados a tráfico de red, la busqueda y elección de datasets que cumplan con este requisito es aun una tarea desafiante en gran parte, debido a la falta de un sitio centralizado y especializado donde sea facil obtenter este tipo de datos.

Para tratar esta dificultad, Cinthya Grajeda et al [7], presentan un overview de datasets relevantes en analisis digital forense. Asi mismo, recopilan toda esta informacion en un repositorio centralizado [2] para facilitar la busqueda, actualización y uso por parte de la comunidad, de datasets relacionados con escenarios de seguridad. De los datasets allí presentados, los únicos que representan algún interés para nuestro caso, son aquellos relacionados con trafico de red. En la tabla 1 se muestran algunos datasets de interes que pueden ser empleados como fuentes de datos para la reproducción de experimentos relacionados con los ataques de denegación de servicios.

Las pricipales caracteristicas mostradas en la tabla 1 para cada dataset, estan rela-

cionadas con el tipo de datos que los componen (archivos pcap, logs, etc) que son de vital importancia por que determinan los parámetros (variables: IP origen, IP destino, etc) asociados a cada dato, el tamaño del dataset, la fecha de disponibilidad y si es Labeled (L) o Unlabeled (U).

El uso de datasets facilita el diseño de pruebas experimentales pues, permite la aplicación de una misma entrada (dataset como tal) ante diferentes condiciones y configuraciónes del sistema de seguridad estudiado. Ademas, los datasets son ampliamente usados en areas de investigacion con machine learning (ML) y sistemas de intrucion (IDS) [9] lo cual hace que valga la pena que estos sean empleados como una fuente de entrada al definir un experimento.

Como se puede enfatizar que este es el ultimo parrafo, se puede deja asi o es necesario hacer este enfasis

Dataset	***Tipo de datos***	***Tamaño***	***Fecha***	***Labeled or Unlabeled***
Digital Corpora	archivos pcap	N/A	2008 - 2009	U
DFRWS 2009 Challenge	archivos pcap	N/A	2009	U
University of New Haven cFREG	archivos pcap	876 KB	2015	U
The CFReDS Project - NIST	trace logs	3.8 MB	2005	?
CAIDA	68 network related datasets	N/A	1998 - 2017	?
University of Oregon Route Views Project	Cisco, Zebra BGP RIBs	N/A	1997 - 2017	?
DARPA	(Raw dataset) TCP/IP Dump files	9.67 GB	1999	L
KDD99	Caracteristicas extraidas y preprocesadas del dataset DARPA usando machine learning	5209460	1999	L
NLS-KSDD	Version reducida del dataset KDD99 (se remueven datos redundantes)	N/A	?	L
CIDDS-001	flujos de red + labels	N/A	?	L

Cuadro 1: Principales caracteristicas de algunos datasets para hacer pruebas con ataques de denegacion de sevicio

3.1.2. Herramientas para lanzar ataques de denegacion de servicio

Existe un gran numero de herramientas para realizar ataques. Mas exactamente en el caso de los ataques de denegación de servicios, la obtención y uso de estas es sumamente facil gracias a la existencia de portales como sectools o distribuciones linux enfocadas en seguridad como kali que traen muchas de estas aplicaciones por defecto. La siguiente tabla [6] muestran algunas aplicaciones empleadas para lanzar ataques de denegación de servicio de manera resumida:

Agregando nueva tabla

Herramienta	*Tipo de trafico*	*Método de ataque*	*Tipo de ataque DoS/DDoS*	*Impacto*
GoldenEye	HTTP	GET Flood, POST Flood, Random Flood	Aplicacion	Recurso
LOIC (Low Orbit Ion Cannon)	HTTP, TCP, UDP	GET Flood, TCP Flood, UDP Flood	Aplicacion	Recurso
R.U.DY (R U Dead Yet?)	HTTP	HTTP POST	Aplicacion	Recurso
Slowloris	HTTP	HTTP GET	Aplicacion	Recurso
Dirt Jumper	HTTP	POST Flood, SYN Flood, HTTP Flood	Aplicacion	Recurso
Tor's Hammer	HTTP	slow POST	Aplicacion	Recurso
Nuclear DDoSer	http	Slowloris, Slow POST	Aplicacion	Recurso
Railgun	http	Slowloris o Slow POST	Aplicacion	Recurso
High Orbit Ion Cannon (HOIC)	http	POST Flood, GET Flood	Aplicacion	Recurso
HULK (HTTP Unbearable Load King)	HTTP	TCP SYN flood, HTTP GET flood	Aplicacion	Recurso

