Integrating Splunk Into Some of Cybersecurity Courses

Fatima Abd Elmajid

Follow this and additional works at: https://lib.dr.iastate.edu/creativecomponents

Part of the Engineering Commons

Recommended Citation

This Creative Component is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Creative Components by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Integrating Splunk Into Some of Cybersecurity Courses

by

Fatima Abd Elmajid

A creative component submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
Master of Science

Major: Computer Engineering and Cybersecurity

Program of Study Committee

Doug Jacobson, Major Professor

Iowa State University
Ames, Iowa
2021
TABLE OF CONTENTS

LIST OF FIGURES .................................................................................................................ii

ACKNOWLEDGMENTS ........................................................................................................ iii

ABSTRACT ......................................................................................................................... iv

CHAPTER 1. Introduction........................................................................................................ 1

CHAPTER 2. Main Functions and Components....................................................................... 3

Splunk Components.............................................................................................................4

Forwarders..........................................................................................................................5

CHAPTER 3. Main Types of Splunk Deployments...................................................................6

CHAPTER 4. Implementation and Results................................................................................ 8

CHAPTER 5. What Students Can Learn..................................................................................11

CHAPTER 6. Conclusion and Future Recommendations .......................................................12

References............................................................................................................................13
LIST OF FIGURES AND TABLES

Figure 1-1 Splunk Captures Different Data Structures .................................................. 1
Figure 2-1 Universal Forwarder ..................................................................................... 5
Figure 4-1 Configuration Steps .................................................................................... 8
Figure 4-2 Network Implementation ............................................................................ 9
Figure 4-3 Results ......................................................................................................... 9
Figure 4-4 Data Summary ............................................................................................ 10
Table 5-1 What Students Can Learn ................................................................. 11
ACKNOWLEDGMENTS

I would like to thank my major professor, Doug Jacobson, my family, and my classmate Shanell Hurst for their assistance and support throughout the course of this project. I would like to thank my advisers and my professors at Iowa State University.
ABSTRACT

Splunk is a powerful software that indexes, manages, and enables people to search data from any application, server, or network device in real time. The information assurance department at Iowa State University offers several cybersecurity courses to both undergraduate and graduate students. Those courses often require some type of remote access to school servers to complete course labs.

Integrating Splunk into Cybersecurity courses would help to manage and monitor machine data from multiple servers and improve both students and faculty experience across all systems used. Moreover, it would help students learn how to detect security risks and network abuse, and to use Splunk Enterprise for statistical purposes.

This project is the second step towards integrating Splunk to cybersecurity courses. A successful implementation of a virtual network was used to test if a keyhole (proxy server) can be used as a relay forwarder between a Splunk Enterprise who has a public IP address and the forwarders who have private IP addresses. At the end of this project, it has been proved that Splunk can be integrated into the courses to improve the students and faculty experience. Recommendations for future improvements have been listed.
CHAPTER 1. Introduction

According to Splunk 90% of data accumulated by today’s organizations is machine data. Splunk is a powerful software that can take any data and put it to an intelligent searchable index. It is used by software developers and other security professionals. It can parse and index very large quantities of structured and unstructured machine data. If we are talking about security, the Splunk platform can help to reduce incidents, improve observability, and reveal deeper insights. For example, machine data in the form of audit logs can be used for identifying malicious login attempts. Moreover, Distributed Denial of Service (DDOS) attacks can be identified by looking for patterns in the client IP field from the audit logs [1].

Splunk can capture essential security-related log data from different devices and software. This can enhance students’ understanding of concepts such as vulnerability scanners, phishing prevention, firewalls, and intrusion detection and prevention. Figure 1-1 shows how Splunk can index different data types [1].

![Fig 1-1 Splunk Captures Different Data Structures. Retrieved from [2]]
With that in mind and a powerful query language such as Splunk Search Processing Language (SPL), students can run effective ad hoc queries and correlate events across many security sources. Moreover, students can collect and index audit logs to create useful reports to comply with, and audit against specific security standards, which can be industry-related, such as PCI compliance of financial transactions; customer-related, such as the National Institute of Standards and Technologies (NIST) requirements in working with the US government; or data privacy-related, such as the Health Insurance Portability and Accountability Act (HIPAA) or the European Union's General Data Protection Regulation (GDPR) [1].

