ZEEK INTRUSION DETECTION SERIES

Lab 1: Introduction to the Capabilities of Zeek

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
Contents

Overview .................................................................................................................................................. 3
Objectives ............................................................................................................................................... 3
Lab topology ........................................................................................................................................... 3
Lab settings ........................................................................................................................................... 3
Lab roadmap .......................................................................................................................................... 4
1 Introduction to Zeek .......................................................................................................................... 4
  1.1 The Zeek event engine ..................................................................................................................... 5
      1.1.1 State management .................................................................................................................... 5
      1.1.2 Transport layer analyzers ........................................................................................................ 6
      1.1.3 Application layer analyzers ...................................................................................................... 6
      1.1.4 Infrastructure .......................................................................................................................... 6
  1.2 The Zeek policy script interpreter .................................................................................................... 6
  1.3 Zeek analyzers .................................................................................................................................. 6
  1.4 Signatures ......................................................................................................................................... 7
  1.5 ZeekControl .................................................................................................................................... 7
2 Using ZeekControl to update the status of Zeek ............................................................................. 8
  2.1 Starting a new instance of Zeek ......................................................................................................... 9
  2.2 Stopping the active instance of Zeek ................................................................................................. 10
3 Introduction to Zeek’s traffic analysis capabilities ........................................................................... 11
  3.1 Processing offline packet capture files ............................................................................................. 11
      3.1.1 Command format for processing packet capture files ............................................................. 11
      3.1.2 Leveraging a script to detect brute force attacks present in a pcap file ................................ 12
  3.2 Generating and analyzing live network traffic capture ................................................................. 13
      3.2.1 Leveraging the Tcpdump utility for capturing live network traffic .......................................... 13
      3.2.2 Capturing live network traffic ................................................................................................. 13
      3.2.3 Analyzing the newly captured network traffic ......................................................................... 15
References ............................................................................................................................................. 17
Lab 1: Introduction to the Capabilities of Zeek

Overview

This lab introduces Zeek, an open-source network analysis framework primarily used in security monitoring and traffic analysis. The focus in this lab is on explaining Zeek’s layered architecture and demonstrating Zeek’s capabilities of performing network traffic analysis.

Objectives

By the end of this lab, students should be able to:

1. Understand Zeek’s layered architecture.
2. Start and manage a Zeek instance using the ZeekControl utility.
3. Use Zeek to process packet captures files.
4. Generate and analyze live network traffic in Zeek.

Lab topology

Figure 1 displays the topology of the lab. This lab will primarily use the Zeek2 machine for offline packet capture processing and analysis.

![Lab topology diagram]

Figure 1. Lab topology.

Lab settings

The information (case-sensitive) in the table below provides the credentials to access the machines used in this lab.
### Table 1. Device credentials for lab workspace.

<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@admin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2, 192.168.3.1</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>Router</td>
<td>192.168.1.1, 192.168.2.1, 203.0.113.2</td>
<td>root</td>
<td>password</td>
</tr>
</tbody>
</table>

### Table 2. Shell variables and their corresponding absolute paths.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/Zeek/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/Zeek/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/Zeek-Labs-Workspace/</td>
</tr>
</tbody>
</table>

### Lab roadmap

This lab is organized as follows:

1. Section 1: Introduction to Zeek.
2. Section 2: Using ZeekControl to update the status of Zeek.
3. Section 3: Introduction to Zeek’s traffic analysis capabilities.

### 1 Introduction to Zeek

Zeek is a passive, open-source network traffic analyzer. It is primarily a security monitor that inspects all traffic on a link in depth for signs of suspicious activity Previous. It can run on commodity hardware with standard UNIX-based systems and can be used as a passive network monitoring tool.
Setting Zeek as a node with an assigned IP address on the monitored network is not mandatory. Figure 2 shows Zeek’s layered architecture. Once Zeek receives packets, its **event engine** converts them into events. The events are then forwarded to the policy script interpreter, which generates logs, notifications, and/or actions.

Zeek uses the standard *libpcap* library for capturing packets to be used in network monitoring and analysis.

### 1.1 The Zeek event engine

The event engine layer performs low-level network packets analysis. It receives raw packets from the network layer (packet capture), sorts them by connection, reassembles data streams, and decodes application layer protocols. Whenever it encounters something potentially relevant to the policy layer, it generates an event.

The event engine consists of several analyzers responsible for well-defined tasks. Typical tasks include decoding a specific protocol, performing signature-matching, identifying backdoors, etc. Usually, an analyzer is accompanied by a default script which implements some general policy adjustable to the local environment. The event engine can be divided into four major parts.

#### 1.1.1 State management

Zeek’s main data structure is a connection which follows typical flow identification mechanisms, such as 5-tuple approaches. The 5-tuple structure consists of the source IP address/port number, destination IP address/port number, and the protocol in use. For a connection-oriented protocol like TCP, the definition of a connection is more clear-cut,
however for others such as UDP and ICMP, Zeek implements a flow-like abstraction to aggregate packets. Each packet belongs to exactly one connection.

1.1.2  Transport layer analyzers

On the transport layer, Zeek analyzes TCP, UDP packets. In TCP, Zeek’s associated analyzer closely follows the various state changes, keeps track of acknowledgments, handles retransmissions and much more.

1.1.3  Application layer analyzers

The analysis of the application layer data of a connection depends on the service. There are analyzers for a wide variety of different protocols, e.g. HTTP, SMTP or DNS, that generally conduct detailed analysis of the data stream.

1.1.4  Infrastructure

The general infrastructure of Zeek includes the event and timer management components, the script interpreter, and data structures.

1.2  The Zeek policy script interpreter

While the event engine itself is policy-neutral, the top layer of Zeek defines the environment-specific network security policy. By writing handlers for events that may be raised by the event engine, the user can precisely define the constraints within the given network. If a security breach is detected, the policy layer generates an alert.

New event handlers can be created in Zeek’s own scripting language. While providing all expected convenience of a powerful scripting language, it has been designed with network intrusion detection in mind. While it is expected that additional policy scripts are written by the user, there are nevertheless several default scripts included with the initial installation of Zeek. These default scripts already perform a wide range of analyses and are easily customizable.

1.3  Zeek analyzers

The majority of Zeek’s analyzers are in its event engine with accompanying policy scripts that can be customized by the user. Sometimes, however, the analyzer is just a policy script implementing multiple event handlers. The analyzers perform application layer decoding, anomaly detection, signature matching and connection analysis. Zeek has been designed so that it is easy to add additional analyzers.
1.4 Signatures

Most network intrusion detection systems (NIDS) match a large set of signatures against the network traffic. Here, a signature is a pattern of bytes that the NIDS tries to locate in the payload of network packets. As soon as a match is found, the system generates an alert.

A well-known IDS system is *Snort*; conversely, Zeek’s general approach to intrusion detection has a much broader scope than traditional signature-matching, yet still contains a signature engine providing a functionality that is similar to that of other systems. Furthermore, while Zeek implements its own flexible signature language, there exists a converter which directly translates Snort’s signatures into Zeek’s syntax, as shown below:

```
alert tcp any any -> [a.b.0.0/16,c.d.e.0/24] 80
  ( msg:"WEB-ATTACKS conf/httpd.conf attempt";
    nocase; sid:1373; flow:to_server,established;
    content:"conf/httpd.conf"; [...] )

signature sid-1373 {
  ipproto = tcp
  dst-ip = a.b.0.0/16,c.d.e.0/24
  dst-port = 80
  # The payload below is actually generated in a
  # case-insensitive format, which we omit here
  # for clarity.
  payload /.*conf\httpd\.conf/;
  tcp-state established,originator
  event "WEB-ATTACKS conf/httpd.conf attempt"
}%
```

Figure 3. Example of signature conversion. (a) Snort’s signature. (b) Bro’s signature.

1.5 ZeekControl

*ZeekControl*, formerly known as BroControl, is an interactive shell for easily operating and managing Zeek installations on a single system or across multiple systems in a traffic-monitoring cluster.
2 Using ZeekControl to update the status of Zeek

Step 1. From the top of the screen, click on the Bro2 button as shown below to enter the Bro2 machine.

Step 2. The Bro2 machine will now open, and the desktop will be displayed. On the left side of the screen, click on the Terminal icon as shown below.
**Step 3.** Using the Terminal, input the following command to enter the *ZeekControl* directory. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key.

```
cd $ZEEK_INSTALL/bin/
```

The active directory will change, as seen on the second line of the Terminal. Note that $ZEEK_INSTALL variable was substituted by its value (/usr/local/zeek) listed in Table 2.

**Step 4.** Use the following command to view the contents of the active directory.

```
ls
```

The directory contents will be displayed. The green file name portrays an executable file.

**Step 5.** Use the following command to launch the *ZeekControl* tool. When prompted for a password, type password and hit Enter.

```
sudo ./zeekctl
```

Once active, *ZeekControl* prompt will be displayed within the Terminal. The help command will display additional information regarding *ZeekControl*.

2.1  Starting a new instance of Zeek
Step 1. To initialize Zeek, enter the following command into the *ZeekControl* prompt.

```
start
```

Step 2. Use the following command to view the status of the currently active Zeek instance to ensure that it is active.

```
status
```

The *running* status indicates that Zeek is currently active and functioning properly. The output of the *status* command includes other useful parameters:

- **Name**: the name of the Zeek instance.
- **Type**: the type of the instance (standalone in our case).
- **Host**: the hostname (localhost).
- **Pid**: the process ID. This ID can be used with other tools like *kill* to send a signal to the process.
- **Started**: the starting date and time of the instance.

### 2.2 Stopping the active instance of Zeek

Step 1. To stop Zeek, enter the following command into the *ZeekControl* prompt.

```
stop
```

Step 2. Use the following command to verify the exit status of Zeek.

```
status
```
The *stopped* status indicates that Zeek is currently stopped.

**Step 3.** To restart Zeek, enter the following command into the *ZeekControl* prompt.

```
start
```

**Step 4.** Use the following command to exit *ZeekControl*.

```
exit
```

Note that exiting the *ZeekControl* tool does not stop Zeek. Zeek is only stopped by explicitly using the `stop` command in the *ZeekControl* prompt.

## 3 Introduction to Zeek’s traffic analysis capabilities

Zeek's broad range of traffic analysis capabilities makes it an exceptional intrusion detection system (IDS) and network analysis framework. Zeek is proficient in processing packet capture (pcap) files and logging traffic on a given network interface.

### 3.1 Processing offline packet capture files

Linux-based systems process packet capture (pcap) files using the *libpcap* library. In Zeek, it is possible to capture live traffic and analyze trace files. In the following example, we analyze a pcap file using a premade script that detects brute force attacks.

#### 3.1.1 Command format for processing packet capture files
The general format for initializing offline packet capture analysis is as follows:

```
zeek -r <pcap_file_location> <script_location>
```

- `zeek`: command to invoke Zeek.
- `-r`: option signifies to Zeek that it will be reading from an offline file.
- `<pcap_file_location>`: indicates the pcap file location.
- `<script_location>`: indicates the script location.

### 3.1.2 Leveraging a script to detect brute force attacks present in a pcap file

Zeek installs a number of default scripts and trace files that can be used for testing purposes. In this section, we use the `bruteforce.pcap` as the input packet capture file and `ZeekBruteforceDetection.zeek` as the detection script. The packet capture file contains network traffic of a brute force password attack, while the script defines the brute forcing event for the Zeek event engine.

**Step 1.** Enter the lab workspace directory. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
cd $ZEEK_LABS
```

**Step 2.** Initialize Zeek offline packet parsing on the packet capture file. Use the `tab` key for autocompletion. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
zeek -r Sample-Captures/bruteforce.pcap ZeekBruteforceDetection.zeek
```

**Step 3.** After running the command, if a brute forcing attack was found, it will be logged in the `notice.log` output log file. We will use the `cat` command to view the file.

```
cat notice.log
```
Examining the proceeding image, brute forcing was detected. The log shows that 20 login attempts failed on an FTP server with an IP 192.168.56.1.

### 3.2 Generating and analyzing live network traffic capture

The Tcpdump command utility is a famous network packet analyzing tool that is used to display TCP/IP and other network packets being transmitted over the network.

#### 3.2.1 Leveraging the Tcpdump utility for capturing live network traffic

The general format for `tcpdump` is the following:

```
sudo tcpdump -i <interface_name> -s <num> -w <pcap_file_location>
```

- `sudo`: option to enable higher level privileges.
- `tcpdump`: program for capturing live network traffic.
- `-i`: option used to specify a network interface.
- `<interface_name>`: denotes the interface name.
- `-s`: option used to specify number of packets to capture.
- `<num>`: denotes the number of packets to capture. 0 equals infinite.
- `-w`: option used to specify that we will be writing to a new file.
- `<pcap_file_location>`: indicates the file location.

#### 3.2.2 Capturing live network traffic

`Bro2` machine's `ens33` interface is used to record sample network traffic.
Lab 1: Introduction to the Capabilities of Zeek

Step 1. Enter the TCP-Traffic directory by using the `cd` command.

```
cd TCP-Traffic/
```

Step 2. Use the following command to begin live packet capture. When prompted for a password, type `password` and hit Enter. If the Terminal session has not been terminated or closed, you may not be prompted to enter the password. Live packet capture will start on interface ens33.

```
sudo tcpdump -i ens33 -s 0 -w ntraffic.pcap
```

Step 3. Click on File, then New Tab to open a new tab within the same terminal window.

Step 4. Generate traffic by using the `ping` utility. Ping operates by sending Internet Control Message Protocol (ICMP) echo request packets to the target host and waiting for an ICMP echo reply. Issue the following command on the newly opened tab to conduct a ping on the Bro1 machine. When prompted for a password, type `password` and hit Enter.

```
sudo ping -c 3 192.168.1.2
```
Step 5. Re-open the first Terminal tab as shown in the figure below.

Step 6. Use the Ctrl+c key combination to stop live traffic capture. Statistics of the capture session will be displayed. 10 packets were recorded by the interface, while 6 were captured and stored in the new ntraffic.pcap file.

3.2.3 Analyzing the newly captured network traffic

Step 1: A new file ntraffic.pcap was generated after using the tcpdump tool. Use the ls command to list the current directory’s contents.

ls
Lab 1: Introduction to the Capabilities of Zeek

Step 2: Initialize Zeek offline packet parsing on the packet capture file. The -r option is used to read from a given pcap file, and the -C option is for disabling checksums validation.

```
zeek -C -r ntraffic.pcap
```

Step 3: Open the conn.log file in a text editor. The following command uses nano text editor to open the file for editing. Unlike the cat tool which displays plain text, nano displays formatted log files, which makes it easier to visualize the output.

```
nano conn.log
```

By default, Zeek will generate tab delimited columns for packet information, with each row corresponding to an individual connection. It may be necessary to maximize the Terminal window to properly view the following IP address.

```
192.168.2.2
```

A sample row is highlighted in the image above, displaying the Bro2 machine’s IP address (192.168.2.2) and the Bro1 machine’s IP address (192.168.1.2). The line shows that the protocol is ICMP since we used the ping tool. Note that the figure displays partial output.
The user can scroll to the right to see more fields. Use the \texttt{ctrl+x} shortcut combination to exit the text editor.

\textbf{Step 4:} Stop Zeek by entering the following command on the terminal. If necessary, type \texttt{password} as the password. If the Terminal session has not been terminated or closed, you may not be prompted to enter a password. To type capital letters, it is recommended to hold the \texttt{Shift} key while typing rather than using the \texttt{Caps} key.

\begin{verbatim}
  cd $ZEEK_INSTALL/bin && sudo ./zeekctl stop
\end{verbatim}

The above command navigates to Zeek’s installation directory and executes the stop command in \texttt{zeekctl}.

Concluding this lab, we have reviewed the Zeek (Bro)’s architecture and event-based engine, as well as introduced both offline and live network traffic capture.

\textbf{References}

ZEEK INTRUSION DETECTION SERIES

Lab 2: An Overview of Zeek Logs

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
## Contents

Overview ................................................................................................................................. 3
Objectives .................................................................................................................................. 3
Lab topology .............................................................................................................................. 3
Lab settings .............................................................................................................................. 3
Lab roadmap .............................................................................................................................. 4

1 Introduction to Zeek Logs ................................................................................................... 4
   1.1 Zeek Logs generated by packet analysis ........................................................................ 5
   1.2 Zeek Logs generated by recurrent network analysis ....................................................... 5
   1.3 Typical uses of Zeek Logs .............................................................................................. 6

2 Starting a new instance of Zeek ............................................................................................ 6

3 Parsing packet capture files into Zeek log files ................................................................. 7
   3.1 Overview of Zeek command options .............................................................................. 7
   3.2 Using Zeek to process offline packet capture files ......................................................... 8
   3.3 Understanding Zeek log files .......................................................................................... 9
   3.4 Basic viewing of Zeek logs ............................................................................................ 9

4 Analyzing Zeek log files ....................................................................................................... 10
   4.1 Leveraging zeek-cut for a more refined view of log files .............................................. 10
      4.1.1 Using zeek-cut in conjunction with the cat, head, and tail commands .................. 10
      4.1.2 Printing the output of zeek-cut to a text file ......................................................... 12
      4.1.3 Printing the output of zeek-cut to a csv file ......................................................... 13
   4.2 Closing the current instance of Zeek .............................................................................. 14

References .................................................................................................................................. 14
Overview

This lab covers Zeek’s logging files. Zeek’s event-based engine will generate log files based on signatures or events found during network traffic analysis. The focus in this lab is on explaining each logging file and introducing some basic analytic functions and tools.

Objectives

By the end of this lab, students should be able to:

1. Generate Zeek log files.
2. Use Linux Terminal tools combined with Zeek’s `zeek-cut` to customize the output of logs files for analysis.

Lab topology

Figure 1 displays the topology of the lab. This lab will primarily use the Zeek2 machine for offline packet capture processing and analysis.

![Lab topology](image1.png)

Figure 1. Lab topology.

Lab settings

The information (case-sensitive) in the table below provides the credentials to access the machines used in this lab.
Table 1. Device credentials for lab workspace.

<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@admin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router</td>
<td>192.168.1.1</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.0.113.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Shell variables and their corresponding absolute paths.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/Zeek/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/Zeek/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/Zeek-Labs-Workspace/</td>
</tr>
</tbody>
</table>

Lab roadmap

This lab is organized as follows:

1. Section 1: Introduction to Zeek logs.
2. Section 2: Starting a new instance of Zeek.
3. Section 3: Parsing packet capture files into Zeek log files.
4. Section 4: Analyzing Zeek log files.