Cuadro 2: Principales caracteristicas de algunos datasets para hacer pruebas con ataques de denegacion de sevicio

Herramienta	***Tipo de trafico***	***Método de ataque***	***Tipo de ataque DoS/DDoS***	***Impacto***
TFN (Tribe Flood Network)	ICMP, TCP, UDP	ICMP Flood, SYN Flood, UDP Flood and Smurf attack	Por volumen	Ancho de banda
trin00 (o trino)	UDP, TCP	SYN Flood, UDP flood	Por volumen	Ancho de banda
stacheldraht	ICMP, TCP, UDP	ICMP Flood, SYN Flood, UDP Flood and Smurf attack	Por volumen	Ancho de banda
Hping3	TCP, UDP, ICMP, RAW-IP	?	Por volumen	Ancho de banda
Ddosim	HTTP, TCP, SMTP	Create full TCP connections, when the connection is stablised send HTTP GET request	Aplicación	Recurso
Pyloris	HTTP	?	Aplicación	Recurso
Davoset	HTTP	Abuse of Functionality	Aplicación	Recurso
Trinity	TCP, UDP	Flood attacks	Aplicación	Ancho de banda, Recurso
XOIC	TCP, UDP, HTTP, ICMP	?	Aplicación	Ancho de banda, Recurso
Owasp Http Dos Post	HTTP	HTTP POST attacks, HTTP GET attacks	Aplicación	Recurso
THC-SSL-DoS	SSL	Malformed SSL request	Presentación (Protocol ?)	Recurso
Brobot	TCP, UDP	HTTP POST attacks, HTTP GET attacks	Por volumen	Ancho de banda

Cuadro 3: Principales caracteristicas de algunos datasets para hacer pruebas con ataques de denegacion de sevicio

Como existen varios tipos de ataques de denegación de servicio; conocer la herramienta, permite definir el tipo de ataque en el que esta se enfoca y por ende es un paso fundamental al definir la entrada que sera empleada en el experimento.

3.1.3. Generadores de trafico

Los generadores de trafico son herramientas que pueden generar trafico tanto legitimo como de ataque. A continuación muestran algunos resumiendo sus caracteristicas mas relevantes [6].

Nota: Ver las herramientas relacionadas —¿http://bittwist.sourceforge.net/doc.html

Generador	***Descripción***	***Parametros de entrada***
Bit-Twist	Es una herramienta de generacion de diferentes tipos de trafico Ethernet. Permite generar paquetes a partir de trazas tcpdump (.pcap). Adicionalmente, esta herramienta permite la edición de edicion de trazas.	TCP, UDP, IP,ARP
packETH	Generador de paquetes ethernet que permite crear y enviar cualquier paquete o secuencia de paquetes a traves de un link ethernet.	TCP, UDP, IP, ARP, ICMP
Nemesis	Utilidad que permite la reacion e inyeccion de paquetes de red. Es ampliamente usado para testear IDS, firewals e IP stacks entre otros.	ARP, DNS, ET-HER-NET, ICMP, IGMP, IP, OSPF, RIP, TCP, UDP
D-ITG (Distributed Internet Traffic Generator)	Es una herramienta con la capacidad de generar trafico de manera mas realista usando procesos estocasticos para IDT (Inter Departure Time) y PS (Packet Size).	HTTP, TCP/IP
curl-loader	Herramienta que simula el comportamiento y carga generada por miles y decenas de miles de clientes HTTP/HTTPS y FTP/FTPs con sus propias direcciones IP. Esta herramienta es util para la medicion de carfas de desempeño de varias aplicaciones, para testeo de servidores web y ftp y para generar trafico.	HTTP, HTTPS, FTP, FTPS

Generador	*Descripción*	*Parametros de entrada*
HTTPerf	httperf es una herramienta para la medición de desempeño en servidores web. Esta aplicación es basicamente un cliente que ejecuta request especificos contra un servidor para luego realizar mediciones y registros de metricas como el tiempo de resupuesta.	HTTP, SSL

Los generadores de trafico son utiles para simular trafido de red, testear firewalls, IDS e IPS, asi mismo para resolver varios problemas de red. (Mejorar esta redaccion y agregar su papel para las entradas)

4. Analisis de trafico

Herramientas para el monitoreo de trafico

Herramienta	*Uso*
Wireshark	Analisis de protocolo
tcpwrite	Edición de archivos de trafico **pcap** que permite reescribir headers TCP/IP y de capa 2, asi mismo permite generar trafico mediante el reuso de paquetes **pcap** ya disponibles.
tcpreplay	Permite el reuso de paquetes de trafico previamente capturados a velocidades arbitrarias en la red
nmap	Herramienta para escaneo de puertos y exploración de redes
tcptrack	Usada para sniffing y despliegue de información (IPs fuente y destino, estado de la conexion, idle time, Puertos fuente y destino y uso del ancho de banda en la conexion entre otros) de las conexiones de red vistas en la interfaz de red.

