

UNIVERSIDAD DE ANTIOQUIA

REPORTE 3

Metricas - Parte 1

Author:
Henry ARCILA

Supervisors:
Prof. Natalia GAVIRIA Prof.
Danny MÚNERA

September 15, 2018



Contents

1	Objetivos	2
2	Introducción	2
3	Entradas	3
3.1	Fuentes de generación de ataques de denegación de servicio	4
3.1.1	Datasets	5
3.1.2	Herramientas para lanzar ataques de denegación de servicio . . .	8
3.1.3	Generadores de tráfico	11
4	Análisis de tráfico	13
5	Experimento	13
6	Salidas	14
7	Conclusiones	16

Abstract

De acuerdo al World internet usage and population statistics, aproximadamente un 54.4% tienen acceso a internet [3]. Como el recurso por excelencia intercambiado a través de internet es la información este debe ser protegido; sin embargo, dicha tarea es cada vez más desafiante debido a la mayor facilidad, número 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 (tráfico 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 restricciones 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).

deficiencia de sistema de seguridad incompleta

1 Objetivos

1. Describir de manera consistente el diagrama de bloques de un sistema de seguridad.
2. Hacer un estudio breve de entradas de tráfico asociado con ataques de denegación de servicio.
3. Hacer un inventario a partir de la literatura de algunas métricas del ataque.
4. Consultar cómo obtener las métricas.

2 Introducción

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

1. **Preprocesamiento:** Componente que procesa los datos de entrada (datos de red sin procesar) para extraer sus principales características con el objetivo de generar una representación equivalente pero más reducida (datos o vectores característicos, estadísticas) y apropiada para etapas de procesamiento posteriores.
2. **Alarma:** tal y como se muestra en la figura 1, este componente toma los datos característicos y lanza alarmas de red (logs que reportan 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, también pueden ser empleadas como entradas adicionales a un componente de procesamiento posterior para posterior análisis.

Puede que sea mejor modificar la figura (ver cuaderno)

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
renom-
bramiento
de este
compo-
nente

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 énfasis en las mas relevantes para los ataques de denegación de servicio. Finalmente, el documento culminará con una sección dedicada a las discuciones y conclusiones.

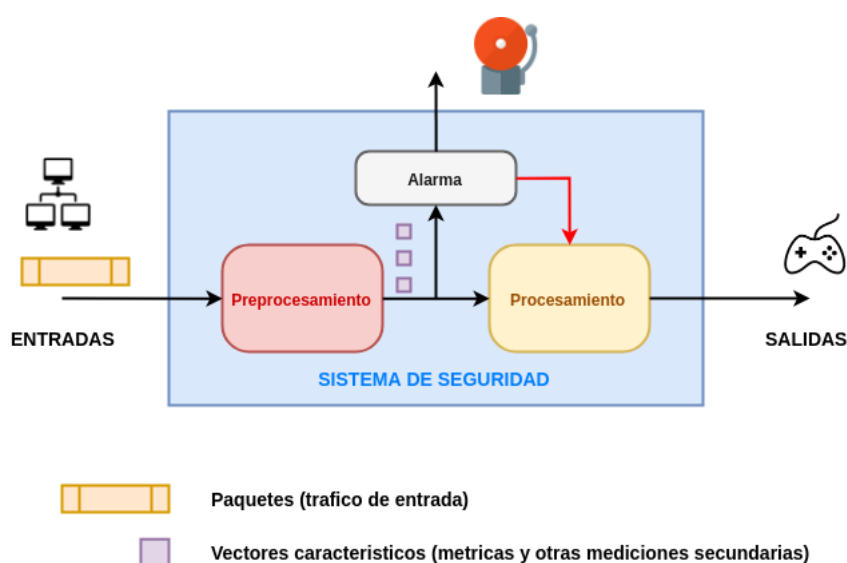


Figure 1: Sistema de seguridad simplificado

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 **tráfico 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 [7, 9] 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 generacion 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 figura, 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.