1 Introduction to Zeek Logs

Zeek’s generated log files include a comprehensive record of every connection seen on the wire; this includes application-layer protocols and fields (e.g., Hyper-Text Transfer Protocol (HTTP) sessions, Uniform Resource Locator (URL), key headers, Multi-Purpose Internet Mail Extensions (MIME) types, server responses, etc.), Domain Name Server
(DNS) requests and responses, Secure Socket Layer (SSL) certificates, key content of Simple Mail Transfer Protocol (SMTP) sessions, and others.

1.1 Zeek Logs generated by packet analysis

A Zeek log is a stream of high-level entries that correspond to network activities, such as a login to SSH or an email sent using SMTP. In Zeek, each event stream has a dedicated file with its own set of features, fields, or columns.

During capture or analysis, Zeek generates a log determined by the protocol type. Due to this architecture, a Session Initiation Protocol (SIP) log for instance, does not contain any other protocols’ packets information like HTTP. Furthermore, each log file contains case-relative fields (e.g., from and subject fields in an SMTP log). Some of these log files are large and contain entries that can be either benign or malicious, whereas others are smaller and contain more actionable information.

1.2 Zeek Logs generated by recurrent network analysis

With every session of packet analysis, either through live packet analysis or the parsing of an offline packet capture file, Zeek generates session-specific log files. In addition to these session-based log files, Zeek creates network-reliant log files as well. These network-reliant files are continually generated and updated when a new session is initialized and started.

The following Zeek log files are updated daily:

- **known_hosts.log**: Log file containing information for hosts that completed TCP handshakes.
- **known_services.log**: Log file containing a list of services running on hosts.
- **known_certs.log**: Log file containing a list of Secure Socket Layer (SSL) certificates.
- **software.log**: Log file containing information about Software being used on the network.

Additionally, a list of detection-based log files is created during each session. The log files relevant to this lab are:

- **notice.log** (Zeek notices): When Zeek detects an anomaly, a corresponding notice will be raised in this file.
- **intel.log** (Intelligence data matches): When Zeek detects traffic flagged with known malicious indicators, a corresponding reference will be logged in this file.
- **signatures.log** (Signature matches): When Zeek detects traffic flagged with known malicious or faulty packet signatures, a corresponding reference will be logged in this file.
1.3 Typical uses of Zeek Logs

By default, Zeek logs all information into well-structured, tab-separated text files suitable for postprocessing. Users can also choose from a set of alternative output formats and backends such as external databases.

The Zeek-native `zeck-cut` utility can be leveraged to further specify and parse the information within the generated log files.

2 Starting a new instance of Zeek

Step 1. From the top of the screen, click on the Bro2 button as shown below to enter the Bro2 machine.

Step 2. The Bro2 machine will now open, and the desktop will be displayed. On the left side of the screen, click on the Terminal icon as shown below.

Step 3. Start Zeek by entering the following command on the terminal. This command enters Zeek’s default installation directory and invokes `zeekctl` tool to start a new instance. When prompted for a password, type `password` and hit `Enter`.

```
cd $ZEEK_INSTALL/bin && sudo ./zeekctl start
```
A new instance of Zeek is now active, and we are ready to proceed to the next section of the lab.

3 Parsing packet capture files into Zeek log files

In this section we introduce Zeek’s capability of generating and viewing log files. Packet capture files used in this lab are preinstalled into the machine, and can be found with the following path:

$ZEEK_LABS/Sample-Captures

These packet capture files were downloaded from Tcpreplay’s sample capture collection:


Tcpreplay is a suite of free Open Source utilities for editing and replaying previously captured network traffic and can be used for testing by security analysts.

3.1 Overview of Zeek command options

When using Zeek, the user specifies a running state option. In this lab, two primarily options are used:

- `-r`: specifies offline packet capture file analysis.
- `-w`: specifies live network capture.

Additional Zeek options can be found using the `-help` option to the command `zeek`:

zeek -help
3.2 Using Zeek to process offline packet capture files

In this step we use Zeek to process the existing offline packet capture file *smallFlows.pcap*. By specifying the `-r` option and the directory path to the pcap file, Zeek can generate the corresponding log files.

**Step 1.** Navigate to the lab workspace directory. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key.

```bash
cd $ZEENK_LABS/Sample-Captures
```

**Step 2.** Use the following command to process the *smallFlows.pcap* file.

```bash
zeek -r smallFlows.pcap
```
Once done processing the packet capture file, Zeek generates a number of log files.

**Step 3.** Use the following command to list the generated log files.

```
ls
```

### 3.3 Understanding Zeek log files

Zeek’s generated log files can be summarized as follows:

- **conn.log**: A file containing information pertaining to all TCP/UDP/ICMP connections, this file contains most of the information gathered from the packet capture.
- **files.log**: A file consisting of analytic results of packets’ counts and sessions’ durations.
- **packet_filter.log**: A file listing the active filters applied to Zeek upon reading the packet capture file.
- **x509.log**: A file containing public key certificates used by protocols.
- **weird.log**: A file containing packet data non-conformant with standard protocols. It also contains packets with possibly corrupted or damaged packet header fields.
- **(protocol).log** (dns.log, dhcp.log, http.log, snmp.log): These are files containing information for packets found in each respective protocol. For instance, `dns.log` will only contain information generated by Domain Name Service (DNS) packets.

More information regarding log files is available in the Zeek official documentation, which can be viewed online through this link:


### 3.4 Basic viewing of Zeek logs

In this section we examine the generated log files and their contents.
Step 1. Use the following command to display the contents of the conn.log file using the `cat` command.

```
cat conn.log
```

The entire conn.log file will be displayed in the Terminal; however, the current formatting wraps around multiple lines, making it unclear and hard to understand. In the following section we introduce the `zeek-cut` utility for enhancing the output of these log files.

4 Analyzing Zeek log files

In this section, we review the utilities that help in creating files with well-formatted outputs.

4.1 Leveraging zeek-cut for a more refined view of log files

Although the produced log file is tab delimited, it is difficult to visualize and parse information from the terminal. The `zeek-cut` utility can be used to parse the log files by specifying which column data to be displayed in a more organized output.

4.1.1 Using zeek-cut in conjunction with the cat, head, and tail commands

Generally, the `zeek-cut` utility is typically coupled with `cat` using the pipe `|` command. In Linux, the pipe command allows sending the output of one command as input to another. Essentially, the output of a left command is passed as input to that on its right, and an unlimited number of commands can be chained together.
Step 1. Use the following command to pipe the contents of `cat` into `zeek-cut`.

```
cat conn.log | zeek-cut id.orig_h id.orig_p id.resp_h id.resp_p
```

The options passed into the `zeek-cut` utility represent the column headers to be extracted from the log file:

- `id.orig_h`: Column containing the source IP address.
- `id.orig_p`: Column containing the source port.
- `id.resp_h`: Column containing the destination IP address.
- `id.resp_p`: Column containing the destination port.

Alternatively, instead of using the `cat` command, the `head` command can be used to display the topmost rows of the log file, which can be very useful to view a large file’s contents.

Step 2. Use the following command to pipe the contents of `head` into `zeek-cut`.

```
head conn.log | zeek-cut id.orig_h id.orig_p id.resp_h id.resp_p
```
Notice that only two records are shown. This is caused by the `head` command taking the 10 topmost rows of `conn.log`, regardless of what that entails, and passing it as input to `zeek-cut`.

Since the log file contains 8 lines of header that explain its format, we will have to specify the first 18 rows of file when displaying the first 10 packets.

**Step 3.** Use the following command to pipe the contents of `head` into `zeek-cut`.

```bash
head -n 18 conn.log | zeek-cut id.orig_h id.orig_p id.resp_h id.resp_p
```

The `-n` option can be passed to the head utility to specify the desired number of rows.

### 4.1.2 Printing the output of `zeek-cut` to a text file

While the results displayed in the Terminal after using the `zeek-cut` utility can be easily viewed for smaller datasets, it is often necessary to save the output into a separate file. Using the `>` character, we can send the output to a new file for further processing by other applications.

**Step 1.** Use the following command to change the output location of `zeek-cut`.

```bash
cat conn.log | zeek-cut id.orig_h id.orig_p id.resp_h id.resp_p > output.txt
```

By including the file extension in `output.txt`, we are choosing to print the output into a plain text file.
Step 2. We can display the contents of the new `output.txt` file by using the `head` command.

```bash
head output.txt
```

4.1.3 Printing the output of zeek-cut to a csv file

In some situations, it is helpful to save the output of zeek-cut in a csv file. In a csv file, data may be imported into other applications, such as databases or machine learning classifiers.

Step 1. The exported output file by zeek-cut is tab-delimited due to the default zeek-cut settings. To export a file with another delimiter, the `-F` option is used.

```bash
cat conn.log | zeek-cut -F ',' id.orig_h id.orig_p id.resp_h id.resp_p > output.csv
```

Step 4. We can now display the contents of the `output.csv` file.

```bash
head output.csv
```
In conclusion, `zeek-cut` is a flexible tool that can be called to format Zeek log files depending on the user’s needs. The `zeek-cut` utility can be utilized with more advanced commands to further increase customization.

### 4.2 Closing the current instance of Zeek

After you have finished the lab, it is necessary to terminate the currently active instance of Zeek. Shutting down a computer while an active instance persists will cause Zeek to shut down improperly and may cause errors in future instances.

**Step 1.** Stop Zeek by entering the following command on the terminal. If required, type `password` as the password. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
cd $ZEEK_INSTALL/bin && sudo ./zeekctl stop
```

Concluding this lab, we have reviewed Zeek’s output log files in more depth while introducing some of the more relevant network-based log files and introduced some basic utilities to view these log files.

**References**

ZEEK INTRUSION DETECTION

Lab 3: Parsing, Reading and Organizing Zeek Files

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
## Contents

Overview ........................................................................................................................................... 3  
Objectives ........................................................................................................................................ 3  
Lab topology .................................................................................................................................... 3  
Lab settings ..................................................................................................................................... 3  
Lab roadmap .................................................................................................................................... 4  
1 Introduction to shell scripts ........................................................................................................... 4  
   1.1 Ubuntu Linux text editors ....................................................................................................... 5  
   1.2 Creating a shell script ........................................................................................................... 5  
2 Advanced zeek-cut log file analysis ............................................................................................... 8  
   2.1 Example 1 ............................................................................................................................. 8  
   2.2 Example 2 ............................................................................................................................ 10  
   2.3 Example 3 ............................................................................................................................ 11  
   2.4 Example 4 ............................................................................................................................ 12  
3 Incorporating the AWK scripting language for log file analysis ............................................. 14  
   3.1 Example 1 ............................................................................................................................. 14  
   3.2 Example 2 ............................................................................................................................ 15  
   3.3 Example 3 ............................................................................................................................ 17  
   3.4 Closing the current instance of Zeek .................................................................................. 19  
References ....................................................................................................................................... 19
Overview

This lab explains how to format and organize Zeek’s log files by combining zeek-cut utility with basic Linux shell commands. Utilities and tools introduced in this lab provide practical examples for logs customization in a real network environment.

Objectives

By the end of this lab, students should be able to:

1. Use Linux tools and commands for text files processing.
2. Practice Linux shell scripts and the AWK scripting language.
3. Incorporate AWK with zeek-cut to provide formatted logs.

Lab topology

Figure 1 shows the lab workspace topology. This lab primarily uses the Bro2 machine for offline packet capture processing and analysis.

![Lab topology diagram]

Figure 1. Lab topology.

Lab settings

The information (case-sensitive) in the table below provides the credentials to access the machines used in this lab.
Lab 3: Parsing, Reading and Organizing Zeek Files

Table 1. Device credentials for lab workspace.

<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@dmin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router</td>
<td>192.168.1.1</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.0.113.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Shell variables and their corresponding absolute paths.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/Zeek/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/Zeek/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/Zeek-Labs-Workspace/</td>
</tr>
</tbody>
</table>

Lab roadmap

This lab is organized as follows:

1. Section 1: Introduction to shell scripts.
2. Section 2: Advanced zeek-cut log file analysis.
3. Section 3: Incorporating the AWK scripting language for log file analysis.

1 Introduction to shell scripts

A shell script is a text file containing commands to be executed by the Unix command-line interpreter. Shell scripts provide a convenient way to manipulate files and automate programs’ executions. Selection and repetition are incorporated into scripts to branch control based on conditioning and looping statements. Running a shell script can
immensely save time and prevent manually entering repetitive commands in recurrent tasks.

1.1 Ubuntu Linux text editors

Linux-based distributions include pre-installed text editors like *nano*, *vi*, *vim*, *gedit*, etc. *nano* is a keyboard-oriented lightweight text editor with a simple Command Line Interface (CLI). Other editors such as *vi* and *vim* are highly customizable and extensible, making them attractive for users that demand a large amount of control and flexibility over their text editing environment. Alternatively, the Graphical User Interface (GUI) text editor *gedit* can be used to visually work outside of the terminal. More information on these text editors can be found on the Ubuntu help pages:


For simplicity, in this lab we use *nano* text editor to view, create and edit text files.

1.2 Creating a shell script

Shell scripts are effective in executing repetitive terminal commands. Unlike executing commands manually in the terminal, scripts can be saved and executed whenever needed simple by invoking their names. We begin by writing some basic shell scripts.

**Step 1.** From the top of the screen, click on the *Bro2* button as shown below to enter the *Bro2* machine.

![Bro1, DTN, Bro2, Client buttons]

**Step 2.** The *Bro2* machine will now open, and the desktop will be displayed. On the left side of the screen, click on the Terminal icon as shown below.
Step 2: In the Linux terminal, navigate to the lab workspace directory typing the following command. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key.

```
cd $ZEEK_LABS
```

Step 3: Use the Nano text editor to create the lab3script.sh file.

```
nano lab3script.sh
```

Step 4: Enter the following commands.

Once the text editor has opened, we will be able to enter the following commands. Each new line will denote a new Terminal command being passed. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key.

```
cd $ZEEK_INSTALL/bin
sudo ./zeekctl start
cd $ZEEK_LABS
zeek -r Sample-Captures/smallFlows.pcap
```
The file’s content is explained as follows:

- Line 1: changes the current directory to the Zeek’s installation directory.
- Line 2: starts a new instance of Zeek through `zeekctl`.
- Line 3: changes the current directory to the lab workspace.
- Line 4: invokes the `zeek` program with the `-r` option process the `smallFlows.pcap` capture file located in the `Sample-Captures` directory.

**Step 5**: When using Nano, the following keyboard shortcuts are used to save a file and then exit the workspace.

- `CTRL + o` — save the file
- `CTRL + x` — save and exit the file, return to terminal

After completing Step 4 and adding the correct commands with proper formatting, we will save and exit the text editor. Press `CTRL + o` and hit `Enter` to save the file’s contents, then `CTRL + x` to exit `nano` and return to the terminal.

**Step 6**: Use the following command to modify the permissions of the script file to make it executable. When prompted for a password, type `password` and hit `Enter`.

```
sudo chmod +x lab3script.sh
```

**Step 7**: Execute the `lab3script.sh` shell script by typing the following command.

```
./lab3script.sh
```
Step 8: Verify that the pcap file accessed on line 4 within the `lab3script.sh` shell script was processed successfully.

```
ls
```

The above output shows the list of log files generated which verifies the successful processing of the pcap file.

## Advanced zeek-cut log file analysis

This section introduces more advanced `zeek-cut` functionality to analyze packet capture statistics. These statistics can be used for planning and anomaly analysis. For instance, if a single port has been targeted and received a large number of network traffic, it may highlight a possible vulnerability. We can use the `zeek-cut` utility to determine if a host sends an abnormal number of packets to a specific destination and further analyze this event.

### Example 1

Example 1: Show the 10 source IP addresses that generated the most network traffic, organized in descending order.

To solve this example, we will be looking at the `id.orig_h` column because it contains the source IP addresses from the packet capture file.

Step 1: Open the `lab3script.sh` file with `nano` text editor.
Step 2: Modify the script file’s contents. Delete all the previous content and type the following command:

```
zeek -cut id.orig_h < conn.log | sort | uniq -c | sort -rn | head -n 10
```

Press \[CTRL + o\] and hit Enter to save the file’s contents, then \[CTRL + x\] to exit \texttt{nano} and return to the terminal. The above command is explained as follows:

- \texttt{zeek-cut id.orig_h < conn.log}: selects the id.orig.h column from the \texttt{conn.log} file.
- \texttt{sort}: uses the \texttt{sort} command to organize the rows in alphabetical order.
- \texttt{uniq -c}: uses the \texttt{uniq} command with the \texttt{-c} option to remove duplicates while returning unique instances and their counts.
- \texttt{sort -rn}: uses the \texttt{sort} command with the \texttt{-rn} option to organize the rows in reverse numerical order.
- \texttt{head -n 10}: uses the \texttt{head} command with the \texttt{-n} option to display the 10 topmost values.

Step 3: Execute the modified shell script.

```
./lab3script.sh
```
The number of duplicates is seen in the left column, while the matching source IP address is seen in the right column. Only 8 unique source addresses were found, and each was returned. From this output, we can conclude that the majority of network traffic was generated by the top 3 source IP addresses.

2.2 Example 2

Example 2: Show the 10 destination ports that received the most network traffic, organized in descending order.

To solve this example, we will be looking at the `id.resp_p` column because it contains the destination ports from the packet capture file.

Step 1: Open the `lab3script.sh` file with `nano` text editor.

```
nano lab3script.sh
```

Step 2: Modify the script file’s contents.

```
zeek-cut id.resp_p < conn.log | sort | uniq -c | sort -rn | head -n 10
```

Press `CTRL + O` and hit `Enter` to save the file’s contents, then `CTRL + X` to exit `nano` and return to the terminal. The above command is explained as follows:

- `zeek-cut id.resp_p < conn.log`: selects the `id.resp_p` column from the `conn.log` file.
- `| sort`: uses the `sort` command to organize the rows in alphabetical order.
- `| uniq -c`: uses the `uniq` command with the `-c` option to remove duplicates while returning unique instances and their counts.
- `| sort -rn`: uses the `sort` command with the `-rn` option to organize the rows in reverse numerical order.
- `| head -n 10`: uses the `head` command with the `-n` option to display the 10 topmost values.
Step 3: Execute the modified shell script.

```bash
./lab3script.sh
```

The number of duplicates is seen in the left column, while the matching destination port is seen in the right column. More than 10 unique destination ports were found, so only the top 10 were returned. From this output we can conclude that port 80 received the most traffic.

2.3 Example 3

Example 3: Show the number of connections per protocol service.

To solve this example, we will be looking at the `service` column because it contains the destination ports from the packet capture file.

Step 1: Open the `lab3script.sh` file with `nano` text editor.