Parametros en los paquetes de red - Representación

Para llevar a cabo la labor de preprocesamiento es necesario hacer una captura de los paquetes que viajan a traves de la red con el proposito de realizar una inspección profunda de sus principales caracteristicas. Interfaces de programación para la captura de paquetes como pcap (packet capture) e interfaces de monitoreo de paquetes como NetFlow o sflow son bastante comunes. La siguiente tabla muestra algunas de las caracteristicas que pueden ser obtenidas con estas:

5. Experimento

Seccion dedicada a describir pasos importantes sobre el experimento (ojo con los enlaces comentados):

Parametro	***Descripción***
Src IP	Direccion IP fuente
Src Port	Puerto fuente
Dest IP	Direccion IP destino
Dest Port	Puerto destino
Proto Transport Protocol	Protocolo de transporte (ICMP, TCP o UDP)
Num	Numero del paquete
Tiempo de llegada	Tiempo de llegada de un paquete
Size	Tamaño del paquete ???
header len	Longitud de la cabecera
total len	Longitud total (la verdad no se de que???)
flags	bandereas

6. Salidas

Extracción de caracteristicas

El conocimiento de estos parametros de red (algunos de los cuales fueron previamente citados) es de extrema utilidad por que permite analisis de trafico de red tanto offline como online. Sin embargo, adicional a este proceso, es necesario llevar a cabo una tarea adicional sobre el trafico con el fin de seleccionar los parametros mas relevantes u obtener medidas secundarias (metricas) para etapas de procesamiento posteriores. La siguiente tabla muestra algunos de los parametros que suelen ser empleados:

- % of same service to same host
- % on same host to same service
- average duration / all services
- average duration /current host
- average duration / current service
- bytes transfered / all services
- bytes transfered / current host
- bytes transfered / current service
- Destination bytes
- Destination IP
- Destination port
- Duplicate ACK rate
- Duration

- Hole rate
- Land packet
- Protocol
- Resent rate
- Source bytes
- Source IP
- Source port
- TCP Flags
- Timestamp
- # different services accessed
- # establishment errors
- # FIN flags
- # ICMP packets
- # keys with outside hosts
- # new keys
- # other errors
- # packets to all services
- # RST flags
- # SYN flags
- # to certain services
- # to privileged services
- # to the same host
- # to the same service
- # to unprivileged services
- # total connections
- # unique keys

- # urgent
- % control packets
- % data packets
- wrong data packet size rate
- variance of packet count to keys

Tras ver todo este gran numero de parametros entran una serie de preguntas que son de vital importancia resolver y que se citan a continuación:

- ¿Como obtener todas estas caracteristicas del trafico de red que se esta analizando ya sea de manera online o de manera offline?
- ¿Que herramientas o librerias pueden existir para facilitar esta tarea?
- ¿Como configurarlas y ponerlas a punto para la extracción de caracteristicas?

7. Conclusiones

El código ejemplo se encuentra disponible en: `https://github.com/tigarto`

Referencias

[1] Awesome public datasets. (Date last accessed 15-September-2018).

[2] Datasets for cyber forensics. (Date last accessed 15-September-2018).

[3] Internet world stats. (Date last accessed 21-August-2018).

[4] Kaggle. (Date last accessed 15-September-2018).

[5] Uci machine learning repository. (Date last accessed 15-September-2018).

[6] Sunny Behal and Krishan Saluja. Characterization and comparison of ddos attack tools and traffic generators -a review. 19:383–393, 04 2017.

[7] Cinthya Grajeda, Frank Breitinger, and Ibrahim Baggili. Availability of datasets for digital forensics – and what is missing. *Digital Investigation*, 22:S94 – S105, 2017.

[8] N. Hoque, Monowar H. Bhuyan, R.C. Baishya, D.K. Bhattacharyya, and J.K. Kalita. Network attacks: Taxonomy, tools and systems. *Journal of Network and Computer Applications*, 40:307 – 324, 2014.

[9] Atilla Özgür and Hamit Erdem. A review of kdd99 dataset usage in intrusion detection and machine learning between 2010 and 2015. *PeerJ Preprints*, 4:e1954v1, April 2016.