ISA 419: Data Driven Security

20 - Web Server Logs

Fadel M. Megahed

Endres Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Automated Scheduler for Office Hours

Fall 2022

Outline

Learning Objectives for Today's Class

Objectives

- **Understand and explain the basic format of a server log**
- **Examine a web Server Log Dataset**
- **Identify one type of attack based on the dataset**

Outline

Web Servers

Let us consider the following URL https://miamioh.edu/fsb/academics/isa/index.html. My web browser will translate this URL into a connect to www.miamioh.edu with the following HTTP request:

```
GET fsb/academics/isa/index.html HTTP
Host: www.miamioh.edu
```

The web server ib www.miamioh.edu will append the given path to the path of its root directory, giving the following result (assuming an Apache Server):

```
home/www/fsb/academics/isa/index.html
```

The web server then reads the file, if it exists, and sends a response to the client's web browser. The response will describe the content of the file and contain the file itself or an error message will return saying that the file does not exist or is unavailable.

Web Server Logs: CLF [1]

Per the Appache Manual, log entries written in a format known as the Common Log Format (CLF) will look something like this

```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0"
200 2326
```

Each part of this log entry is described below:

- **127.0.0.1:** This is the IP address of the client (remote host) which made the request to the server.
- **-:** The "hyphen" in the output indicates that the requested piece of information is not available.

Web Server Logs: CLF [2]

- **frank:** This is the userid of the person requesting the document as determined by HTTP authentication. If the status code for the request (see below) is 401, then this value should not be trusted because the user is not yet authenticated.
- **[10/Oct/2000:13:55:36 -0700]:** The timestamp the request was received, with -0700 indicating the time zone
- **"GET /apache_pb.gif HTTP/1.0":** The request line from the client is given in double quotes:
  - First, the method used by the client is GET.
  - Second, the client requested the resource /apache_pb.gif, and
  - Third, the client used the protocol HTTP/1.0

Web Server Logs: CLF [3]

- **200:** This is the status code that the server sends back to the client. This information is very valuable, because it reveals whether the request resulted in a successful response (codes beginning in 2), a redirection (codes beginning in 3), an error caused by the client (codes beginning in 4), or an error in the server (codes beginning in 5).
- **2326:** The size of the object returend to the client, not including the response headers.

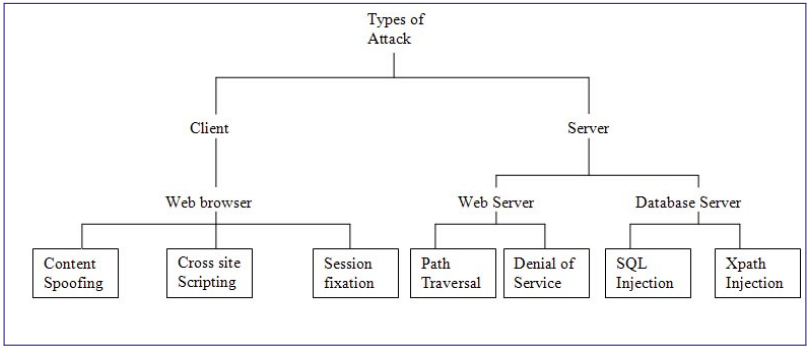
Common Web Networking Protocols

Protocol	Common Port
FTP (File Transfer Protocol)	20, 21
SSH (Secure Shell)	22
Telnet	23
SMTP (Simple Mail Transfer Protocol)	25
DNS (Domain Name Service)	53
TFTP (Trivial File Transfer Protocol)	69
HTTP (Hypertext Transfer Protocol)	80
POP3 (Post Office Protocol version 3)	110
NNTP (Network News Transport Protocol)	119
NTP (Network Time Protocol)	123
IMAP4 (Internet Message Access Protocol version 4)	143
HTTPS (Hypertext Transfer Protocol Secure)	443

Types of Attacks Targeting Web Systems [1]

The Survey on Attacks targeting Web based System through Application Layer provides a nice overview of different types of attacks affecting web based systems. The following two slides are utilizing information from that survey; you should read the page and do your examination of some of these concepts.

Types of Attacks Targeting Web Systems [2]



Types of Attacks Targeting Web Systems [3]

Attacks	Target	Effects	Severity	Prevention
Content Spoofing	Client	Loss of confidential information like username and password	moderate	Educating user
Cross site Scripting	Client	Unauthorized access to client's private information like cookies	moderate	Disabling scripts at client's browser
Session fixation	Client	Unauthorized access to user account information	moderate	Issuing session id after authentication
Path traversal	Server	Unauthorized access to files and directories stored on server.	moderate	checking "../" and its variants.
Denial of Service	Server	Web application unavailable for normal users	moderate	Fixing bugs, reverse turing test, resource replication
SQL injection	Server	Unauthorized operations and access to database.	Severe	checking user input, giving minimum privileges
Xpath injection	Server	Unauthorized access to all data entities.	Severe	checking user input

Outline

Class Activity

Consider the dataset in Stanford Server Logs. We know that one type of attack occurred around 8 am. By building on the code below, please address the following:

- Please identify the exact date and time.
- What are the characteristics of that attack? (i.e. what anomalies are shown in the feature(s))
- What is the type of attack deployed by the attackers?

```
pacman::p_load(tidyverse, magrittr, lubridate, hms, iptools)
df = read.table("http://web.stanford.edu/class/cs259d/hw/server-log.txt",
                sep=" ", stringsAsFactors = F) %>% .[,-1]
colnames(df) = c('StartDate', 'StartTime', 'Duration', 'Server',
                 'SourcePort', 'DestinationPort', 'SourceIP',
                 'DestinationIP')
```

Outline

Outline

Project Description and Expectations

Let us discuss this in more detail (see Canvas Project Description for more details).

Outline

What you should have learned from today's class?

Objectives

- **Understand and explain the basic format of a server log**
- **Examine a web Server Log Dataset**
- **Identify one type of attack based on the dataset**

ISA 419: Data Driven Security

20 - Web Server Logs

Fadel M. Megahed

Endres Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Automated Scheduler for Office Hours

Fall 2022