```bash
nano lab3script.sh
```

Step 2: Modify the script file's contents.

```bash
zeek-cut service < conn.log | sort | uniq -c | sort -n
```
Press \texttt{CTRL + o} and hit \texttt{Enter} to save the file’s contents, then \texttt{CTRL + x} to exit \texttt{nano} and return to the terminal. The above command is explained as follows:

- \texttt{zeek-cut service < conn.log}: selects the \texttt{service} column from the \texttt{conn.log} file.
- \texttt{sort}: uses the \texttt{sort} command to organize the rows in alphabetical order.
- \texttt{uniq -c}: uses the \texttt{uniq} command with the \texttt{-c} option to remove duplicates while returning unique instances and their counts.
- \texttt{sort -n}: uses the \texttt{sort} command with the \texttt{-n} option to organize the rows in numerical order.

**Step 3:** Execute the modified shell script.

\texttt{./lab3script.sh}

The number of duplicates is seen in the left column, while the matching destination port is seen in the right column. From this output we can see that 331 packets did not have a marked protocol. This can be caused by a number of anomalies and is an example of how you can use the \texttt{zeek-cut} utility to return anomalies that require further identification.

### 2.4 Example 4

Example 4: Print the distinct browsers used by the hosts in this packet capture file to a separate file.

To solve this example, we will be looking at the \texttt{user_agent} column because it contains the destination ports from the packet capture file.

**Step 1:** Open the \texttt{lab3script.sh} file with \texttt{nano} text editor.
Lab 3: Parsing, Reading and Organizing Zeek Files

```
nano lab3script.sh
```

Step 2: Modify the script file’s contents.

```
zeek -cut user_agent < http.log | sort -u > browser.txt
```

Press `CTRL + o` and hit `Enter` to save the file’s contents, then `CTRL + x` to exit `nano` and return to the terminal. The above command is explained as follows:

- `zeek -cut user_agent < http.log` selects the `user_agent` column from the `http.log` file.
- `| sort -u > browser.txt` uses the `sort` command to sort the lines in the file and the `-u` option checks for strict ordering. The output is then saved into the `browser.txt` file.

Step 3: Execute the modified shell script.

```
./lab3script.sh
```

Step 4: Use a text editor to view the contents of the `browser.txt` file.

```
nano browser.txt
```

Step 5: View the distinct browser information.
Each browser found within the packet capture file is printed with related information extracted from the traffic by Zeek.

3 Incorporating the AWK scripting language for log file analysis

AWK is a terminal scripting language used to parse, filter and modify text files. AWK is specifically useful when processing rows and columns found in a Comma Separated Value (CSV) file. Additionally, AWK’s integrated string manipulation functions allow for the searching and modifying of specific output.

Like **cat** and **head** commands, AWK output can be piped into the **zeek-cut** utility, allowing more advanced parsing and formatting options. AWK reads each column in a file through its position. The first input column is accessed using $1 while the second column is accessed using $2 and so on. AWK also allows creating simple variables to store and read script values. AWK reads the input data as a loop, starting from the top of the file and finishing at the end of the file. Each row is considered an instance within the script.

3.1 Example 1

Example 1: Find the source and destination IP address of all UDP and TCP connections that lasted more than one minute.

**Step 1:** Open the **lab3script.sh** file with **nano** text editor.

```
nano lab3script.sh
```

**Step 2:** Modify the script file’s contents.
Lab 3: Parsing, Reading and Organizing Zeek Files

awk '$9 > 60' conn.log | zeek - cut id.orig_h id.resp_h

Press [CTRL + o] and hit Enter to save the file’s contents, then [CTRL + x] to exit nano and return to the terminal. The above command is explained as follows:

- awk `$9 > 60` conn.log selects the rows that have their 9th column value greater than 60 from the conn.log file. The 9th field represents the connection duration, and we are checking if the value is greater than 60 seconds (or 1 minute).
- `zeek-cut id.orig_h id.resp_h` returns the source and destination IP addresses.

**Step 3:** Execute the modified shell script.

`./lab3script.sh`

The source IP address is seen in the left column, while the matching destination IP address is seen in the right column. The pairs will only be displayed if the connection lasted at least one minute.

### 3.2 Example 2

Example 2: Show the top source host addresses in terms of total traffic (in bytes) sent in descending order.

The current directory contains an AWK script named `lab_sec3-2.awk`. This script will be used to solve this example and can be viewed with the following command.
The script is explained as follows. Each number represents the respective line number:

1. The `{' character is used to begin nested statements. This instance is the main functionality of the script.
2. The host variable, which will be used to store the source IP addresses found in the first column ($1), is checked against the current data entry in the column. If it is not equal, we will enter the next statement. Because we only want one instance of each source IP address, but the summed value of bytes sent, we will use this check to prevent duplicate entries.
3. This line contains a check to make sure the current packet is not empty and does contain a payload. If the current packet contains a payload of more than 0 bytes, we will proceed to line 4.
4. The current source IP address and its byte payload will be printed or returned to the next statements.
5. Now that we know the current source IP address is not yet stored in the host variable, we will create a new entry into the variable.
6. The size variable is reset back to zero
7. The `'}' character is used to end nested statements. Therefore, the first case of a source IP address not being contained in host is complete.
8. If the host variable contains the current data entry, we will proceed to line 9.
9. Here we will sum the unique source IP address’ total bytes by adding the payload from the second column ($2).
10. The `'}' character is used to end nested statements. This is the ending of the main functionality of the script.
11. The `END` statement denotes what the script will do once it has reached the end of the file, and there are no more input data rows to be read.
12. If a source IP address contains a total payload of more than 0 bytes, we will proceed to line 13.
13. AWK will return the source IP address found in the first column, as well as the size variable, containing the total payload in relation to that source IP address.

**Step 1:** Input the following command.

```
zeek-cut id.orig_h orig_bytes < conn.log | sort | awk -f lab3_sec3-2.awk | sort -k 2 | head -n 10
```

- `zeek-cut id.orig_h orig_bytes < conn.log`: selects the `id.orig_h` and `orig_bytes` columns from the `conn.log` file.
- `| sort`: uses the `sort` command to organize the rows in alphabetical order.
- `| awk -f lab3_sec3-2.awk`: will execute awk with the `-f` option to denote using the script find within the `lab3_sec3-2.awk` file.
- `| sort -k 2`: uses the `sort` command with the `-k` option to organize the rows based on the values found in the second column – the total number of bytes.
- `| head -n 10`: uses the `head` command with the `-n` option to display the 10 topmost values.

The left column contains the source IP address, while the right column contains the number of bytes produced by the paired source IP address.

### 3.3 Example 3

Example 3: Are there any web servers operating on non-standardized ports?

To solve this example, we will be looking at the `service` column to view the packets using the Hyper Text Transport Protocol (HTTP) protocol. The standard ports for the HTTP protocol are 80 and 8080, so we will be searching for the network traffic that does not reach those ports.

**Step 1:** Open the `lab3script.sh` file with `nano` text editor.

```
geedit lab3script.sh
```
Step 2: Delete the previous command and type the following command.

```bash
zeek-cut service id.resp_p id.resp_h < conn.log | awk 'S1 == "http" && ! ($2 == 80 || $2 == 8080) {print $3}' | sort -u
```

Press [CTRL + o] and hit Enter to save the file’s contents, then [CTRL + x] to exit nano and return to the terminal. The above command is explained as follows:

- `zeek-cut service id.resp_p id.resp_h < conn.log`: selects the service, id.resp_p and id.resp_h columns from the conn.log file.
- `| awk S1 == "http" && ! ($2 == 80 || $2 == 8080) {print $3}`: passes the input into the following AWK command:
  - `$1 == "http"`: performs a check on the first column to make sure the active data entry is running on the http service.
  - `&& ! ($2 == 80 || $2 == 8080)`: performs a second check if the first check is successfully passed. The ports will be checked and if they are not equal to either of the standard http ports (80 and 8080), they will be passed to the print statement.
  - `{print $3}`: prints the destination IP address of any host that passes both of the previous checks.
- `| sort -u`: uses the sort command to sort the lines in the file and the -u option checks for strict ordering.

Step 3: Execute the modified shell script.

```.lab3script.sh```
The destination IP addresses that received traffic on non-standardized ports are displayed.

### 3.4 Closing the current instance of Zeek

After you have finished the lab, it is necessary to terminate the currently active instance of Zeek. Shutting down a computer while an active instance persists will cause Zeek to shut down improperly and may cause errors in future instances.

**Step 1.** Stop Zeek by entering the following command on the terminal. If required, type `password` as the password. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```bash
cd $ZEEK_INSTALL/bin && sudo ./zeekctl stop
```

Concluding this lab, we have reviewed the process of creating shell scripts to be used for network analysis. We introduced more complex commands for the `zeek-control` utility, as well as used the AWK scripting language to retrieve information from Zeek log files.

**References**

ZEEK INTRUSION DETECTION SERIES

Lab 4: Generating, Capturing and Analyzing Network Scanner Traffic

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
Contents

Overview .................................................................................................................................................. 3
Objective .................................................................................................................................................. 3
Lab topology .............................................................................................................................................. 3
Lab settings ............................................................................................................................................... 3
Lab roadmap .............................................................................................................................................. 4
1 Introduction to Internet scanning and probing ...................................................................................... 4
2 Generating real time network scans ..................................................................................................... 5
  2.1 Starting a new instance of Zeek ......................................................................................................... 5
  2.2 Setting up the Bro2 machine for live network capture ...................................................................... 6
  2.3 Using the Bro1 machine for network scanning activities .................................................................... 7
    2.3.1 nmap options ............................................................................................................................... 8
    2.3.2 TCP SYN scans ........................................................................................................................... 8
    2.3.3 TCP connect scans ..................................................................................................................... 9
    2.3.4 TCP NULL scans ......................................................................................................................... 9
    2.3.5 TCP XMAS scans ....................................................................................................................... 10
    2.3.6 Terminating live network capture .............................................................................................. 11
3 Analyzing collected network traffic ...................................................................................................... 11
  3.1 Example Query 1 ............................................................................................................................... 12
  3.2 Example Query 2 ............................................................................................................................... 13
4 Detailing the importance of the Zeek interface topology ......................................................................... 14
  4.1 Scanning the Client machine ............................................................................................................ 14
  4.2 Scanning the DTN server ................................................................................................................ 16
  4.3 Closing the current instance of Zeek ................................................................................................. 17
References ................................................................................................................................................. 18
Overview

This lab is designed to provide an in-depth guide to scanning and probing network traffic. The lab demonstrates the generation of scan-based traffic and uses Zeek to process the collected traffic.

Objective

By the end of this lab, students should be able to:

1. Perform Internet scanning and probing events.
2. Utilize the Nmap software.
3. Generate and collect scan traffic.

Lab topology

Figure 1 shows the lab workspace topology. This lab primarily uses the Zeek1 machine to generate scan-based traffic, and the Zeek2 machine to perform live network capture.

Lab settings

The information (case-sensitive) in the table below provides the credentials to access the machines used in this lab.
### Table 1. Device credentials for lab workspace.

<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@dmin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router</td>
<td>192.168.1.1</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.0.113.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Shell variables and their corresponding absolute paths.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/Zeek/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/Zeek/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/Zeek-Labs-Workspace/</td>
</tr>
</tbody>
</table>

### Lab roadmap

This lab is organized as follows:

1. Section 1: Introduction to Internet scanning and probing.
2. Section 2: Generating real time network scans.
3. Section 3: Analyzing collected network traffic.
4. Section 4: Detailing the importance of the Zeek interface topology.

### 1 Introduction to Internet scanning and probing

Internet scanning is the process of generating crafted traffic used to identify active devices on a network. A variety of software utilities and tools are used to replicate scan-related traffic for testing purposes. These crafted packets can be both stealthy and
versatile. It is hard to determine scan-like activities when scanning traffic follows protocols’ standards and specifications.

Malicious scanning is a reconnaissance technique used to collect information about a target’s machine or network to facilitate an attack against it. Scanning is used by attackers to discover what ports are open, what services are running and identify system software, all to enable an attacker to more easily detect and exploit known vulnerabilities within a target machine.

This lab uses nmap, and its documentation can be found on the nmap website:

https://www.nmap.org/

nmap has a wide array of scan-related functionalities such as the customization of a scan’s transport protocol, ports, IP ranges, etc.

2 Generating real time network scans

Zeek’s default packet capture processing generates log files containing organized network traffic statistics. By leveraging Bro1’s machine to scan Bro2’s machine, we can better define and understand the steps it takes to both generate and capture scan traffic.

2.1 Starting a new instance of Zeek

Step 1. On the top of the lab workspace, click on the Bro2 button as shown below to enter the Bro2 machine.

Step 2. On the left side of the Bro2 desktop, click on the Terminal icon as shown below.
Step 3. Start Zeek by entering the following command on the terminal. This command enters Zeek’s default installation directory and invokes `zeekctl` tool to start a new instance. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key. When prompted for a password, type `password` and hit `Enter`.

```
cd $ZEEK_INSTALL/bin && sudo ./zeekctl start
```

A new instance of Zeek will now be active, and we can proceed to the next section of the lab.

### 2.2 Setting up the Bro2 machine for live network capture

**Step 1.** Navigate to the lab workspace directory.

```
cd $ZEEK_LABS/TCP-Traffic
```

**Step 2.** Start live packet capture on interface `ens33` and save the output to a file called `scantraffic.pcap`. To type capital letters, it is recommended to hold the `Shift` key while
typing rather than using the Caps key. Take notice that the 0 in the following command is the number zero.

```
sudo tcpdump -i ens33 -s 0 -w scantraffic.pcap
```

The Bro2 machine is now ready to begin collecting live network traffic. Next, we use the Bro1 machine to generate scan-based network traffic.

### 2.3 Using the Bro1 machine for network scanning activities

**Step 1.** On the top of the lab workspace, click on the Bro1 button as shown below to enter the Bro1 machine.

**Step 2.** On the left side of the Bro1 desktop, click on the Terminal button as shown below.

**Step 3.** On a machine running Linux, `nmap` is executed through the Terminal. Verify that `nmap` is functioning properly by viewing the currently installed version.

```
nmap -version
```
The figure above shows that the currently installed version of nmap is 7.60. With both the Bro2 and Bro1 machines configured correctly, we can proceed with the exercises.

### 2.3.1 nmap options

Nmap is used to discover hosts and services on a computer network by sending packets and analyzing the responses. The `nmap` command has a list of options for every scan type and covers several protocols. This lab focuses on TCP scans with their default settings. Two additional options are used during this lab are:

- `-A`: enables operating system and version detection.
- `-T4`: faster execution, can strain the initiator’s machine on larger scans.

More information is available on the following nmap documentation page. Because the lab environment is not connected to the Internet, viewing the documentation must be done on a personal computer not located within the lab topology.


The `nmap` tool should be invokes with the `sudo` command. When prompted for a password, type `password` and hit `Enter`.

### 2.3.2 TCP SYN scans

TCP SYN scan is one of the most common types of scans used for vulnerability detection. During SYN scanning, the initiating host sends a single TCP SYN packet to the destination. The receiving host interprets the request as a new TCP connection where the standard three-way TCP handshake is to be established. If a SYN/ACK packet is sent back, the initiator can infer that the port is open. The initiator can then send an RST (reset) packet to terminate the established connection.

**Step 1.** Use the following command to conduct a TCP SYN scan.

```
sudo nmap -sS 192.168.2.2
```
The `-sS` option is used to indicate a TCP SYN scan.

After the scan is completed, `nmap` produces a report on the performed scan. This includes the scan starting time, the number of ports, the total time, etc. We can see here the TCP SYN scan took 1.52 seconds, and none of the scanned ports were open.

### 2.3.3 TCP connect scans

TCP Connect scans are an alternative to TCP SYN scans. Rather than starting a TCP handshake, the initiator’s operating system attempts to establish a connection with the target victim through a system call. If a connection is successfully created, the initiator can infer that the receiver is open.

**Step 1.** Use the following command to conduct a TCP connect scan.

```bash
sudo nmap -sT 192.168.2.2
```

The `-sT` option is used to indicate a TCP Connect scan.

The report in the above figure shows that the scan was completed in 0.29 seconds, and none of the scanned ports were open.

### 2.3.4 TCP NULL scans
TCP NULL scans are another form of TCP scanning. In general, all TCP packets contain flags. Firewalls are configured to drop packets containing certain flags. The TCP NULL scan attempts to bypass these firewalls by excluding the header. With a sequence number of 0, packets in a TCP NULL scan will have no flags and can potentially infiltrate a network’s firewall.

**Step 1.** Use the following command to conduct a TCP NULL scan.

```
sudo nmap -sN 192.168.2.2
```

The `\-sN` option is used to indicate a TCP NULL scan.

The report in the above figure shows that the scan was completed in 1.50 seconds, and none of the scanned ports were open.

### 2.3.5 TCP XMAS scans

TCP Xmas scans, also known as Christmas tree scans, have their name derived from their set flags. In TCP Xmas scans, the PSH, URG and FIN flags are all set in the TCP header. This combination of flags is used in an attempt to infiltrate a strict network’s firewall.

**Step 1.** Use the following command to conduct a TCP XMAS scan.

```
sudo nmap -sX 192.168.2.2
```

The `\-sX` option is used to indicate a TCP XMAS scan.
The report in the above figure shows that the scan was completed in 1.51 seconds, and none of the scanned ports were open or vulnerable.

### 2.3.6 Terminating live network capture

**Step 1.** On the top of the lab workspace, click on the Bro2 button as shown below to enter the Bro2 machine.

**Step 2.** Within the terminal, use the `Ctrl+c` key combination to stop the live traffic capture session.

Statistics regarding the capture session will be displayed. 8045 packets were generated by these TCP-based scans. Note that the number of packets captured may vary per session and for the purpose of this lab, it is okay to continue.

### 3 Analyzing collected network traffic

After successfully conducting a number of TCP-based scans, the `scanpackets.pcap` packet capture file now contains all traffic. In this section we analyze the collected network traffic using Zeek.
**Step 1.** Use the following Zeek command to process the packet capture file.

```
zeek -C -r scantraffic.pcap
```

- The `-C` option indicates that Zeek will ignore corrupted checksums – which is highly probable as we are forcibly ending packet collection and did not initialize an absolute stopping point with the `tcpdump` utility.
- The `-r` option indicates that Zeek will be reading from an offline pcap file.

Because we only generated TCP packets, a warning will be displayed. This will not affect the log files or output.

**Step 2.** List the generated Zeek log files.

