

UNIVERSIDAD DE ANTIOQUIA

REPORTE 3

---

# Metricas - Parte 1

---

*Author:*

Henry ARCILA

*Supervisors:*

Prof. Natalia GAVIRIA

Prof. Danny MÚNERA

August 23, 2018



# Contents

<b>1</b>	<b>Objetivos</b>	<b>2</b>
<b>2</b>	<b>Introducción</b>	<b>2</b>
<b>3</b>	<b>Entradas</b>	<b>5</b>
3.1	Fuentes de generación de ataques de denegación de servicio . . .	5
3.1.1	Datasets . . . . .	6
3.1.2	Herramientas para lanzar ataques de denegacion de servicio . . . . .	8
3.1.3	Generadores de trafico . . . . .	12
<b>4</b>	<b>Analisis de trafico</b>	<b>14</b>
<b>5</b>	<b>Salidas</b>	<b>15</b>
<b>6</b>	<b>Conclusiones</b>	<b>18</b>

## Abstract

De acuerdo al World internet usage and population statistics, aproximadamente un 54.4% tienen acceso a internet [1]. Como el recurso por excelencia intercambiado a través de internet es la información, datos como estos demuestran que la información es un elemento fundamental para el funcionamiento de la sociedad. Al ser información un recurso vital 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 (datos de red: 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 existen una gran cantidad de ataques y defensas es necesario definir con claridad lo que hará el sistema de seguridad. 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).

## 1 Objetivos

1. Describir de manera consistente 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 mas reducida (vectores característicos) mas apropiada para etapas de procesamiento posteriores.
2. **Alarma:** este componente los datos característicos resultantes de la etapa de preprocesamiento 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 al componente de procesamiento tal y como se muestra en la figura 1.
3. **Procesamiento:** este componente lleva a cabo acciones de control (bloquear trafico, limitar ancho de banda, reconfigurar la red, aislar equipos infectados, etc) con el fin de mitigar problemas en la red sin intervención humana.

En el presente reporte se hará énfasis en las fuentes empleadas para generar datos de entrada y las diferentes herramientas empleadas para su análisis, restringiendo el estudio a los ataques de denegación de servicio. Finalmente, el documento culminará con una sección dedicada a las discusiones y conclusiones.



Figure 1: Sistema de seguridad simplificado

### 3 Entradas

Un sistema de seguridad puede procesar diferentes tipos de entradas (tráfico de red, carga de memoria, logs, puertos abiertos, etc). Existen además, diferentes tipos de sistemas de seguridad (antivirus, firewalls, IDS, IPS, etc.) los cuales, de acuerdo a unas entradas determinadas realizan tareas para proteger, desde equipos de computo hasta redes de posibles ataques. De todas las entradas posibles, solo unas cuantas son empleadas y dependen específicamente de la tarea llevada a cabo por el sistema de seguridad; así por ejemplo, las entradas para un antivirus no serán las mismas que para un sistema de detección de intrusiones (IDS).

Lo anterior plantea una primera restricción ¿Que sistema de seguridad se va a implementar? Pues bien, para el caso, se hará énfasis en los IDS que son sistemas 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.

Cuando se define un IDS, es necesario definir el tráfico de entrada a emplear. Por un lado por que esto facilitará las pruebas y evaluación de dicho sistema y por el otro, por que permitira un mejor control de las condiciones experimentales. Adicionalmente, debido al gran numero de ataques es bueno restringir el ataque se hará énfasis, siendo el ataque de Denegación de servicio (DoS) el elegido para este caso. Para poder generar tráfico DoS existen un gran numero de herramientas de diferentes tipos [3, 4]. En el presente documento, tomando como base la literatura relacionada con el tema, se definieron las siguientes tres fuentes de tráfico tal y como se muestra en la figura 2:

- 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 denegación de servicio

Cuando nos referimos a las fuentes de generación de ataques, en el siguiente documento queremos hacer referencia a aquellas herramientas y conjuntos de