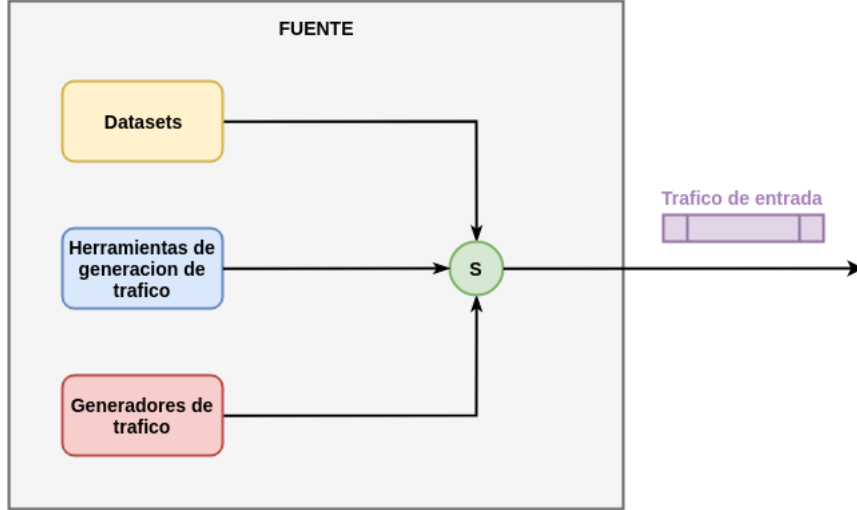


Figure 2: Sistema de seguridad simplificado

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 [8].

En la red existe diferentes fuentes de las cuales se pueden obtener datasets de manera libre [6, 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 [8], 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 características 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 configuraciones del sistema de seguridad estudiado. Además, los datasets son ampliamente usados en áreas de investigación con machine learning (ML) y sistemas de intrusión (IDS) [11] 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 último párrafo, se puede dejar así o es necesario hacer este énfasis

<i>Dataset</i>	<i>Tipo de datos</i>	<i>Tamaño</i>	<i>Fecha</i>	<i>Labeled or Unlabeled</i>
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	Características extraídas y preprocesadas del dataset DARPA usando machine learning	5209460	1999	L
NLS-KSDD	Version reducida del dataset KDD99 (se remueven datos redun- dantes)	N/A	?	L
CIDDS-001	flujos de red + labels	N/A	?	L

Table 1: Principales características de algunos datasets para hacer pruebas con ataques de denegación de servicio

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 [7] muestran algunas aplicaciones empleadas para lanzar ataques de denegación de servicio de manera resumida:

Aqui vamos - Nota falta ponerle titulo a la tabla

<i>Dataset</i>	<i>Descripción</i>	<i>Tipo de ataque</i>
LOIC (Low Orbit Ion Cannon)	herramienta openasource diseñada que permite realizar ataques de denegacion de servicio enviando un gran cantidad de paquetes de red con el proposito de saturar el ancho de banda	tcp, udp, icmp,http
HOIC (High Orbit Ion Cannon)	herramienta para lanzar hasta 256 ataques de denegacion (a diferentes dominios) de servicio de manera paralela mediante el envio de GET flood and POST requeststs	http
HULK (HTTP Unbearable Load King)	herramienta para dar de baja servidores web empleando TCP SYN flood and multi threaded HTTP GET flood requests. Esta herramienta usa varias obfuscation technique para limitar la capacidad del objetivo de mitigar el ataque.	http
Tor's Hammer	Es una herramienta escrita en python diseñada para agotar el ancho de banda y los recursos de la victima. Se caracteriza por correr en una red ToR usando direcciones IPs aleatorias con el fin de dificultar el rastreo del atacante.	http