```
ls
```

With the log files generated, we can now use the `zeek-cut` utility for further analysis.

### 3.1 Example Query 1

Example 1: Show the source IP addresses that generated the most network traffic, organized in descending order.

**Step 1.** Enter the following command.

```
zeek-cut id.orig_h < conn.log | sort | uniq -c | sort -rn | head -n 10
```

The above command is explained as follows:

- `zeek-cut id.orig_h < conn.log` selects the `id.orig_h` column from the `conn.log` file.
• **sort**: uses the `sort` command to organize the rows in alphabetical order.
• **uniq -c**: uses the `uniq` command with the `-c` option to remove duplicates while returning unique instances and their counts.
• **sort -rn**: uses the `sort` command with the `-rn` option to organize the rows in reverse numerical order.
• **head -n 10**: uses the `head` command with the `-n` option to display the 10 topmost values.

We can see the majority of the packets were received from the *Bro1* machine denoted by the IP address 192.168.1.2.

### 3.2 Example Query 2

Example 2: Show the 10 destination ports that received the most network traffic, organized in descending order.

**Step 1.** Enter the following command.

```
zeek-cut id.resp_p < conn.log | sort | uniq -c | sort -rn | head -n 10
```

The above command is explained as follows:

• **zeek-cut id.resp_p < conn.log**: selects the `id.resp_p` column from the `conn.log` file.
• **sort**: uses the `sort` command to organize the rows in alphabetical order.
• **uniq -c**: uses the `uniq` command with the `-c` option to remove duplicates while returning unique instances and their counts.
• **sort -rn**: uses the `sort` command with the `-rn` option to organize the rows in reverse numerical order.
• **head -n 10**: uses the `head` command with the `-n` option to display the 10 topmost values.
The number of duplicates is seen in the left column, while the matching destination port is seen in the right column. More than 10 unique destination ports were found, so only the top 10 were returned.

4 Detailing the importance of the Zeek interface topology

In this section we review the aggregation of log files from scanning multiple devices within the same network. This section introduces the collection of traffic on machines other than the one hosting the Zeek interface. Zeek should be positioned at the entry point of a network to collect all incoming/outgoing traffic.

4.1 Scanning the Client machine

In this section we use the Bro1 machine to scan the Client machine which has the IP address 192.168.3.2.

Step 1. In Bro2 machine’s lab workspace directory, enter the TCP-Traffic directory typing the following command:

```bash
cd $ZEEK_LABS/TCP-Traffic
```

Step 2. Begin live packet capture on the ens33 interface and save the packet captures to a pcap file called scanclient.pcap. Take notice that the 0 in the following command is the number zero.

```bash
sudo tcpdump -i ens33 -s 0 -w scanclient.pcap
```
Step 3. Return to Bro1 machine and scan the Client machine using Nmap with the TCP SYN scan option \((-sS)\) activated.

```
sudo nmap -sS 192.168.3.2
```

The report in the above figure shows that the scan was completed in 4.14 seconds, and a list of open 6 ports with their related services are displayed.

Step 4. Return to the Bro2 machine and use the `Ctrl+c` key combination to stop the live traffic capture session.
Note that 2058 packets were collected. The Zeek interface running on the Bro2 machine was able to collect the network traffic that passed through it. Therefore, packets sent to the Client machine are also collected. Note that the number of packets captured may vary per session and for the purpose of this lab, it is okay to continue.

4.2 Scanning the DTN server

In this section we use the Bro1 machine to scan the DTN server machine which has the IP address 192.168.1.2.

**Step 1.** Navigate to the Bro2 machine’s lab workspace directory. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key.

```
cd $ZEEK_LABS/TCP-Traffic
```

**Step 2.** Begin live packet capture. Take notice that the 0 in the following command is the number zero. If necessary, type password as the password.

```
sudo tcpdump -i ens33 -s 0 -w scanserver.pcap
```

**Step 3.** Return to the Bro1 machine and scan the DTN server with the TCP SYN scan option (-sS) activated.

```
sudo nmap -sS 192.168.1.3
```
The report in the above figure shows that the scan was completed in 18.00 seconds, and that only one port was open – TCP port 21, hosting the File Transfer Protocol (FTP).

**Step 4.** Return to the Bro2 machine and use the Ctrl+c key combination to stop the live traffic capture session.

Note that 0 packets were collected. The Zeek interface running on the Bro2 machine was unable to collect the network traffic that did not pass through it, leading to no packets being collected. In a real-time network environment, having undetected pathways can lead to malicious activity going untraced.

### 4.3 Closing the current instance of Zeek

After you have finished the lab, it is necessary to terminate the currently active instance of Zeek. Shutting down a computer while an active instance persists will cause Zeek to shut down improperly and may cause errors in future instances.

**Step 1.** Stop Zeek by entering the following command on the terminal. If required, type `password` as the password. If the Terminal session has not been terminated or closed,
you may not be prompted to enter a password. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
cd $ZEEK_INSTALL/bin && sudo ./zeekctl stop
```

Concluding this lab, we have reviewed the steps required to generate scan traffic as well as enable live traffic capture using `zeek`. Once collected, the trace files can be studied, and empirical data can be investigated regarding the current state of a network and its devices. Lastly, we reviewed the importance of choosing which machine or server to have the Zeek interface running, as it is possible to leverage a network’s topology to circumvent the collecting interface.

**References**

ZEEK INTRUSION DETECTION SERIES

Lab 5: Generating, Capturing and Analyzing DoS and DDoS-centric Network Traffic

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
Lab 5: Generating, Capturing and Analyzing DoS and DDoS-centric Network Traffic

Contents

Overview .......................................................................................................................................................... 3
Objective ....................................................................................................................................................... 3
Lab topology ................................................................................................................................................. 3
Lab settings .................................................................................................................................................. 3
Lab roadmap ................................................................................................................................................ 4
1 Introduction to DoS and DDoS activity ........................................................................................................ 4
   1.1 DoS attack characteristics .................................................................................................................. 5
   1.2 DDoS attack characteristics ................................................................................................................. 5
2 Generating real-time DoS traffic .................................................................................................................. 5
   2.1 Starting a new instance of Zeek ......................................................................................................... 6
   2.2 Setting up the Bro2 machine for live network capture .................................................................. 7
   2.3 Launching LOIC ................................................................................................................................. 7
   2.4 Using the Bro1 machine to launch a TCP-based DoS attack ......................................................... 9
   2.5 Using the Bro1 machine to launch a UDP-based DoS attack ...................................................... 10
   2.6 Using the Bro1 machine to launch an ICMP-based DoS attack ................................................... 12
3 Analyzing collected network traffic .......................................................................................................... 14
   3.1 Analyzing TCP based traffic .......................................................................................................... 14
   3.2 Analyzing UDP based traffic .......................................................................................................... 15
   3.3 Analyzing ICMP based traffic ....................................................................................................... 16
   3.4 Closing the current instance of Zeek .............................................................................................. 17
References ..................................................................................................................................................... 18
Overview

This lab covers Denial of Service (DoS)-based network traffic. The lab introduces the generation of DoS-based traffic for testing purposes and uses Zeek to process the collected traffic.

Objective

By the end of this lab, students should be able to:

2. Experiment with the Low Orbit Ion Canon (LOIC) software.
3. Analyze collected DDoS traffic.

Lab topology

Figure 1 shows the lab workspace topology. This lab primarily uses the Zeek1 machine to generate scan-based traffic, and the Zeek2 machine to perform live network capture.

Lab settings

The information (case-sensitive) in the table below will be needed in order to complete the lab. The task sections below provide details on the use of this information.
<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@dmin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router</td>
<td>192.168.1.1</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.0.113.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/Zeek/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/Zeek/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/Zeek-Labs-Workspace/</td>
</tr>
</tbody>
</table>

**Lab roadmap**

This lab is organized as follows:

2. Section 2: Generating real-time DoS traffic.
3. Section 3: Analyzing collected network traffic.

1 **Introduction to DoS and DDoS activity**

Denial-of-Service (DoS) is an attack launched by a malicious user to render a target machine or network resource unavailable to its intended users. Distributed Denial-of-Service (DDoS) is an attack originated from different sources to flood the victim’s resources. A DDoS attack is more effective than a normal DoS and is harder to mitigate since unlike DoS, it is impossible to stop the attack simply by blocking a single source.
The different types of DoS attacks can be grouped by the traffic they generate, the bandwidth they consume, the services they disrupt, etc. Traffic-based DoS attacks aim at flooding the target with a large volume unsolicited traffic. Bandwidth-based DoS attacks involve transmitting a massive amount of junk data to overload the victim and render its network equipment congested.

1.1 **DoS attack characteristics**

DoS attacks generally involve flooding a targeted victim with network traffic to cause a crash and make it unavailable to benign users. In this lab we explore two common DoS attacks:

- **SYN flood**: an attacker attempts to overwhelm the server machine by sending a constant stream of TCP connection requests, forcing the server to allocate resources for each new connection until all resources are exhausted.  
- **ICMP flood**: the attacker abuses ICMP Ping and floods the victim computer with Echo Request messages. When a computer receives an ICMP Echo Request message it responds with an ICMP Echo Reply message.

1.2 **DDoS attack characteristics**

DDoS attacks involve using a large number of devices to flood a victim. With an increased number of exploited machines, the amount of resources available to the attacker is far higher. Some relevant DDoS attacks are:

- **HTTP flood**: simple attack but requires a large number of resources. An attacker who controls several devices (botnet) can continually flood a server with HTTP requests until the server becomes unavailable and unable to respond to additional incoming requests.  
- **SYN flood**: similar to the DoS SYN flood, a botnet initiates several sessions without completing a TCP handshake, causing the victim to consume its available resources.  
- **Amplification attack**: attackers abuse UDP-based network protocols to launch DDoS attacks that exceed hundreds of Gbps in traffic volume. This is achieved via reflective DDoS attacks where an attacker does not directly send traffic to the victim but sends spoofed network packets to a large number of systems that reflect the traffic to the victim. Domain Name System (DNS) and Network Time Protocol (NTP) are examples of application-layer protocols that act as potential amplification attack vectors.

DoS and DDoS attacks can cause catastrophic fallout and monetary losses to a victim.

2 **Generating real-time DoS traffic**
This lab uses the Low Orbit Ion Canon (LOIC), open-source network stress testing and DoS attack generator. LOIC can be found in the following Github repository. Because the lab environment is not connected to the Internet, viewing the documentation must be done on a personal computer not located within the lab topology.

https://github.com/NewEraCracker/LOIC

Similar to the nmap utility, LOIC can be used to replicate DoS or DDoS activity for testing purposes. LOIC has a Graphical User Interface (GUI), which facilitates the attack’s customization.

In this lab, Zeek’s default packet capture processing will generate log files containing organized network traffic statistics. In this section, Bro2 machine is used for live capture and Bro1 machine is used to generate DoS-related traffic.

2.1 Starting a new instance of Zeek

Step 1. On the top of the lab workspace, click on the Bro2 button as shown below to enter the Bro2 machine.

Step 2. On the left side of the Bro2 desktop, click on the Terminal icon as shown below.

Step 3. Start Zeek by typing the following command on the terminal. This command navigates to Zeek’s default installation directory and invokes zeekctl tool to start a new instance. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key. When prompted for a password, type password and hit Enter.
A new instance of Zeek will now be active, and we can proceed to the next section of the lab.

### 2.2 Setting up the Bro2 machine for live network capture

**Step 1.** Navigate to the lab workspace directory and enter the `TCP-Traffic` directory. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```bash
cd $ZEEK_LABS/TCP-Traffic
```

**Step 2.** Start live packet capture on interface `ens33` and save the output to a file called `tcprtraffic.pcap`. Take notice that the 0 in the following command is the number zero. If necessary, type `password` as the password.

```bash
sudo tcpdump -i ens33 -s 0 -w tcprtraffic.pcap
```

The *Bro2* machine is now ready to begin collecting live network traffic. Next, we use the *Bro1* machine to generate DoS-based network traffic.

### 2.3 Launching LOIC
Step 1. On the top of the lab workspace, click on the Bro1 button as shown below to enter Bro1 machine.

Step 2. On the left side of the Bro1 desktop, click on the Terminal icon as shown below.

Step 3. Navigate to the Documents directory.

```
cd Documents/
```

Step 4. Execute the loic.sh shell script by entering the following command in the terminal.

```
./loic.sh run
```
The figure above shows the LOIC interface. Important features highlighted with colored boxes are explained as follows:

1. **Red Box:** target IP address. After entering an IP address, clicking the *Lock on* button will select the IP as the target destination address.
2. **Green Box:** target port. Can be changed depending on which method is used to launch the DoS attack.
3. **Yellow Box:** target method. Can be changed to define which protocol is used to launch the DoS attack.
4. **Blue Box:** number of threads. Indicates the amount of resources LOIC will allocate on the host machine.
5. **Purple Box:** number of sockets per thread. Increasing the number of sockets per thread will exponentially increase the speed of the DoS attack; however, it also requires more resources on the host machine.
6. **Brown Box:** packet payload. Used to define what each packet will contain as payload.
7. **Orange Box:** start button. After customizing a desired attack, this button is used to launch the attack.

### 2.4 Using the Bro1 machine to launch a TCP-based DoS attack

**Step 1:** Customize the DoS attack by entering the following values in their respective input boxes.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>192.168.2.2</td>
</tr>
<tr>
<td>Port</td>
<td>80</td>
</tr>
<tr>
<td>Method</td>
<td>TCP</td>
</tr>
<tr>
<td>Threads</td>
<td>20</td>
</tr>
<tr>
<td>Sockets</td>
<td>25</td>
</tr>
<tr>
<td>Payload</td>
<td>TCP TEST</td>
</tr>
</tbody>
</table>
Step 2. Click the Lock on button to save the current configurations. Click the Start (IMMA CHARGIN MAH LAZER) button to begin the DoS attack. Wait for 10 seconds and click the Stop (Stop flooding) button to stop the DoS attack.

Step 3. Return to the Bro2 machine and use the Ctrl+c key combination to stop the live traffic capture session.

Within the 10 seconds timeframe, 1,028,840 packets were generated and collected. This number of packets verifies that DoS attacks generate an immense amount of network traffic. Note that the number of packets captured may vary per session and for the purpose of this lab, it is okay to continue.

2.5 Using the Bro1 machine to launch a UDP-based DoS attack

Step 1. In the Bro2 machine, navigate to the lab workspace directory and enter the UDP-Traffic directory. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key.

cd $ZEEK_LABS/UDP-Traffic
Step 2. Start live packet capture on interface ens33 and save the output to a file called udptraffic.pcap. Take notice that the 0 in the following command is the number zero. If required, type `password` as password.

```
sudo tcpdump -i ens33 -s 0 -w udptraffic.pcap
```

Step 3. Return to the *Bro1* machine to customize the DoS attack. Enter the following values in their respective input boxes.

<table>
<thead>
<tr>
<th>IP</th>
<th>192.168.2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port</td>
<td>20</td>
</tr>
<tr>
<td>Method</td>
<td>UDP</td>
</tr>
<tr>
<td>Threads</td>
<td>20</td>
</tr>
<tr>
<td>Sockets</td>
<td>25</td>
</tr>
<tr>
<td>Payload</td>
<td>UDP TEST (Must be changed before updating Method feature)</td>
</tr>
</tbody>
</table>

Step 4. Click the **Lock on** button to save the current configurations. Click the **Start (IMMA CHARGIN MAH LAZER)** button to begin the DoS attack. Wait for 10 seconds and click the **Stop (Stop flooding)** button to stop the DoS attack.
Step 5. Return to the Bro2 machine and use the `Ctrl+c` key combination to stop the live traffic capture session.

Within the 10 second time frame, 208,952 packets were generated and collected.

2.6 Using the Bro1 machine to launch an ICMP-based DoS attack

Step 1. While the Bro2 machine is active, navigate to the lab workspace directory and enter the ICMP-Traffic directory.

```
cd $ZEEK_LABS/ICMP-Traffic
```

Step 2. Start live packet capture on interface ens33 and save the output to a file called `udptraffic.pcap`.

```
sudo tcpdump -i ens33 -s 0 -w icmptraffic.pcap
```

Step 3. Return to the Bro1 machine to customize the DoS attack. Enter the following values in their respective input boxes.

- IP: 192.168.2.2
- Port: 0
- Method: ICMP
- Threads: 20
- Sockets: 25
- Payload: ICMP TEST (Must be changed before updating Method feature)
Step 4. Click the *Lock on* button to save the current configurations. Click the *Start (IMMA CHARGIN MAH LAZER)* button to begin the DoS attack. ICMP-based attacks require each socket to successfully connect before transmitting packets. Therefore, this step will take longer than the previous. Wait until all 20 threads have connected.

Step 5. Wait for 15 seconds and click the *Stop (Stop flooding)* button to stop the DoS attack.

Step 6. Return to the *Bro2* machine and use the `Ctrl+c` key combination to stop the live traffic capture session.

Within the 15 seconds timeframe, 84 packets were generated and collected – significantly less than the number generated during the simulated TCP and UDP DoS attacks. Note
that the number of packets captured may vary per session and for the purpose of this lab, it is okay to continue.

So far, we have launched TCP, UDP and ICMP DoS attacks. From the results we can see that the number of packets received for TCP and UDP DoS were exponentially larger compared against scan-based traffic explored in the previous lab.

3 Analyzing collected network traffic

We now use the `zeek-cut` utility commands to display the capture traffic.

3.1 Analyzing TCP based traffic

Step 1. In the *Bro2* machine, navigate to the lab workspace directory and enter the *TCP-Traffic* directory. To type capital letters, it is recommended to hold the *Shift* key while typing rather than using the *Caps* key.

```
cd $ZEER_LABS/TCP-Traffic
```

Step 2. Process the `tcptraffic.pcap` packet capture file using Zeek. The `-r` option indicates that Zeek will be reading from an offline pcap file, and the `-C` is used to disable checksum verification. Note that the execution time of this command depends on the size of the pcap file.