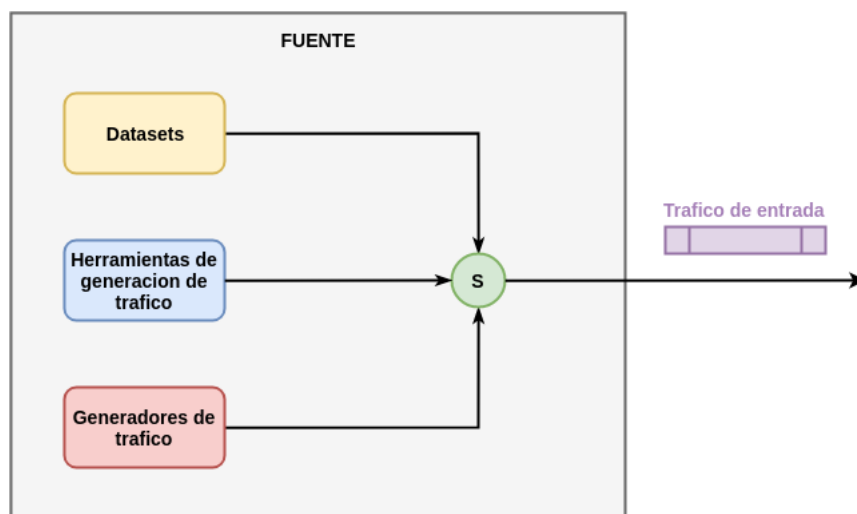


Figure 2: Sistema de seguridad simplificado

datos que pueden ser empleados como fuente de generacion de trafico no solo para lanzar ataques, sino tambien como sucede mas especificamente en lo que queremos, para evaluar un sistema de seguridad propuesto. Como hablar de ataques es ahondar en un tema muy amplio, la busqueda se restringio a herramientas para lanzar ataques de denegación de servicio para lo cual haciendo una revisión de la literatura se realizo la siguiente clasificación.

### 3.1.1 Datasets

En lo que respecta al trafico de red existen varios datasets (conjunto de datos) relacionados con difentes ataques tales como los que se muestran en la siguiente tabla:

<b><i>Dataset</i></b>	<b><i>Descripción</i></b>
MIT Lincoln Laboratory IDS Datasets & Scan Data Repository	Large collection of Internet-wide scanning data from Rapid7, the University of Michigan, and others link
UCI	Repositorio de Machine Learnin de la universidad de Irvine link
NSA Cyber Defense Exercise Dataset	Dataset Snort, DNS, web server, and Splunk logs link
Internet-Wide Scan Data Repository	Large collection of Internet-wide scanning data from Rapid7, the University of Michigan, and others link
LIBSVM	This page contains many classification, regression, multi-label and string data sets stored in LIBSVM format. Many are from UCI, Statlog, StatLib and other collections link
Center for Applied Internet Data Analysis (CAIDA) Datasets	Internet measurement with collaboration of numerous institutions, academics, commercial and noncommercial contributors, including anonymized Internet traces, Code Red worm propagation, passive traces on high-speed links link



En cada data set se definen unas parametros (features) como la IP origen, la IP destino, el tipo de ataque, etc. Un ejemplo se puede encontrar en el KDD Cup 1999 Data Set[2, 5].

La utilidad de los datasets radica en su amplio uso en areas de investigacion relacionadas con machine learning (ML) y sistemas de intrusion (IDS) [6] lo que da pie para la utilización de uno de estos dataset con el fin de llevar a cabo la evaluación del sistema de seguridad propuesto. Sin embargo esto abre interrogantes como:

- ¿Como llevar a cabo un montaje experimental que haga uso de un dataset?
- ¿Será necesario el uso de todos los parametros de un dataset?, ¿Cuales serian los mas relevantes?
- ¿Que se puede lograr empleando un dataset?

### **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 [3] muestran algunas aplicaciones empleadas para lanzar ataques de denegación de servicio de manera resumida:

<b><i>Dataset</i></b>	<b><i>Descripción</i></b>	<b><i>Tipo de ataque</i></b>
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><b>Dataset</b></i>	<i><b>Descripción</b></i>	<i><b>Tipo de ataque</b></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><b>Dataset</b></i>	<i><b>Descripción</b></i>	<i><b>Tipo de ataque</b></i>
Slowloris	es una herramienta para lanzar ataques DoS enviando continuamente TCP SYN requests a la victima con el proposito de establecer muchas conexiones abiertas con la victima 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 loc cuales crean conexiones full TCP con la victima. Una vez establecida la conexion 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 [3].

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

<i><b>Generador</b></i>	<i><b>Descripción</b></i>	<i><b>Parametros de entrada</b></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 reacion e inyeccion de paquetes de red. Es ampliamente usado para testear IDS, firewals e IP stacks entre otros.	ARP, DNS, ETH- ER- NET, ICMP, IGMP, IP, OSPF, RIP, TCP, UDP
D-ITG (Dis- tributed Inter- net Traffic Gener- ator)	Es una herramienta con la capaci- dad 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 com- portamiento y carga generada por miles y decenas de miles de clientes HTTP/HTTPS y FTP/FTPs con sus propias direcciones IP. Esta her- ramienta es util para la medicion de carfas de desempeño de varias aplica- ciones, 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 <b>pcap</b> que permite reescribir headers TCP/IP y de capa 2, asi mismo permite generar trafico mediante el reuso de paquetes <b>pcap</b> 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:

<i><b>Parametro</b></i>	<i><b>Descripción</b></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	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

## 5 Salidas

### Extracción de características

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 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?

## 6 Conclusiones

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

## References

- [1] Internet world stats. (Date last accessed 21-August-2018).
- [2] Kdd cup 1999 data data set. (Date last accessed 9-August-2018).
- [3] Sunny Behal and Krishan Saluja. Characterization and comparison of ddos attack tools and traffic generators -a review. 19:383–393, 04 2017.
- [4] 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.
- [5] 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.
- [6] 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.