<i>Dataset</i>	<i>Descripción</i>	<i>Tipo de ataque</i>
RUDY (R-U-dead-Yet)	Es una herramienta escrita en python usada para realizar slow attacks con el fin de comprometer la disponibilidad de servidores web. Los ataques realizados empleando esta herramienta, se caracterizan por abrir conexiones POST HTTP simultáneas un servidor HTTP y, retrasar el envío del cuerpo de la solicitud POST lo cual hace que los recursos del servidor se saturen. El envío paquetes pequeños a un ritmo tan lento (para mantener la conexión abierta y al servidor ocupado) hace que la detección de este tipo de ataques sea difícil de detectar comparado con los ataques DoS por inundación.	http
Golden-Eye	Es una herramienta escrita en python que permite realizar ataques http flood con el objetivo de agotar los recursos de la víctima, es ideal para probar websides pero no es efectiva en el mundo real los ataques lanzados con esta herramienta son fáciles de detectar por defensas perimetrales.	http
TFN(Tribe flood network)	es una herramienta con la capacidad de lanzar múltiples ataques DDoS usando UDP flood, Smurf attack y TCP SYN flood.	tcp,udp,icmp

<i>Dataset</i>	<i>Descripción</i>	<i>Tipo de ataque</i>
Slowloris	es una herramienta para lanzar ataques DoS enviando continuamente TCP SYN requests a la víctima con el proposito de establecer muchas conexiones abiertas con la víctima tanto como sea posible lo que satura la capacidad esta para recibir nuevas conexiones de los clientes.	http
Ddosim	Es una herramienta para generar ataques DDoS que usa direcciones IP aleatorias para simular varios zombies los cuales crean conexiones full TCP con la víctima. Una vez establecida la conexión la herramienta lanza ataques como HTTP-GET flood para hacerla no disponible.	tcp,smtp ,http,udp
Pyloris	Esta herramienta permite al usuario realizar ataques DoS sobre un servicio. Puede trabajar con varios protocolos, incluyendo HTTP, FTP, SMTP, IMAP, y Telnet.	—

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 características mas relevantes [7].

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

<i>Generador</i>	<i>Descripción</i>	<i>Parametros de entrada</i>
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 reaccion e inyeccion de paquetes de red. Es ampliamente usado para testear IDS, firewals e IP stacks entre otros.	ARP, DNS, ETHER-NET, ICMP, IGMP, IP, OSPF, RIP, TCP, UDP
D-ITG (Dis-tributed Internet Traffic Genera-tor)	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 car-fas de desempeño de varias aplicaciones, para testeo de servidores web y ftp y para generar trafico.	HTTP, HTTPS, FTP, FTPS

<i>Generador</i>	<i>Descripción</i>	<i>Parametros de entrada</i>
HTTPerf	httpperf 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

<i>Herramienta</i>	<i>Uso</i>
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 características. 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 características que pueden ser obtenidas con estas:

5 Experimento

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

<i>Parametro</i>	<i>Descripción</i>
Src IP	Dirección IP fuente
Src Port	Puerto fuente
Dest IP	Dirección IP destino
Dest Port	Puerto destino
Proto Transport Protocol	Protocolo de transporte (ICMP, TCP o UDP)
Num	Número 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 características

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

7 Conclusiones

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

References

- [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] Kdd cup 1999 data data set. (Date last accessed 9-August-2018).
- [6] Uci machine learning repository. (Date last accessed 15-September-2018).
- [7] Sunny Behal and Krishan Saluja. Characterization and comparison of ddos attack tools and traffic generators -a review. 19:383–393, 04 2017.
- [8] Cinthya Grajeda, Frank Breitingner, and Ibrahim Baggili. Availability of datasets for digital forensics – and what is missing. *Digital Investigation*, 22:S94 – S105, 2017.

- [9] 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.
- [10] Abhishek Verma and Virender Ranga. Statistical analysis of cids-001 dataset for network intrusion detection systems using distance-based machine learning. *Procedia Computer Science*, 125:709 – 716, 2018. The 6th International Conference on Smart Computing and Communications.
- [11] 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.