```
zeek -C -r tcptraffic.pcap
```

Because we only generated TCP packets, a warning will be displayed. This will not affect the log files or output.

Step 3. Show the destination ports that received traffic the most.
We can see that 649,273 packets were received by the Bro2 machine on port 80, which is the port we specified for the Bro1 machine to target. Additional ports may be discovered during processing, slightly variable due to LOIC attempting to establish connections; however, it is clear the most targeted port is the one we specified in the DoS attack. Note that the results may vary per session and for the purpose of this lab, it is okay to continue.

### 3.2 Analyzing UDP based traffic

**Step 1.** Navigate to the lab workspace directory and enter the **UDP-Traffic** directory. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
cd $ZEEK_LABS/UDP-Traffic
```

**Step 2.** Process the `udptraffic.pcap` packet capture file using Zeek. The `-r` option indicates that Zeek will be reading from an offline pcap file, and the `-C` is used to disable checksum verification.

```
zeek -C -r udptraffic.pcap
```

**Step 3.** Show the list of ports that received network traffic.

```
cat conn.log | zeek-cut id.resp_p
```

We can see that despite the large number of packets collected, very few were recorded by Zeek’s event-based engine. We specified port 20 as the targeted port during our DoS attack; however, the number of identified packets is significantly lower than expected.

The primary cause of the decreased packet count is due to the number of UDP packets being dropped. Primarily due to firewalls, UDP packets may be traced on the interface, but may not reach the target destination. Furthermore, the default Zeek customization is primarily focused on TCP traffic, and is not designed to handle UDP traffic in such an in-depth manner, requiring additional scripts and policies that will be introduced in later labs.

### 3.3 Analyzing ICMP based traffic

**Step 1.** Navigate to the lab workspace directory and enter the `ICMP-Traffic` directory. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
cd $ZEEK_LABS/ICMP-Traffic
```
Step 2. Process the `icmptraffic.pcap` packet capture file using Zeek. The `-r` option indicates that Zeek will be reading from an offline pcap file, and the `-C` is used to disable checksum verification.

```
zeek -C -r icmptraffic.pcap
```

Step 3. Show the list of ports that received network traffic.

```
cat conn.log | zeek-cut id.resp_p
```

Similar to the UDP packet capture file, we can see that despite the number of packets collected, very few were recorded by Zeek’s event-based engine. Only one instance of traffic recorded on Port 0 was recorded, and it is the port we specified for the `Bro1` machine to target. Additional Zeek event streams can be developed with the Zeek scripting language to further increase the depth of Zeek’s ICMP processing.

3.4 Closing the current instance of Zeek

After you have finished the lab, it is necessary to terminate the currently active instance of Zeek. Shutting down a computer while an active instance persists will cause Zeek to shut down improperly and may cause errors in future instances.

Step 1. Stop Zeek by entering the following command on the terminal. If required, type `password` as password. If the Terminal session has not been terminated or closed, you may not be prompted to enter a password. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
cd $ZEEK_INSTALL/bin && sudo ./zeekctl stop
```

Concluding this lab, we have introduced DoS and DDoS events, as well as generated and captured Dos traffic in the lab workspace environment. Networks require some form of denial-of-service mitigation or prevention tools since attacks can devastate unsecured networks.

References

ZEEK INTRUSION DETECTION SERIES

Lab 6: Introduction to Zeek Scripting

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
Overview

This lab covers Zeek’s scripting language. It introduces the major keywords and components required in a Zeek script. The lab then uses these scripts to analyze processed log files.

Objectives

By the end of this lab, students should be able to:

1. Develop scripts using Zeek’s scripting language.
2. Analyze processed log files using Zeek scripts.
3. Modify log streams for creating additional events and notices.

Lab topology

Figure 1 shows the lab workspace topology. This lab primarily uses the Zeek2 machine for offline Zeek script development and offline packet capture processing and analysis.

Lab settings

The information (case-sensitive) in the table below provides the credentials to access the machines used in this lab.
Table 1. Device credentials for lab workspace.

<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@admin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router</td>
<td>192.168.1.1</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.0.113.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Shell variables and their corresponding absolute paths.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/Zeek/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/Zeek/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/Zeek-Labs-Workspace/</td>
</tr>
</tbody>
</table>

Lab roadmap

This lab is organized as follows:

1. Section 1: Introduction to scripting with Zeek.
2. Section 2: Log file analysis using Zeek scripts.
3. Section 3: Modifying Zeek log streams.
1 Introduction to scripting with Zeek

Zeek includes its own event-driven scripting language which provides the primary means for an organization to extend and customize Zeek’s functionality. By modifying Zeek’s log streams, a more in-depth analysis can be performed on network events.

Since Zeek’s scripting language is event-driven, we define which events we need Zeek to respond to when encountered during network traffic analysis.

1.1 Zeek script events

The script below shows events that will be explored during this lab. When developing a Zeek script, the script’s functionalities are wrapped within respective events.

```zeek
1  event zeek_init(){
2      /* code */
3  }
4  event zeek_done(){
5      /* code */
6  }
7  event tcp_packet(){
8      /* code */
9  }
10 event udp_request(){
11     /* code */
12  }
13 event udp_reply(){
14     /* code */
15  }
```

- `zeek_init` event: activated when Zeek is first initialized.
- `zeek_done` event: activated before Zeek is terminated.
- `tcp_packet` event: activated when a packet containing a TCP header is processed.
- `udp_request` event: activated when a packet containing a UDP request header is processed.
- `udp_reply` event: activated when a packet containing a UDP reply header is processed.

Additional events and their required parameters are outlined and explained in Zeek’s official documentation.

1.2 Zeek module workspace

The script below uses the `module` keyword which assigns the script to a `namespace`. Codes from other scripts can be accessed by including a matching module. The `export` keyword is used to export the code entered in its block with the module workspace.
• **module ZeekScript**: changes the module workspace to ZeekScript.

• **export** block: code entered here will be exported with the module workspace.

Exporting code with a module workspace allows more advanced scripts to be built on top of other scripts.

### 1.3 Zeek log streams

The script below shows the log stream functionality. When developing a Zeek script, all processed outputs will be sent to a specific log stream. These log streams will contain the format of the corresponding log file output. We can create new streams, modify original streams or append additional parameters to existing streams.

```zeek
1  event connection_established() {
2    Log::create_stream(LOG, format, path);
3    Log::write(Logstream, data);
4  }
```

• **connection_established** event: activated when a host makes a connection to a receiver.

• **Log::create_stream**: creates a new log stream, will a name, format structure and path.

• **Log::write**: writes included data to the specified log stream.

Additional log stream commands are explained in detail in Zeek’s official documentation.

### 2 Log file analysis using Zeek scripts

With Zeek’s event-driven scripting language, we can create specific event-based filters to be applied during packet capture analysis. This section shows example scripts for network analysis.

#### 2.1 Starting a new instance of Zeek

**Step 1.** On the top of the lab workspace, click on the Bro2 button as shown below to enter the Bro2 machine.
Step 2. On the left side of the Bro2 desktop, click on the Terminal icon as shown below.

![Terminal Icon]

Step 3. Start Zeek by entering the following command on the terminal. This command enters Zeek’s default installation directory and invokes `zeekctl` tool to start a new instance. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key. When prompted for a password, type password and hit Enter.

```bash
cd $ZEEK_INSTALL/bin && sudo ./zeekctl start
```

A new instance of Zeek will now be active, and we can proceed to the next section of the lab.

### 2.2 Executing a UDP Zeek script

Step 1. Navigate to the workspace directory. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key.

```bash
cd $ZEEK_LABS
```
Step 2. Display the content of the `lab6_sec2-2.zeek` Zeek script using `nl` command. `nl` shows the line numbers in the file.

```bash
nl lab6_sec2-2.zeek
```

The script is explained as follows. Each number represents the respective line number:

1. Event `udp_request` activated when a packet containing a UDP Request header is processed. The related packet header information is stored in the connection data structure passed to the function through the `$id$` variable.
2. Prints the specified string. `%s` is a format specifier for strings with `fmt`. It indicates the position of the corresponding variable’s information in the string. `$id$resp_h` retrieves the destination IP address from the UDP packet.
3. End of the `udp_request` event.
4. Event `udp_reply` activated when a packet containing a UDP Reply header is processed. The related packet header information is stored in the connection data structure passed to the function through the `$id$` variable.
5. Prints the specified string. `$id$resp_h` retrieves the destination IP address from the UDP packet.
6. End of the `udp_reply` event.

Step 3. Process a packet capture file using the Zeek script.

```bash
zeek -r Sample-Captures/smallFlows.pcap lab6_sec2-2.zeek
```
The packet capture file is processed into output log files. Since we did not create a new log stream, the script’s output is displayed on the standard output (the screen). When `udp_request` or `udp_reply` events are triggered, the resulting packet information is displayed.

### 2.3 Executing a TCP Zeek script

**Step 1.** Display the content of the `lab6_sec2-3.zeek` Zeek script using `nl` command. `nl` shows the line numbers in the file.

```
$ nl lab6_sec2-3.zeek
```

The script is explained as follows. Each number represents the respective line number:

1. Event `tcp_packet` activated when a packet containing a TCP header is processed. The related packet header information is stored in the connection data structure passed to the function through the `$u` variable. Additional TCP-related information is passed in a similar manner.
2. Prints the specified string. `%s` is a format specifier for strings with `fmt`. It indicates the position of the corresponding variable’s information in the string. `$u$id$` retrieves the destination IP address from the TCP packet.
3. End of the `tcp_packet` event.

**Step 2.** Process a packet capture file using the Zeek script.

```zeek
.zeek -r Sample-Captures/smallFlows.pcap lab6_sec2-3.zeek
```

The following output is produced:

When the `tcp_packet` event is triggered, the resulting packet information is displayed. Highlighted is an example of Port 80 and Port 443 traffic.

These examples highlight Zeek’s capabilities of tracking specific traffic. For instance, a script can be designed to collect all Port 80 traffic daily and to export it to a log file. In the following section we introduce log streams.

### 3 Modifying Zeek log streams

Zeek log streams determine where an event’s output will be returned, as well as how it is formatted. It is possible to append new streams, modify default streams, or remove streams.
Before continuing, we must clear the lab workspace directory.

**Step 1.** Display the contents of the *lab_clean.sh* shell script using `nl` command.

```bash
nl lab_clean.sh
```

The shell script removes a list of files expected to be generated by Zeek’s processing using default log streams. Executing this shell script will clear the directory of log files generated previously. Output messages from running this script are not displayed in the Terminal, instead the code `>` /dev/null `2>&1` will set errors and notices to be sent to a null folder, effectively eliminating them.

**Step 2.** Execute the *lab_clean.sh* shell script. If required, type ```password``` as the password.

```bash
./lab_clean.sh
```

With the workspace directory cleared, we can move to the next section.

### 3.1 Renaming the conn.log stream

In this example, we will rename the *conn.log* file to be *UpdatedConn.log*. Renaming log streams can help with files organization, especially if a log file has been modified from its original functionality.

**Step 1:** Display the contents of the *lab6_sec3-1.zeek* Zeek script using the `nl` command.

```bash
nl lab6_sec3-1.zeek
```
The script is explained as follows. Each number represents the respective line number:

1. Event **zeek_init** activated when Zeek is first initialized.
2. Creates a local variable **update** initialized to the default **Conn::LOG** filter.
3. Sets the **update** variable’s path to **UpdatedConn.log**.
4. Appends the new filter to the active log streams.
5. End of the **zeek_init** event.

**Step 2.** Process a packet capture file using the Zeek script.

```
zeek -r Sample-Captures/smallFlows.pcap lab6_sec3-1.zeek
```

**Step 3.** List the generated log files in the current directory.

```
ls
```

Note the **UpdatedConn.log** file in the fourth column. Since we did not change any formatting, it is an exact replica of the original **conn.log** file.

### 3.2 Updating the conn.log stream
In this example, we modify the `conn.log` file to generate an additional `conn-http.log` file. This modification will split the `conn.log` contents between two log files, which is useful in organizing specific events.

**Step 1.** Execute the included `lab_clean.sh` shell script. If required, type `password` as the password.

```
./lab_clean.sh
```

**Step 2.** Display the contents of `lab6_sec3-1.zeek` Zeek script using the `nl` command.

```
nl lab6_sec3-2.zeek
```

The script is explained as follows. Each number represents the respective line number:

1. Boolean function that has the parameter `rec`, an instance of Conn::Info.
2. Returns True if the service stored in `rec` is the HTTP protocol.
3. End of the function.
4. Event [zeek_init](#) activated when Zeek is first initialized.
5. Creates a local filter with `http` related naming and pathing.
6. Appends the new filter to the active log streams.
7. End of the [zeek_init](#) event.

**Step 2:** Process a packet capture file using the Zeek script.

```
zeek -r Sample-Captures/smallFlows.pcap lab6_sec3-2.zeek
```
Step 3: List the generated log files in the current directory.

```bash
ls
```

Note the `conn-http.log` file in the first column. This file will have the same formatting as the `conn.log` file; however, it will only contain HTTP traffic.

### 3.3 Closing the current instance of Zeek

After you have finished the lab, it is necessary to terminate the currently active instance of Zeek. Shutting down a computer while an active instance persists will cause Zeek to shut down improperly and may cause errors in future instances.

**Step 1.** Stop Zeek by entering the following command on the terminal. If required, type `password` as the password. If the Terminal session has not been terminated or closed, you may not be prompted to enter a password. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```bash
cd $ZEEK_INSTALL/bin && sudo ./zeekctl stop
```

Concluding this lab, we have introduced the Zeek scripting language. Using event-driven functionality, Zeek scripts can be used to customize the output log streams. Besides
renaming existing files, you can also split the files to generate a more protocol or event-specific log file.

References

ZEEK INTRUSION DETECTION SERIES

Lab 7: Advanced Zeek Scripting for Anomaly and Malicious Event Detection

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
Contents

Overview .................................................................................................................................................. 3
Objectives.................................................................................................................................................. 3
Lab topology ................................................................................................................................................. 3
Lab settings .................................................................................................................................................. 3
Lab roadmap .................................................................................................................................................. 4
1 Zeek’s default anomaly detection scripts .............................................................................................. 5
  1.1 Zeek scan-event .................................................................................................................................. 5
  1.2 Zeek bruteforce-event ....................................................................................................................... 6
2 Generating customized malicious network traffic .................................................................................. 7
  2.1 Starting a new instance of Zeek ......................................................................................................... 7
  2.2 Setting up the Bro2 machine for live network capture ...................................................................... 8
  2.3 Using the Bro1 machine to launch customized TCP-based scans .................................................. 9
3 Applying Zeek scripts to filter network traffic ...................................................................................... 11
  3.1 Applying the ZeekScanDetection filter ............................................................................................. 12
  3.2 Applying the ScanFilter filter .......................................................................................................... 14
  3.3 Closing the current instance of Zeek .................................................................................................. 18
References .................................................................................................................................................. 19
Overview

This lab covers Zeek’s scripting language and introduces more advanced scripting capabilities. This lab simulates a new zero-day scanning technique and explains a Zeek script that captures this new event. The lab is designed to further highlight the customization properties of Zeek scripting.

Objectives

By the end of this lab, students should be able to:

1. Use precompiled Zeek scripts for identifying network traffic anomalies.
2. Develop a Zeek script for identifying and organizing specific malicious traffic events.
3. Generate customized malicious traffic to be used for testing purposes.

Lab topology

Figure 1 shows the lab workspace topology. This lab primarily uses the Zeek2 machine for offline Zeek script development and offline packet capture processing and analysis.

![Lab topology diagram]

Figure 1. Lab topology.

Lab settings

The information (case-sensitive) in the table below provides the credentials to access the machines used in this lab.
Lab 7: Advanced Zeek Scripting for Anomaly and Malicious Event Detection

<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@dmin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router</td>
<td>192.168.1.1</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.0.113.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Device credentials for lab workspace.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/Zeek/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/Zeek/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/Zeek-Labs-Workspace/</td>
</tr>
</tbody>
</table>

Table 2. Shell variables and their corresponding absolute paths.

Lab roadmap

This lab is organized as follows:

1. Section 1: Zeek’s default anomaly detection scripts.
2. Section 2: Generating customized malicious network traffic.
3. Section 3: Applying Zeek scripts to filter network traffic.
# Zeek’s default anomaly detection scripts

Zeek’s scripting language can be used to identify and report network anomalies by using event-driven functions. This section introduces two default Zeek script filters that are installed by default after Zeek installation.

While these default Zeek scripts might not correctly identify every unique anomaly, they provide a comprehensive starter code that can be customized further for anomaly-based detection.

## 1.1 Zeek scan-event

The first default Zeek script is the `scan.zeek` script. More information on this script can be found in Zeek’s documentation pages.


The file has been copied into the Zeek lab workspace directory and renamed to `ZeekScanDetection.zeek` for ease of access and name-reference clarity.

This Zeek script is used to identify scan-related traffic. Internet scanning can be split into three main categories:

1. Vertical Scanning: an attacker scans many ports on a single destination host address.
2. Horizontal Scanning: an attacker scans a single port on many destination host addresses.
3. Block Scanning: an attacker interweaves vertical and horizontal scanning techniques to increase complexity and become harder to track.

The script shown in the figure below list the first few lines of the `ZeekScanDetection.zeek` file.

```plaintext
1 #!TCP Scan detection
2 # ..Authors: Sheharbano Khattak
3 #    Seth Hall
4 #    All the authors of the old scan.bro
5 @load base/frameworks/notice
6 @load base/frameworks/sumstats
7 @load base/utils/time
```

As shown in the figure above, loading other scripts is done through the `@load` statement with the following format:

```plaintext
@load <zeekscriptfile>
```

Lines 5, 6 and 7 include the functionalities found within the export blocks of the respectively included Zeek scripts.
The script leverages thresholds to determine if scan-like activities are present when processing network capture. If all the thresholds are exceeded, traffic is inferred to be scan-related.

For real time deployment, these thresholds will need to be modified dependent on the network size. For instance, a smaller network containing less IP addresses will need a lower threshold of scan packets to identify a scan-event. However, modifying these thresholds may result in an increase of false positives and true negatives, so it highly recommended to simulate and test network traffic before modification.

The figure above shows the thresholds in the ZeekScanDetection.zeek file. The thresholds are explained as follows. Each number represents the respective line number:

28. `const addr_scan_interval = $min & redef;` threshold to check a source IP address for varying destination IP address scan-related traffic. The default interval is 5 minutes.

32. `const port_scan_interval = $min & redef;` threshold to check a source IP address for varying destination port scan-related traffic. The default interval is 5 minutes.

35. `const addr_scan_threshold` threshold of unique destination IP addresses that a single host attempts to contact. The default threshold is 25 unique destination IP addresses.

38. `const port_scan_threshold` threshold of unique destination ports that a single host attempts to contact. The default threshold is 15 unique destination ports.

### 1.2 Zeek bruteforce-event

The second default Zeek script is the detect-bruteforcing.zeek script. More information on this script can be found in Zeek’s documentation pages.


The file has been copied into the Zeek lab workspace directory and renamed to ZeekBruteforceDetection.zeek for ease of access and name-reference clarity.
This Zeek script is used to identify brute-force password attacks. Brute-force attacks can be identified by several failed login attempts. This denotes that an attacker is attempting to systematically submit credentials until the correct credentials are found. The motivation behind this attack is to gain authorized access to an account, machine or server.

The script leverages the following thresholds to determine if scan-like activities are present when processing network capture. During real time deployment, these thresholds should be modified depending on the network size. The number of failed login attempts (or duration) should be modified to increase the script’s accuracy.