In this project, it was proved that Splunk can be integrated into cybersecurity courses by implementing a small enterprise deployment in a VMWare environment.
CHAPTER 2. Splunk Functions and Components

2.1 Splunk Main Functions

Splunk has five main functions. This includes collect and index, search, knowledge object, add knowledge, and monitor and alert.

2.1.1 Collect and Index

Splunk offers several ways to receive data from machine data sources. It collects data for any source so that events can be put inside the indexer to be searched.

The most common way is using Splunk Forwarders (see 2.2.3), a Splunk forwarder is a piece of software that consumes data and sends it to an indexer. Forwarders require minimal resources and have little impact on performance, so they can usually reside on the machines where the data originates. The collected data is parsed and indexed in indexes, which are directories that consist of raw data and time-series index files. For Splunk Enterprise, the data can be on the same machine as an indexer (local data) or on another machine (remote data).

2.1.2 Search

In the search function, Splunk provides Search Processing Language (SPL) to query the indexed data. SPL provides a family of commands known as transforming commands which can be used to generate reports and visualizations just with one or two commands. It gives you the ability to search data that has a specific value across multiple source types.

2.1.3 Knowledge object

In the Add Knowledge function, Splunk offers a variety of tools such as lookups, field extractions, tags, event types, workflow actions, and data models. These tools help to make sense of the machine data in the quickest and most effective way for other users. You can add it to your data to affect how your data is interpreted.
2.1.4 Monitor & Alert

In the Monitor and Alert function, Splunk enables you to monitor your systems in real time to identify issues and attacks and trigger alerts based on the thresholds you set.

2.1.5 Report and Visualize

In the Report and Visualize function, Splunk offers the capability to produce stunning reports and dashboards and even schedule them for automatic delivery. This allows you to collect reports and visualizations on data into dashboards [1][3][4].

2.2 Splunk Components

Splunk has 3 basic components which are: indexers, forwarders, and search heads.

2.2.1 Indexer

The indexer’s mainly function is to process machine data. It stores the results in indexes as events, enabling fast search and analysis. As it indexes data, it creates several files organized in sets of directories by age.

2.2.2 Search heads

Its main function to transform the data without changing the underlying index data. This allows users to use the Search language to search the indexed data. In addition, it distributes user search requests to the indexers. There are knowledge objects on the search heads that can be created to extract additional fields.

2.2.3 Forwarders

Forwarders are Splunk Enterprise instances that consume and send data to the index. They require minimal resources and have little impact on performance. Typically, they reside on the machines where the data originates. This is the primary way data is supplied to indexers.
2.2.3.1 Universal Forwarder

Contains only the components that are necessary to forward data. It cannot search, index, or produce alerts with data, and it does not parse data. The universal forwarder cannot be used to route data to different indexers based on its contents.

Fig 2-1 Universal Forwarder

2.2.3.2 Intermediate Forwarder

Intermediate forwarders can be used when you want to reduce or limit network bandwidth usage on specific network segments. In addition, you can use it if you have some need to limit access to the indexer machine, for instance, for security reasons. It can be used in other situations, for example, as in our case for this project.

2.2.3.4 Heavy Forwarder

A heavy forwarder is a full Splunk Enterprise instance that can index, search, change, and forward the data.

2.2.3.5 Light Forwarder

Light forwarder is also a full Splunk Enterprise instance, with more features disabled to achieve as small resource footprint as possible [1][3][4].
CHAPTER 3. Main Types of Splunk Deployment

3.1 Important Considerations Before Deployment

When deploying Splunk, there are several questions that need to be asked and answering those questions, would help to identify which type of deployment we need to use. These include:

✓ What data sources will be accessed and managed?
✓ How much data volume do these sources will generate in a typical day?
✓ What type of server hardware will be used to deploy Splunk?
✓ Who will have access to and to which data sources?
✓ How long will you want to retain your indexed data for search and for longer term storage? [1][3][4]

3.2 Main Types of Deployments

There are four basic types of deployments, which are, departmental, small enterprise, medium, and large enterprise [5].