```
1  ## FTP brute-forcing detector, triggering when too many rejected usernames or
2  ## failed passwords have occurred from a single address.
3  @load base/protocols/ftp
4  @load base/frameworks/sumstats
5  @load base/utils/time
6  module FTP;
7  * export {
8    * redef enum Notice::Type += {
9      *     # Indicates a host bruteforcing FTP logins by watching for too
10     *     # many rejected usernames or failed passwords.
11     *     Bruteforcing
12     *   };
13     *   ## How many rejected usernames or passwords are required before being
14     *   ## considered to be bruteforcing.
15     *   const bruteforce_threshold: double = 20 &redef;
16     *   ## The time period in which the threshold needs to be crossed before
17     *   ## being reset.
18     *   const bruteforce_measurement_interval = 15mins &redef;
19   }
```

The thresholds are explained as follows. Each number represents the respective line number:

15. `const bruteforce_threshold`: threshold for the number of failed authentications attempts a source IP address can make. The default value is 20 failed attempts within the related time interval threshold.

18. `const bruteforce_measurement_interval`: threshold for the time to check a source IP address for failed authentication attempts. The default interval is 15 minutes.

## Generating customized malicious network traffic

This section introduces creating and using a new Zeek script, tailored to react to more specific events.

### 2.1 Starting a new instance of Zeek

**Step 1:** On the top of the lab workspace, click on the *Bro2* icon as shown below to enter the *Bro2* machine.
Step 2: On the left side of the Bro2 desktop, click on the Terminal button as shown below.

![Terminal button on Bro2 desktop](image)

Step 3. Start Zeek by entering the following command on the terminal. This command enters Zeek’s default installation directory and invokes `zeekctl` tool to start a new instance. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key. When prompted for a password, type `password` and hit `Enter`.

```
cd $ZEEK_INSTALL/bin && sudo ./zeekctl start
```

A new instance of Zeek will now be active, and we can proceed to the next section of the lab.

2.2 Setting up the Bro2 machine for live network capture

Step 1: Navigate to the lab workspace directory and enter the `TCP-Traffic` directory. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
cd $ZEEK_LABS/TCP-Traffic
```
Step 2. Start live packet capture on interface `ens33` and save the output to a file named `ntraffic.pcap`. Take notice that the 0 in the following command is the number zero. If necessary, type `password` as the password.

```
sudo tcpdump -i ens33 -s 0 -w ntraffic.pcap
```

The `Bro2` machine is now ready to begin collecting live network traffic. Next, we use the `Bro1` machine to generate unique scan-based network traffic.

### 2.3 Using the `Bro1` machine to launch customized TCP-based scans

In this section we use the `nmap` software to generate TCP-based scan traffic. TCP flags are included within the packets’ headers.

We will replicate a `zero-day` exploit within our subnet’s firewall. By including specific TCP flag combinations, packets can pass the subnet’s firewall unobstructed.

Step 1. On the top of the lab workspace, click on the `Bro1` button as shown below to enter the `Bro1` machine.

Step 2. On the left side of the `Bro1` desktop, click on the Terminal button as shown below.
Step 3. Launch a TCP connect scan against the Client machine. If necessary, type password as the password.

```
sudo nmap -sT 192.168.3.2
```

Step 4. Launch a scan against the Client machine with the SYN, FIN and RST flags set. We will label this scan as Case1.

```
sudo nmap --scanflags SYN,FIN,RST 192.168.3.2
```

By specifying the `--scanflags` option, we can control which TCP flags are included in the packet header.
**Step 5.** Launch a scan against the *Client* machine with the SYN, RST and ACK flags set. We will label this scan as Case2.

```bash
sudo nmap --scanflags SYN,FIN,RST 192.168.3.2
```

**Step 6.** Return to the *Bro2* machine and use the `Ctrl+c` key combination to stop the live traffic capture session.

The capture session’s results show that 6050 packets were captured. We will now process the *nttraffic.pcap* file using Zeek while including Zeek filters to organize and identify these new *zero-day* exploits. Note that the number of packets captured may vary per session and for the purpose of this lab, it is okay to continue.

### 3 Applying Zeek scripts to filter network traffic
Now that we have collected traffic containing the zero-day exploits, we will process the packet capture file using Zeek.

### 3.1 Applying the ZeekScanDetection filter

**Step 1:** Process the `ntraffic.pcap` packet capture file while in the `TCP-Traffic` directory. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key. Note that using the `tab` key to autofill the directory may cause errors in finding the `ZeekScanDetection.zeek` filter and it is recommended to manually type the entire file path.

```
zeek -C -r ntraffic.pcap /$ZEEK_LABS/ZeekScanDetection.zeek
```

**Step 2:** Display the contents of the `notice.log` file using the `cat` command.

```
cat notice.log
```
Within the notice.log file, we can see the Bro1 machine has been identified for creating scan-based network traffic and exceeding the 15-ports threshold configured earlier.

**Step 3:** Display the contents of the conn.log file using the following command.

```
head -n 25 conn.log | zeek-cut ts id.orig_h id.orig_p id.resp_h id.resp_p history
```

The Terminal command is explained as follows:

- `head -n 25 conn.log`: returns the top 25 rows of the conn.log file, specified by the `-n` option.
- `| zeek-cut ts id.orig_h id.orig_p id.resp_h id.resp_p history`: uses the `zeek-cut` utility to return the specified columns and remove padding.

The `history` column (last column in the figure above) contains information regarding which TCP flags were found within a packet header:

- 😄: SYN flag.
- 😄: SYN+ACK flags.
- 😄: ACK flag.
- 😄: FIN flag.
- 😄: RST flag.
- 😄: URG flag.
- 😄: Multiple flags set.

The event is attributed to the host when the flag letter is uppercase; otherwise, it is attributed to the receiver.
3.2 Applying the ScanFilter filter

Step 1: Navigate to the lab workspace directory.

```bash
cd $ZEEK_LABS
```

Step 2: Display the contents of the `ScanFilter.zeek` file using `nl`.

```bash
nl ZeekFilter.zeek
```

The script is explained as follows. Each number represents the respective line number:

1. Declares a new module workspace.
2. Export block allows code to be accessed outside the current module workspace.
3. Creates and appends the `CASE1LOG` to the list of Log files.
4. Creates and appends the `CASE2LOG` to the list of Log files.
5. Block that includes all the columns and features to be included in these new log files. Each will contain a variable type and output location:
   - `ts`: time that the packet was received.
   - `id`: packet identification number.
   - `orig_h`: source IP address.
- `orig_p`: source port.
- `resp_h`: destination IP address.
- `resp_p`: destination port.
- `history`: string of flag characters.

16. Initialization event.  
17. Creates a new log stream using the previously introduced `CASE1LOG` LOG ID, column formatting and a file name path.  
18. Creates a new log stream using the previously introduced `CASE2LOG` LOG ID, column formatting and a file name path.  
19. Event triggered when a TCP packet is processed.  
20. Creates a local variable `rec` to store the column-related information, using the current packet data, accessed with the `$id$<column>` format.  
21. Checks if the SFR flag combination is present in the packet. This relates to the history column, containing SYN-FIN-RST flags.  
22. If the SFR flag combination is present, the packet will be written to the `CASE1LOG` log stream with the packet information passed through the local variable `rec`.  
23. Checks if the SRA flag combination is present in the packet. This relates to the history column, containing SYN-RST-ACK flags.  
24. If the SRA flag combination is present, the packet will be written to the `CASE2LOG` log stream with the packet information passed through the local variable `rec`.  

**Step 3:** Process the `ntraffic.pcap` packet capture file while in the lab workspace directory.

```
zeek -r TCP-Traffic/ntraffic.pcap ScanFilter.zeek
```
Step 4: List the generated log files in the current directory.

```
ls
```

Note the `Case1.log` and `Case2.log` files generated by including the `ScanFilter.zeek` filter during processing.

Step 5: View the contents of the `Case1.log` file.

```
head -n 25 Case1.log | zeek-cut ts id.orig_h id.orig_p id.resp_h id.resp_p history
```
The Terminal command is explained as follows:

- `head -n 25 Case1.log`: returns the top 25 rows of the `conn.log` file, specified by the `-n` option.
- `| zeek-cut ts id.orig_h id.orig_p id.resp_h id.resp_p history`: uses the `zeek-cut` utility to only return the specified columns, and removes padding.

Unlike the default example, we can see the `history` column contains the exact same flag. Our filter was successful in organizing the traffic related to the `Case1` exploit.

**Step 6:** Display the contents of the `Case2.log` file.

```
head -n 25 Case2.log | zeek-cut ts id.orig_h id.orig_p id.resp_h id.resp_p history
```
The Terminal command is explained as follows:

- `head -n 25 Case2.log`: returns the top 25 rows of the `conn.log` file, specified by the `-n` option.
- `| zeek-cut ts id.orig_h id.orig_p id.resp_h id.resp_p history`: uses the `zeek-cut` utility to only return the specified columns, and removes padding.

Unlike the default example, we can see the `history` column contains the exact same flag. Our filter was successful in organizing the traffic related to the `Case2` exploit.

### 3.3 Closing the current instance of Zeek

After you have finished the lab, it is necessary to terminate the currently active instance of Zeek. Shutting down a computer while an active instance persists will cause Zeek to shut down improperly and may cause errors in future instances.

**Step 1.** Stop Zeek by entering the following command on the terminal. If required, type `password` as the password. If the Terminal session has not been terminated or closed, you may not be prompted to enter a password. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
cd $ZEEK_INSTALL/bin && sudo ./zeekctl stop
```
Concluding this lab, we introduced default frameworks for anomaly-detection scripts. We generated malicious network traffic to simulate a zero-day exploit, and then processed the traffic using a customized a Zeek script. With the resulting Zeek log files, these exploits can be studied for additional analysis and mitigation.

References

ZEEK INTRUSION DETECTION SERIES

Lab 8: Preprocessing of Zeek Output Logs for Machine Learning

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
Contents

Overview ............................................................................................................................. 3
Objective .............................................................................................................................. 3
Lab topology ......................................................................................................................... 3
Lab settings ......................................................................................................................... 3
   Lab roadmap .................................................................................................................... 4
1 Introduction to machine learning in network security ......................................................... 5
   1.1 ARFF file format .......................................................................................................... 5
2 Aggregating network capture datasets .............................................................................. 6
   2.1 Starting a new instance of Zeek .................................................................................. 7
   2.2 Setting up the Bro2 machine for live network capture .............................................. 7
   2.3 Using the Bro1 machine to launch a TCP-based scans ............................................. 8
3 Preprocessing of Zeek log files ....................................................................................... 10
   3.1 Preprocessing the malicious dataset .......................................................................... 10
   3.2 Preprocessing of the benign dataset .......................................................................... 13
   3.3 Creation of the test and training datasets .................................................................... 15
   3.4 Adding the .arff file headers .................................................................................... 17
   3.5 Closing the current instance of Zeek ......................................................................... 19
References .......................................................................................................................... 19
Lab 8: Preprocessing of Zeek Output Logs for Machine Learning

Overview

This lab introduces the application of machine learning in the network security field. After using Zeek's scripting language to generate anomaly-based output files, it is necessary to format these datasets to be used by machine learning classifiers.

Objective

By the end of this lab, students should be able to:

1. Explain the benefits of leveraging machine learning for network analysis.
3. Aggregate and preprocess a dataset to be used by a machine learning classifier.

Lab topology

Figure 1 shows the lab workspace topology. This lab primarily uses the Zeek2 machine for offline Zeek log file processing and reformatting.

Lab settings

The information (case-sensitive) in the table below will be needed in order to complete the lab. The task sections below provide details on the use of this information.
Table 1. Device credentials for lab workspace.

<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@admin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>Router</td>
<td>192.168.1.1</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.0.113.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Shell variables and their corresponding absolute paths.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/ZEEK/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/ZEEK/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/ZEEK-Labs-Workspace/</td>
</tr>
</tbody>
</table>

**Lab roadmap**

This lab is organized as follows:

1. Section 1: Introduction to machine learning in network security.
2. Section 2: Aggregating network capture datasets.
3. Section 3: Preprocessing of Zeek log files.
1 Introduction to machine learning in network security

Machine learning is programming computers to optimize a performance criterion using example data or past experience\(^1\). Machine learning is particularly useful for computing empirical correlations, and in cases where it is difficult to write a computer program to solve a given problem. In recent years, technological advances in machine learning have propelled its application on various domains and sectors. Cyber-security is a critical area in which machine learning (ML) is increasingly becoming significant.

By using Zeek and text processing languages, it is possible to identify the presence of an anomaly. Once an anomaly is detected, Zeek’s scripts can be implemented to extract relevant fields and build a dataset.

In this lab, we train machine learning classifiers using these anomaly-based datasets in order to build a model which can be used for future predictions.

This lab focuses on reformatting Zeek log files into Attribute-Relation File Format (ARFF) files, to be used by Weka software. Weka is a workbench for machine learning that is intended to help in the application of machine learning techniques to a variety of real-world problems\(^2\).

Supervised learning is a common approach used in machine learning. Supervised learning consists of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). When training a machine learning classifier using supervised learning, it is important to include both a training and test dataset:

- **Training dataset**: dataset used by the classifier to “learn” correlations and feature weights. Data should include instances of both variable and control group, while containing a classification label.
- **Testing dataset**: dataset used by the classifier to test accuracy. If the classifier is able to accurately predict labels for the training dataset but not for the testing dataset, then it is necessary to adjust and retrain the classifier.

1.1 ARFF file format

The Weka software contains a variety of different machine learning algorithms to train a number of classifiers. Each classifier will require different datasets; for instance, decision trees can only handle numeric or nominal values, and strings cannot be used as an input without being listed nominally.

The majority of machine learning classifiers accept numeric data inputs. Therefore, we will need to preprocess our log file datasets to contain only numeric and nominal data. Additionally, Weka requires each input dataset to be formatted in an .arff file format.
ARFF files contain comma-separated values and additional headers and labels. Below is a sample of a properly formatted .arff file that we will be developing in this lab.

```
@RELATION networktraffic

@ATTRIBUTE time NUMERIC
@ATTRIBUTE sourceip NUMERIC
@ATTRIBUTE destip NUMERIC
@ATTRIBUTE sourceport NUMERIC
@ATTRIBUTE destport NUMERIC
@ATTRIBUTE protocol {tcp, udp, icmp}
@ATTRIBUTE duration NUMERIC
@ATTRIBUTE class {1, 0}

@DATA
1361916108568086,17216133116,17216139250,53244,5440,tcp,0006155,0
136191610729437,1721613337,17216139250,60719,5440,tcp,0005408,0
136191610681641,1721613354,17319747104,64846,80,tcp,115325661,0
136191610877712,17216133116,17216139250,53322,5440,tcp,0005775,0
1361916230836068,1721613344,23049244,64843,80,tcp,15533492,0
1361916103539053,1721613397,17216139250,59315,5440,tcp,0005457,0
1361916103713440,1721613393,17216139250,60929,5440,tcp,0004760,0
13619161916967378,1721613387,964314622,59818,443,tcp,255274202,0
1361916432301319,17216128109,17216133254,4782,161,udp,0241787,0
1361916107865575,1721613397,17216139250,59340,5440,tcp,0015299,0
1361916430354514,172161336,8868,50578,53,udp,0064045,0
1361916191105910,1721613348,16417319013,60566,0,udp,05995783,0
1361916363440100,1721613382,17216139250,61595,5440,tcp,0005439,0
```

The ARFF file headers can be summarized as follows:

- **@RELATION**: name of the dataset.
- **@ATTRIBUTE**: specifies the label and the data type for each column:
  - NUMERIC: integer data type.
  - NOMINAL: values match entries defined within the brackets `{}`.
- **@DATA**: lists the input data.

Now that we have introduced ARFF files and understand what an input dataset should look like, we can start aggregating and preprocessing a dataset using Zeek.

2. **Aggregating network capture datasets**

To create our dataset, we need to make sure there is a certain level of entropy in the data to guarantee that the machine learning classifier will learn properly. Therefore, we need to combine both benign and malicious datasets.

In this lab, we use the `smallFlows.pcap` file as the control group, identified as benign traffic with a class label of 0. We then generate a new `maltraffic.pcap` file to be used as the variable group, identified as malicious traffic with a class label of 1.
2.1 Starting a new instance of Zeek

**Step 1:** On the top of the lab workspace, click on the *Bro2* button as shown below to enter the *Bro2* machine.

![Bro2 button](image)

**Step 2:** On the left side of the *Bro2* desktop, click on the Terminal icon as shown below.

![Terminal icon](image)

**Step 3.** Start Zeek by typing the following command on the terminal. This command navigates to Zeek’s default installation directory and invokes `zeekctl` tool to start a new instance. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key. When prompted for a password, type `password` and hit `Enter`.

```
cd $ZEEK_INSTALL/bin && sudo ./zeekctl start
```

A new instance of Zeek will now be active, and we can proceed to the next section of the lab.

2.2 Setting up the Bro2 machine for live network capture

**Step 1:** Navigate to the lab workspace directory and enter the lab workspace directory. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.
Step 2. Start live packet capture on interface ens33 and save the output to a file called `maltraffic.pcap`. Take notice that the 0 in the following command is the number zero. If necessary, type `password` as the password.

```
sudo tcpdump -i ens33 -s 0 -w maltraffic.pcap
```

The *Bro2* machine is now ready to begin collecting live network traffic. Next, we use the *Bro1* machine to generate unique scan-based network traffic.

### 2.3 Using the *Bro1* machine to launch a TCP-based scans

Similar to the previous labs, we will be using the `nmap` software to generate TCP-based scan traffic. For this section, we will be specifying specific TCP flags to be contained within the packet header.

**Step 1.** On the top of the lab workspace, click on the *Bro1* button as shown below to enter the *Bro1* machine.

**Step 2.** On the left side of the *Bro1* desktop, click on the Terminal button as shown below.
Step 3. Launch a TCP SYN scan against the Bro2 machine. If necessary, type `password` as the password.

```
sudo nmap -sS 192.168.2.2
```

Step 4. Return to the Bro2 machine and use the `Ctrl+c` key combination to stop the live traffic capture session.

The figure above shows that 2028 packets were collected. Since the `smallFlows.pcap` file is already downloaded, we now have both the malicious and benign datasets. In the following section we will begin formatting our datasets into ARFF files. Note that the
number of packets captured may vary per session and for the purpose of this lab, it is okay to continue.

3 Preprocessing of Zeek log files

To generate ARFF files, we first need to process our packet capture files using Zeek’s default configuration.

In a real-time environment, at this stage you may include anomaly-specific scripts. Once an anomaly has been processed by Zeek, the resulting log files will need to be reformatted.

Afterwards, we need to select which features we wish to extract from the Zeek log files to be used in our training and testing datasets. It is important to carefully select the relevant features when training a classifier. If features are not strategically selected, classifiers may create unreliable correlations which may lead to poor accuracy in the detection process. In this lab we extract a small number of general packet features.

3.1 Preprocessing the malicious dataset

Step 1. Process the maltraffic.pcap file.

```
zeek -C -r maltraffic.pcap
```

Step 2. Display the contents of the conn.log file.

```
column -s, -t conn.log | less -#2 -N -S
```

Examining the previous command:

- `column -s, -t conn.log`: calls the `column` utility to read and columnize the file contents of the `conn.log` file. The `-s` option specifies the separator and the `-t` option enables the output to be created as a table.
• `less -#2 -N -S` accepts the output of the column utility and calls the `less` utility. The `-#2` specifies the default number of positions to scroll horizontally in the RIGHTARROW and LEFTARROW keys, the `-N` option marks each row with a line number and the `-S` option causes the display to remove any data that would not fit on the current Terminal screen rather than overflowing to a new line.

The previous command results in the following output.

![Terminal output showing Zeek log files](image)

We can see in the previous image that the `conn.log` file is nowhere near the `.arff` file format. We will need to remove the Zeek padding, column names, change the tab delimiter and remove excess column features.

Press the `q` key on your keyboard to exit and return to the Terminal.

**Step 3:** Display the contents of `lab8_malicious.sh` shell script using the `nl` command.

```
nl lab8_malicious.sh
```

![Terminal output showing command result](image)
The script is explained as follows. Each number represents the respective line number:

1. Using the `cat` utility, the contents of the `conn.log` file will be passed into the `zeek-cut` utility to remove the log file header and only include the specified columns. The output of the `zeek-cut` utility will be saved to a new file named `packet.csv`. The feature columns we will be using to train our example machine learning classifier are:

- `ts`: time the packet was received.
- `id.orig_h`: source IP address.
- `id.resp_h`: destination IP address.
- `id.orig_p`: source port.
- `id.resp_p`: destination port.
- `proto`: transport protocol.
- `duration`: connection or session length.

2. Using the `cat` utility, the contents of the `packet.csv` file will be passed into the `tr` utility. The `tr` utility will replace the `packet.csv` file’s tab-delimited structure with a comma-delimited structure, and the output will be saved to a new file named `packet2.csv`.

3. Using the `sed` utility, all instances of a period `.` will be removed. This will allow for the IP addresses to be input as a numeric data type rather than a string, and the output will be saved to a new file named `packet3.csv`.

4. Using the `sed` utility, all instances of a dash `-` will be replaced by `?`. Currently, when a column is empty, Zeek writes a dash `-`. However, Weka reads question marks `?` as an empty column. The output will be saved to a new file named `packet4.csv`.

5. Using the `awk` utility, every row will have an additional `,1` appended to the end of the row. This will represent the class label; we used `1` to denote the malicious traffic. The output will be saved to a new file named `malicious.csv`.

6. The file contents of `malicious.csv` will be displayed. This command is introduced in the Step 1 of this subsection.

**Step 4:** Execute the `lab8_malicious.sh` shell script. If prompted for a password, type `password` and hit `Enter`.

```
./lab8_malicious.sh
```

After executing all commands in the script, the `malicious.csv` file contents will be displayed on the Terminal as shown in the figure below.
We can see from the above image that the malicious.csv file is now properly formatted to fit in the DATA section of an ARFF file. Each row contains an equal number of comma-delimited columns with only numeric characters.

Press the key on your keyboard to exit and return to the Terminal.

Now that we have our malicious dataset created, we can begin formatting our benign dataset.

**Step 5:** Execute the lab_clean.sh shell script. If prompted for a password, type password and hit Enter.

```
./lab_clean.sh
```

3.2 Preprocessing of the benign dataset

**Step 1:** Process the smallFlows.pcap file using the zeek -r command.

```
zeek -r Sample-Captures/smallFlows.pcap
```
Step 2: Display the contents `lab8_benign.sh` shell script using the `nl` command.

```
nl lab8_benign.sh
```

With the exception of Line 5, the script is exactly the same as the one explained in Step 3 of the previous section.

Line 5 has been modified to append `,0` to the end of each row. This value represents the benign class label. The output will be saved to a new file named `benign.csv`.

Step 3: Execute the `lab8_benign.sh` shell script.

```
./lab8_benign.sh
```

After executing all commands in the script, the `benign.csv` file contents will be displayed on the Terminal as shown in the figure below.
We can see from the above image that the benign.csv file is now properly formatted to fit in the @DATA section of an ARFF file. Each row contains an equal number of comma-delimited columns with only numeric characters.

Press the `q` key on your keyboard to exit and return to the Terminal.

Now that we have our both of our datasets created, we are ready to combine them into the training and test input datasets.

### 3.3 Creation of the test and training datasets

**Step 1:** Combine the malicious.csv and benign.csv files into the dataset.csv file.

```bash
cat malicious.csv benign.csv > dataset.csv
```

The dataset.csv file will now contain the benign.csv rows appended to the end of the malicious.csv rows. We now need to randomize the file contents and apply further formatting by executing the lab8_create_sets.sh shell script.

**Step 2:** Display the contents of lab8_create_sets.sh shell script using the `nl` command.
The script is explained as follows. Each number represents the respective line number:

1. Using the `shuf` utility, the contents of the `dataset.csv` file will be shuffled, and the output will be saved to a new file named `randomized.csv`.
2. Using the `head` utility, the top 300 rows from the `randomized.csv` file were saved to a new file named `test.csv`.
3. Using the `sed` utility, rows 1-300 are removed from the `randomized.csv` file and the output is saved to the new `trainset.arff` file.
4. Using the `sed` utility, the last column of the `test.csv` file is removed. We are removing the label of each instance of the test dataset so that we can have the classifier attempt to predict these labels. The output is saved to the new `testset.arff` file.
5. Using the `wc` utility, the number of rows within the `testset.arff` file are displayed. We can compare this value against the value found in Line 8 to make sure no packet data was lost.
6. Using the `wc` utility, the number of rows within the `trainset.arff` file are displayed. We can compare this value against the value found in Line 7 to make sure no packet data was lost.

**Step 3:** Execute the `lab8_create_sets.sh` shell script.

The figure above shows the line count of the `testset.arff` and `trainset.arff` files. The `testset.arff` file contains 300 rows while the `trainset.arff` file contains 1401 rows. The `trainset.arff` file size may be variable due to the number of packets generated during the original TCP SYN scans; however, the `testset.arff` file should always be equal to 300 rows due to the executed script.
Now that we have generated our testing and training .arff files, the final step for preprocessing the Zeek datasets is to add the .arff file headers to each file.

### 3.4 Adding the .arff file headers

**Step 1:** Using the `nano` text editor, open the `trainset.arff` file for editing.

```bash
nano trainset.arff
```

**Step 2:** Prepend the following headers to the `trainset.arff` file. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```arff
@RELATION networktraffic

@ATTRIBUTE time NUMERIC
@ATTRIBUTE sourceip NUMERIC
@ATTRIBUTE destip NUMERIC
@ATTRIBUTE sourceport NUMERIC
@ATTRIBUTE destport NUMERIC
@ATTRIBUTE protocol {tcp, udp, icmp}
@ATTRIBUTE duration NUMERIC
@ATTRIBUTE class {1,0}

@DATA
```

```bash
156191969960814,19216813,19216822,49526,1755,tcp,0000003,1
1295981663537961,17216255,1892644128,50983,3192,udp,0259354,0
12959816488591455,17216255,2049613158,10638,80,tcp,0150006,0
156191969995516,19216813,19216822,49526,9002,tcp,0000003,1
1561919699872844,19216813,19216822,49526,15660,udp,0000004,0
156191969995584,19216813,19216822,49526,2607,udp,0000004,0
1295981648201900,17216255,25525525525,68,67,udp,7,0
```

The input training dataset is now a properly formatted .arff file and can be input into a machine learning algorithm to train a classifier.

**Step 3:** Using the **nano** text editor, open the `testset.arff` file for editing.

```bash
nano testset.arff
```

**Step 2:** Prepend the following headers to the `testset.arff` file.

The headers are the same as those added to the `trainset.arff` file, so they can be copied and pasted directly into the `testset.arff` file. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```plaintext
@RELATION networktraffic

@ATTRIBUTE time NUMERIC
@ATTRIBUTE sourceip NUMERIC
@ATTRIBUTE destip NUMERIC
@ATTRIBUTE sourceport NUMERIC
@ATTRIBUTE destport NUMERIC
@ATTRIBUTE protocol {tcp, udp, icmp}
@ATTRIBUTE duration NUMERIC
@ATTRIBUTE class {1,0}

@DATA
```

```bash
nano testset.arff
```

```
@RELATION networktraffic

@ATTRIBUTE time NUMERIC
@ATTRIBUTE sourceip NUMERIC
@ATTRIBUTE destip NUMERIC
@ATTRIBUTE sourceport NUMERIC
@ATTRIBUTE destport NUMERIC
@ATTRIBUTE protocol {tcp, udp, icmp}
@ATTRIBUTE duration NUMERIC
@ATTRIBUTE class {1,0}

@DATA
```

```
admin@bro2:~/Zeek-Labs-Workspace
GNU nano 2.9.3
```

The input test dataset is now a properly formatted `.arff` file and can be input into a machine learning classifier to test the classifier’s accuracy.

### 3.5 Closing the current instance of Zeek

After you have finished the lab, it is necessary to terminate the currently active instance of Zeek. Shutting down a computer while an active instance persists will cause Zeek to shut down improperly and may cause errors in future instances.

**Step 1.** Stop Zeek by entering the following command on the terminal. If required, type `password` as the password. If the Terminal session has not been terminated or closed, you may not be prompted to enter a password. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
   cd $ZEEK_INSTALL/bin && sudo ./zeekctl stop
```

![Terminal screenshot](image.png)

**References**

ZEEK INTRUSION DETECTION SERIES

Lab 9: Developing Machine Learning Classifiers for Anomaly Inference and Classification

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
Contents

Overview ........................................................................................................................................ 3
Objectives ....................................................................................................................................... 3
Lab topology .................................................................................................................................. 3
Lab settings ................................................................................................................................... 3
Lab roadmap .................................................................................................................................... 4
1  Introduction to Weka .................................................................................................................. 4
   1.1  Starting Weka ....................................................................................................................... 5
2  Importing a dataset into Weka .................................................................................................... 7
   2.1  Loading the training dataset ................................................................................................. 8
   2.2  Filtering the training dataset ................................................................................................ 10
   2.3  Training a decision table classifier ....................................................................................... 14
   2.4  Training a decision tree classifier ......................................................................................... 16
       2.4.1  Updating the decision tree classifier ............................................................................. 18
3  Reviewing the classifier’s predictions on a test dataset ............................................................... 23
   3.1  Saving the decision table ....................................................................................................... 23
   3.2  Using the classifier to predict labels for the test dataset ..................................................... 24
   3.3  Viewing the predicted labels for the test dataset ................................................................. 27
References ...................................................................................................................................... 30
Overview

This lab introduces the application of machine learning in the network security field. The lab explains how to generate a decision table and decision tree to infer scan-related network traffic. The lab is designed to train and test a machine learning classifier using network traffic dataset.

Objectives

By the end of this lab, students should be able to:

1. Train a decision table to classify scan-related network traffic.
2. Train a decision tree to classify scan-related network traffic.
3. Test and modify the trained classifiers and review their output classifications on a test dataset.

Lab topology

Figure 1 shows the lab workspace topology. This lab primarily uses the Zeek2 machine for offline Zeek log file processing and reformatting.

Lab settings

The information (case-sensitive) in the table below will be needed in order to complete the lab. The task sections below provide details on the use of this information.
Table 1. Device credentials for lab workspace.

<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@dmin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router</td>
<td>192.168.1.1</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.0.113.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Shell variables and their corresponding absolute paths.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/Zeek/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/Zeek/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/Zeek-Labs-Workspace/</td>
</tr>
</tbody>
</table>

Lab roadmap

This lab is organized as follows:

1. Section 1: Introduction to Weka.
2. Section 2: Building a decision classifier with Weka.
3. Section 3: Reviewing the classifier’s predictions on a test dataset.

1 Introduction to Weka

After formatting Zeek output logs into the ARFF files, Weka is now able to process them. Weka contains the algorithms necessary to develop a number of machine learning classifiers. More information on the Weka software can be found on their documentation pages. Because the lab environment is not connected to the Internet, viewing the documentation must be done on a personal computer not located within the lab topology.
In the following sections, we train a *DecisionTable* and a *J48 Decision Tree* classifier.

### 1.1 Starting Weka

Weka is already installed in the lab environment. This section shows how to launch Weka.

**Step 1.** On the top of the lab workspace, click on the *Bro2* button as shown below to enter the *Bro2* machine.

**Step 2.** On the left side of the *Bro2* desktop, click on the Terminal icon as shown below.

**Step 3.** Navigate to the Weka workspace directory. To type capital letters, it is recommended to hold the *Shift* key while typing rather than using the *Caps* key.

```bash
cd $ZEEK_LABS/weka
```

**Step 4.** Using Java, launch the Weka software.

```bash
java -jar weka.jar
```
Step 5. Once Weka has been loaded, a notification containing Weka related information will be displayed. Select the OK button to continue to the Weka GUI Chooser panel.

The Weka GUI Chooser panel will look similar to the following image.

Step 6. For this lab, we will be using the Explorer application. Click the Explorer button to launch the application.
Weka has been successfully launched and we can proceed to the next section.

2 Importing a dataset into Weka

Once the Explorer application opens, a new GUI window will be displayed. Initially, this window has all options greyed out, indicating that we have not yet opened or loaded a dataset.

The Explorer panel contains a Menu Bar located at the top of the GUI window. There is a total of 6 additional panels, which contain related information necessary to train, test and visualize classifiers developed while using Weka. By default, the Preprocess panel will be selected.

The Preprocess panel is used to import a training dataset to be used for training a machine learning classifier. Features can be removed, randomized or appended within this panel.
2.1 Loading the training dataset

Step 1. On the top left of the Preprocess window the Open file button can be found. Click the Open file button to load the training dataset.

Step 2. Enter the path to the trainset.arff file. Alternatively, use the GUI to navigate to the lab workspace directory to select the file. Use the Open button to load the trainset.arff file into Weka.

/home/vlab/Zeek-Labs-Workspace/Sample-Captures/trainset.arff
After clicking the Open button, the Preprocess panel will be updated to contain the trainset.arff file statistics.

The updated Preprocess panel will look similar to the following image.

Each section header has been highlighted with a red box. We can see that the Current relation, Attributes and Selected attribute sections have been updated to contain trainset.arff file data.
Step 3. Within the Attributes section, click the class feature to change the active attribute.

By selecting the class feature within the Attributes section, the Explorer panel will be updated to display the active feature.

Within the Current relation section, our dataset’s name, networktraffic, is displayed. Additionally, it is shown that the dataset contains 1401 unique data objects (instances).

Within the Selected attribute section, the class labels added to the dataset in the previous lab are counted. The trainset.arff dataset contains 831 data objects labeled with a 1, belonging to the malicious class, while 570 data objects are labeled with a 0, belonging to the benign class.

At this point, trainset.arff dataset has been successfully loaded into Weka and we can begin filtering the data before training a machine learning classifier.

2.2 Filtering the training dataset

The majority of machine learning classifiers are unable to handle string attributes. For network analysts, source and destination IP addresses are valuable features that are often necessary for traffic analysis. However, these IP addresses are unable to be stored as string values when training a classifier.
There is a number of ways to address this issue. If a network analyst were to know all of the unique IP addresses, when generating their ARFF dataset, they can create the nominal values similar to how we created the nominal protocol values.

Because Internet-scale traffic contains a very large number of unique IP addresses, the aforementioned process may not be feasible. Therefore, in the previous lab, we converted our source and destination IP addresses into numerical values. In this section, we will be using all unique iterations of the numerical values to generate a nominal list. By reformatting the IP addresses into numeric values using Terminal utilities, the Weka software will be able to select all unique IP addresses and convert them into a nominal feature set.

**Step 1.** Within the Preprocess tab, under the Filter section, click the Choose button.

**Step 2.** Under the unsupervised option, select the attribute option to display a list of attribute-based filters.
Step 3. Scroll to the *NumericToNominal* filter and double click to select it.

![Image of filter selection]

Step 4. Within the *Attributes* section, select the *sourceip* and *destip* attributes. Within the *Filter* section, click the *first-last* description to modify the filter.

![Image of attribute selection]

Step 5. Update the *Indexes of the Attributes* to be filtered. Click the *Apply* button to edit the indexes.
Step 6. On the right side of the Filter section, click the Apply button to apply the modified filter.

The source and destination IP addresses will now be converted to the Nominal feature type.

Step 7. Within the Attributes section, click the sourceip feature to change the active attribute.

By selecting the sourceip feature within the Attributes section, the Explorer panel will be updated to display the active feature.

Within the Selected attribute section, the sourceip feature will now display the Nominal data objects. In the following image, the highlighted sourceip related to the Bro2 machine’s IP address, with 831 unique instances being recorded.
Additionally, the Selected attribute section will be updated to show new statistics for each feature. The updated Selected attribute section is displayed in the previous image.

2.3 Training a decision table classifier

Step 1. Within the Explorer panel, click the Classify tab located at the top of the Explorer panel to switch to the Classify panel.

Step 2. Once the Classify panel has loaded, click the Choose button within the Classifier section to select which machine learning classifier we are developing.

Step 3. Under the rules collection, double-click with your mouse to select the DecisionTable classifier.
Step 4. Under the Test options section, click the Start button to begin training the classifier. Notice the Classifier section has been updated to display the DecisionTable classifier.

Step 4. See the Decision Table classifier’s results.
Within the Result list section we can see our new Decision Table that has been trained with the transet.arff dataset. Within the Classifier output section, we can see the prediction results for the Decision Table classifier. The Confusion Matrix depicts that the classifier had a 100% accuracy when predicting labels after being trained.

### 2.4 Training a decision tree classifier

**Step 1.** Click the Choose button within the Classifier section to select which machine learning classifier we are developing.

**Step 2.** Under the trees collection, double-click with your mouse to select the J48 decision tree classifier.
Step 3. Under the Test options section, click the Start button to begin training the classifier. Notice the Classifier section has been updated to display the J48 classifier.

Step 4. See the J48 Decision Tree classifier’s results.
Within the Result list section we can see our new J48 Decision Tree that has been trained with the transet.arff dataset. Within the Classifier output section, we can view the prediction results for the Decision Tree classifier. The Confusion Matrix depicts that the classifier did not have a 100% accuracy when predicting labels after being trained and misclassified a single malicious data packet as benign.

2.4.1 Updating the decision tree classifier

Because our J48 Decision Tree has made an error in predicted a label, we can attempt to remove or add additional features to increase the classifier’s accuracy.

Step 1. Right click the J48 Decision Tree under the Result list section to display more options. Click to Visualize the J48 Decision Tree.
Step 2. View the Visualized J48 Decision Tree. After viewing the Tree nodes, close the window before proceeding to the next step.

We can see the time feature column was the only decision node within the tree. For the purposes of this lab, the datasets were collected at varying times; therefore, the decision tree had an over reliance on the time feature to determine when the malicious and benign events took place.

Step 3. Within the Explorer panel, click the Preprocess tab located at the top of the Explorer panel to switch to the Preprocess panel.
Step 4. Once the Preprocess tab has loaded, under the Attributes section, uncheck the sourceip and destip attributes. Select and check the time attribute, followed by selecting the Remove button.

With the time feature removed, we can retrain our decision tree to view the new accuracy.

Step 5. Within the Explorer panel, click the Classify tab located at the top of the Explorer panel to switch to the Classify panel.

Step 6. The J48 Decision Tree should still be selected. Under the Test options section, click the Start button to begin training the classifier. Notice the Classifier section has been updated to display the new J48 classifier.
Step 7. See the J48 Decision Tree classifier’s results.

Within the Result list section we can see our new J48 Decision Tree that has been trained with the transet.arff dataset. Within the Classifier output section, we can view the prediction results for the Decision Tree classifier. The Confusion Matrix depicts that the classifier actually had a worse accuracy than the previously trained J48 Decision Tree.

In this example, we highlight the importance of choosing the best fit features when training a classifier. In a real-time network environment, it may take multiple tests before discovering which features are necessary for classifying a specific anomaly.
Step 6. Right click the newest J48 Decision Tree under the Result list section to display more options. Click to Visualize the J48 Decision Tree.

![Image of J48 Decision Tree visualization options]

Step 7. View the Visualized J48 Decision Tree.

![Visualized J48 Decision Tree]

Because the Decision Tree has a larger number of nodes, we are unable to see some of the decision thresholds. The following steps will explain how to scale the Visualized tree.

Step 8. Right click on the Visualized J48 Decision Tree and select the Auto Scale option.
Step 8. View the Visualized J48 Decision Tree.

Here we can see the new J48 Decision Tree has multiple layers and decision nodes. The duration feature has replaced the time feature as the root node, and the sourceip feature is used to further classify the dataset. However, because this tree has reduced accuracy, we will be continuing the lab by using the Decision Table created initially.

3 Reviewing the classifier’s predictions on a test dataset

Now that we have determined that the Decision Table was a more accurate classifier, we can begin testing the classifier’s accuracy using the test dataset.

3.1 Saving the decision table

It is possible to save a trained classifier to be reused in future instances of testing and classification. This section will introduce how to save a trained classifier.

Step 1. Under the Result list section, right click on the Decision Table and select Save model.
Step 2. Navigate to the Lab workspace directory and save the *Decision Table*.

Once saved, we can proceed to testing the classifier’s accuracy on predicting labels for the test dataset.

### 3.2 Using the classifier to predict labels for the test dataset

**Step 1.** Under the *Test options* section, select the *Set* button to load the test dataset. Within the *Test Instances* window, click the *Open file* button.
Step 2. Select the testset.arff file and click the Open button to load the test dataset.

Step 3. Click the Close button to close the Test Instances window. Under the Test options section, select the More options button to configure the classifier to match the following image. Uncheck the Output model, Output per-class stats, Output confusion matrix, and Store predictions for visualization options. Click the Choose button for the Output predictions and select PlainText. Once finished, click the OK button.
Step 4. Under the Result list section, right click on the Decision Table and select Re-evaluate model on current test set.

Step 5. After filtering the sourceip and destip features into Nominal attributes, the testset.arff file will not be properly formatted. Weka will need to update the testset.arff dataset to be used by the classifier. Select the Yes button on the ClassifierPanel pop-up panel.
The classifier will generate new predictions, which can be viewed by saving the resulting .arff file.

### 3.3 Viewing the predicted labels for the test dataset

To save the resulting .arff file,

**Step 1.** Within the *Explorer* panel, click the *Visualize* tab located at the top of the *Explorer* panel to switch to the *Visualize* panel.

Displayed will be resulting graphs from attribute correlations solved by the machine learning classifier.
Step 2. Select the *duration x duration* graph, found in the sixth column (*duration*) and second row (*duration*).

Step 3. Click the *Save* button.

Step 3. Click the *Up One Level* button highlighted by the red box. This will navigate to the Lab workspace directory. Input the following name to save the file as *DecisionTableResults* in *.arff* format. To type capital letters, it is recommended to hold the *Shift* key while typing rather than using the *Caps* key.
Step 4. Close all of the Weka tabs with the orange x on the top right corner of each panel.

Step 5. Return to the Terminal and navigate to the lab workspace directory.

```
cd $ZEEK_LABS
```

Step 5. Using a text editor, view the `DecisionTableResults.arff` file.

```
gedit DecisionTableResults.arff
```

The file will be opened, and each data object will contain a new classification label.
Traffic generated by the Bro1 machine with the source IP address of 192.168.1.3 was labeled with a 1, as malicious traffic. Traffic generated by a machine found in the smallFlows.pcap file was labeled with a 0, as benign traffic.

Concluding this lab, we have introduced the capabilities of implementing a machine learning classifier to detect specific anomalies or events. Multiple classifiers can be used for training network security classifiers, and the features within each training dataset can have a profound impact on the classifier’s accuracy. By removing, modifying or adding new features you can test the accuracy of a classifier. In this lab, we generated a Decision Table that was capable of labeling malicious and benign traffic.

References

ZEEK INTRUSION DETECTION SERIES

Lab 10: Profiling and Performance Metrics of Zeek

Document Version: 02-01-2020

Award 1829698
“CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers”
# Contents

Overview ......................................................................................................................... 3  
Objectives ....................................................................................................................... 3  
Lab topology .................................................................................................................. 3  
Lab settings .................................................................................................................... 3  
Lab roadmap .................................................................................................................... 4  
1 Introduction to Zeek profiling .................................................................................... 5  
2 Generating customized malicious network traffic ....................................................... 5  
  2.1 Starting a new instance of Zeek ............................................................................. 6  
  2.2 Setting up the Bro2 machine for live network capture ....................................... 6  
  2.3 Using the Bro1 machine to launch customized TCP-based scans ....................... 7  
3 Generating and viewing Zeek profiling log files ....................................................... 9  
  3.1 Applying the profiling filter .................................................................................. 9  
4 Implementing tools to test Zeek’s performance ......................................................... 12  
  4.1 Using sysstat sar utility ....................................................................................... 12  
  4.2 Using the top utility ............................................................................................. 13  
  4.3 Viewing the resource consumption of Zeek ....................................................... 14  
  4.4 Closing the current instance of Zeek ................................................................. 15  
References ...................................................................................................................... 16
Overview

With Zeek’s event-based framework, anomalies can be detected, processed and analyzed with external software. In this lab, we explain Zeek’s profiling log stream and Zeek’s resource consumption.

Objectives

By the end of this lab, students should be able to:

1. Enable Zeek’s profiling log stream for session-based statistics.
2. Generate customized traffic to be captured by Zeek’s profiling.
3. Implement tools necessary for testing Zeek’s resource consumption.

Lab topology

Figure 1 shows the lab workspace topology. This lab primarily uses the Bro2 machine for Zeek log file development and the Bro1 machine to generate customized traffic.

![Lab topology](image)

Figure 1. Lab topology.

Lab settings

The information (case-sensitive) in the table below provides the credentials to access the machines used in this lab.
Table 1. Device credentials for lab workspace.

<table>
<thead>
<tr>
<th>Virtual Machine</th>
<th>IP Address</th>
<th>Account</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeek1</td>
<td>192.168.1.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td>DTN</td>
<td>192.168.1.3</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td>Client</td>
<td>192.168.3.2</td>
<td>root</td>
<td>@dmin123</td>
</tr>
<tr>
<td>Zeek2</td>
<td>192.168.2.2</td>
<td>admin</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Border router</td>
<td>192.168.1.1</td>
<td>root</td>
<td>password</td>
</tr>
<tr>
<td></td>
<td>192.168.2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.0.113.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Shell variables and their corresponding absolute paths.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ZEEK_INSTALL</td>
<td>/usr/local/zeek</td>
</tr>
<tr>
<td>$ZEEK_TESTING_TRACES</td>
<td>/home/vlab/Zeek/testing/btest/Traces/</td>
</tr>
<tr>
<td>$ZEEK_PROTOCOLS_SCRIPT</td>
<td>/home/vlab/Zeek/scripts/policy/protocols/</td>
</tr>
<tr>
<td>$ZEEK_LABS</td>
<td>/home/vlab/Zeek-Labs-Workspace/</td>
</tr>
</tbody>
</table>

Lab roadmap

This lab is organized as follows:

1. Section 1: Introduction to Zeek profiling.
2. Section 2: Generating customized malicious network traffic.
3. Section 3: Generating and viewing Zeek profiling log files.
4. Section 4: Implementing tools to test Zeek’s performance.
1  Introduction to Zeek profiling

Zeek includes the option of enabling profiling. When profiling is enabled, a new log stream will be created to store session-related statistics. The Profile log file will contain a large variety of information, including but not limited to running time, memory usage, connection information and packet protocol statistics.

To enable profiling while using Zeek for offline packet capture file processing, you will need to implement the following functionality in a Zeek script.

```
1 module Profiling;
2 redef profiling_file = open_log_file("Statistics");
3 redef profiling_interval = 3 secs;
4 redef expensive_profiling_multiple = 5;
5 event zeek_init() {
6    set_buf(profiling_file, F);
7 }
```

The script is explained as follows. Each number represents the respective line number:

1. Sets the module workspace as Profiling.
2. Specifies the name of the new profiling log file. In this example we have named the file Statistics.log.
3. Specifies the time interval for Zeek to record empirical information. In this example the time interval is 3 seconds.
4. Specifies the number of profiling intervals defined in Line 5. In this example, the profiling interval is 5 instances.
5. Initialization event.
6. Appends the new log stream information.
7. End of initialization event.

Profiling is enabled by calling the Zeek script during packet processing, as reviewed in the previous labs.

```
zeek -r <packet capture file> <Profiling Script>
```

- `<packet capture file>`: denotes the input packet capture file.
- `<Profiling Script>`: denotes the Zeek script to be run during packet processing.

In the following section we generate customized malicious traffic to be viewed within a Zeek profiling log.

2  Generating customized malicious network traffic
This section introduces creating and using a Zeek profiling script, which will enable session-based statistics for Zeek packet capture file processing.

### 2.1 Starting a new instance of Zeek

**Step 1.** On the top of the lab workspace, click on the Bro2 button as shown below to enter the Bro2 machine.

**Step 2.** On the left side of the Bro2 desktop, click on the Terminal icon as shown below.

**Step 3.** Start Zeek by typing the following command on the terminal. This command navigates to Zeek’s default installation directory and invokes `zeekctl` tool to start a new instance. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key. When prompted for a password, type `password` and hit `Enter`.

```bash
cd $ZEEK_INSTALL/bin && sudo ./zeekctl start
```

A new instance of Zeek will now be active, and we can proceed to the next section of the lab.

### 2.2 Setting up the Bro2 machine for live network capture
Step 1. Navigate to the lab workspace directory. To type capital letters, it is recommended to hold the Shift key while typing rather than using the Caps key.

```
cd $ZEEK_LABS
```

Step 2. Start live packet capture on interface ens33 and save the output to a file named ntraffic.pcap. Take notice that the 0 in the following command is the number zero. If necessary, type password as the password.

```
sudo tcpdump -i ens33 -s 0 -w ntraffic.pcap
```

The Bro2 machine is now ready to begin collecting live network traffic. Next, we use the Bro1 machine to generate unique scan-based network traffic.

2.3 Using the Bro1 machine to launch customized TCP-based scans

In this section we use the nmap software to generate TCP-based scan traffic.

This section introduces two new options for the nmap software.

- `-f`: specifies to send packet fragments. By fragmenting packets, a scanner can attempt to bypass firewalls that check for entire packet signatures.
- `-mtu <num>`: specifies the max number of bytes to be sent in a fragmented packet. The number variable must be a multiple of 8.

Step 1. On the top of the lab workspace, click on the Bro1 button as shown below to enter the Bro1 machine.

Step 2. On the left side of the Bro1 desktop, click on the Terminal icon as shown below.
Step 3. Launch a fragmented TCP SYN scan against the Bro2 machine. If necessary, type `password` as the password.

```bash
sudo nmap -sS -f 192.168.2.2
```

Step 4. Launch a fragmented TCP SYN scan with a packet size of 8 bytes against the Bro2 machine.

```bash
sudo nmap -sS -mtu 8 192.168.2.2
```

Step 5. Launch a fragmented TCP SYN scan with a packet size of 64 bytes against the Bro2 machine.

```bash
sudo nmap -sS -mtu 64 192.168.2.2
```
Step 6. Return to the Bro2 machine and use the `Ctrl+c` key combination to stop the live traffic capture session.

The capture session’s results show that 10050 packets were captured. We will now process the `ntrafic.pcap` file with profiling enabled.

3 Generating and viewing Zeek profiling log files

Now that we have collected fragmented traffic, we can begin processing the packet capture file with Zeek.

3.1 Applying the profiling filter

Step 1. View the `EnableProfiling.zeek` Zeek script.

```bash
nl EnableProfiling.zeek
```
Similar to the example in the introduction, the *EnableProfiling.zeek* Zeek script is used to create a new log file named *Statistics.log* containing Zeek profiling statistics. The script enables the intervals to be 5 seconds apart, with 15 interval counts.

**Step 2.** Process the *ntraffic.pcap* packet capture file.

```
zeek -C -r ntraffic.pcap EnableProfiling.zeek
```

**Step 3.** Display the contents of the *Statistics.log* file.

```
sudo gedit Statistics.log
```

The *Statistics.log* file will be displayed.
Viewing the *Statistics.log* file, each profiling_interval will be displayed between a line separator made by dashes `---`.

Within the *Statistics.log* file, we can see the total memory used while processing the packet capture file, the Run-time, as well as a number of TCP flags, connections and Triggers. Within the first iteration of profiling_interval we see that no TCP packet flags have been recorded.

**Step 4.** Go to the next iteration of profiling_interval within the *Statistics.log* file.
By scrolling through the Statistics.log file, we can see information found in the next iteration of a profiling_interval. We can see the total number of TCP-States:Syn has updated the Rst flag count to include 1002 packets due to our previous scans. Additional Triggers have been included, and the total memory usage has been updated.

Zeek profiling is a great tool for generating more detailed session-based statistics while processing packet capture files with Zeek.

4 Implementing tools to test Zeek’s performance

While Zeek profiling will display the resulting statistics after processing a packet capture file, it is important to monitor Zeek resource consumption during network traffic analysis.

A number of Linux-based software utilities can be used to track system resource consumption in real time.

4.1 Using sysstat sar utility

The sar command can be used to display a number of system resources over specific time intervals. The following steps will highlight the ways to enable sar resource tracking.

Step 1. Enable sar utility to track CPU consumption.

```
sar 2 30
```
• **`sar`**: calls the `sar` utility, belonging to the `syystat` packages.
• **2**: indicates each iteration of CPU statistics is separated by a 2 second time interval.
• **30**: indicates that a total of 30 iterations of CPU statistics should be displayed.

### Step 2.
Enable **`sar`** utility to track memory consumption.

```
sar -r 3 25
```

• **`sar`**: calls the `sar` utility, belonging to the `syystat` packages.
• **-r**: indicates memory consumption in kilobytes.
• **3**: indicates each iteration of memory statistics is separated by a 3 second time interval.
• **25**: indicates that a total of 25 iterations of memory statistics should be displayed.

### 4.2 Using the top utility

Alternative to the syystat sar utility, the top utility can be used to display the resource consumption of every active process.
Step 1. Enable the top utility to track resource consumption.

```
top -i
```

- **top**: calls the **top** utility.
- **-i**: toggles idle processes off, so that only active processes will be displayed.

After entering the command, the Terminal will display the resource consumption.

Each row will belong to a unique process and display the related CPU and memory resource usage.

4.3 Viewing the resource consumption of Zeek

Step 1. Using the File drop down options, create a New Tab within the Terminal.

Step 2. Navigate to the lab workspace directory and enter the Sample-Captures folder.

```
  cd $ZEEK_LABS/Sample-Captures
```
Step 3. In the second tab, begin packet capture file processing of the `bigFlows.pcap` file using Zeek.

```
zeek -r bigFlows.pcap
```

Step 4. Return to the first Terminal tab and view the active processes.

```
29277 admin 20 0 953704 177400 18312 R 99.0 7.0 0:09.18 zeek
```

4.4 Closing the current instance of Zeek

After you have finished the lab, it is necessary to terminate the currently active instance of Zeek. Shutting down a computer while an active instance persists will cause Zeek to shut down improperly and may cause errors in future instances.

Step 1. Stop Zeek by entering the following command on the terminal. If required, type `password` as password. If the Terminal session has not been terminated or closed, you may not be prompted to enter a password. To type capital letters, it is recommended to hold the `Shift` key while typing rather than using the `Caps` key.

```
cd $ZEEK_INSTALL/bin && sudo ./zeekctl stop
```
Concluding this lab, we introduced Zeek’s profiling capabilities and generated fragmented traffic to be processed into a profiling log file. Lastly, we introduced Terminal utilities that can be used to track Zeek’s resource consumption per process.

References