3.2.1 Departmental

A departmental deployment is a single instance that combines indexing and search functions. It would be a good option for organizations indexing < 20GB/day within less than 20 users. In this type, forwarders would need to be installed at data source and forwarders are responsible to collect data and send it to Splunk servers.

3.2.2 Small Enterprise

In small enterprise a small search head cluster is used with several indexers. This would be a good option for organizations indexing less than 100 GB/day and with less than 100 users. The small enterprise may have several hundred forwarders. Here, search management and index functions need to be split across multiple machines.

3.2.3 Medium Enterprise
A medium enterprise allows adding a search head Cluster and typically it requires less than or equal to three search heads. This type of deployment allows users and searches to share resources and to coordinate activities to handle search requests. Here we need to use a deployer to manage and distribute applications to the members of the search head cluster.

3.2.4 Large Enterprise

In this deployment a large search head cluster with large numbers of indexers is used. It can be either traditional index clusters which prevent data loss and promote availability, or a non-replicating index clusters which offer simplified management and do not provide availability or data recovery [1][3][4][5].
CHAPTER 4. Implementation and Results

In my project implementation, I designed a network in a VMWare environment based on the departmental deployment. The network consists of Splunk Enterprise (SE) server, 3 forwarders, 1 relay forwarder, and a Pfsense firewall. All the boxes were Linux ubuntu boxes, Pfsense firewall was installed in a FreeBSD.

4.1 Configuration steps

First, I configured the universal forwarder to connect to a receiving indexer. I used the command, 
./splunk add forward-server IP:9997, from the forwarder.

Second, I configured a data input on the forwarder. I used the command, ./splunk add monitor /var/log, to determine what data to collect. Third, I configured the receiver and the intermediate forwarder [6][7].

4.2 Network Implementation

In the network implementation I configured all the servers, ensuring they are connected to the network that corresponds with given IP range. Additionally, I built and managed both the forwarders and the Splunk enterprise server to be connected through relay forwarder using Splunk CLI. I made sure that the forwarders can access the Splunk enterprise server both directly and through the relay forwarder. Figure 4-2 shows the network implementation.
4.3 Results

The results showed that data from the forwarders was received successfully to the Splunk server through the intermediate forwarder as shown in Figure 4-3 and Figure 4-4.
Figure 4-4 Data Summary

<table>
<thead>
<tr>
<th>Host</th>
<th>Count</th>
<th>Last Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forwarder1</td>
<td>4,799</td>
<td>11/10/20 12:31:27 PM</td>
</tr>
<tr>
<td>forwarder2-virtual-machine</td>
<td>1,408</td>
<td>10/28/20 10:16:17 AM</td>
</tr>
<tr>
<td>machine virtual machine</td>
<td>64,070</td>
<td>11/10/20 12:34:27 AM</td>
</tr>
<tr>
<td>ubuntu</td>
<td>6,012</td>
<td>10/13/20 11:00:33 PM</td>
</tr>
</tbody>
</table>
CHAPTER 5. What Students Can Learn

Students can practice use Splunk Processing Language (SPL) to help them understand the material better. For example, transforming command are used to transform the result of a search into such data structures which will be useful in representing the statistics and data visualizations. Transforming commands include chart, timechart, stats, top, rare, contingency, and highlight [4].

Below in Table 5-1, I listed a couple of scenarios that can be integrated in a Cybersecurity course assignment [4].

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Why</th>
<th>How using SPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Use the rare command</td>
<td>See what files get accessed the least amount in a web application</td>
<td>(index=main sourcetype=access_combined_wcookie status=200</td>
</tr>
<tr>
<td></td>
<td>Discover backdoor that few people know about</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Find files that are accessible online, but get very little traffic</td>
<td></td>
</tr>
<tr>
<td>2. Use the stats command</td>
<td>Run a report of the browsers that users are using to access the web application from</td>
<td>(index=main sourcetype=access_combined_wcookie</td>
</tr>
<tr>
<td></td>
<td>Use the stats list function to generate a list of all useragent values that have accessed the web application</td>
<td></td>
</tr>
<tr>
<td>3. Use the stats command</td>
<td>Get a report of users’ IP trying to access forbidden pages(status (403)) in the web application</td>
<td>(index=main sourcetype=access_combined_wcookie status=403)</td>
</tr>
<tr>
<td>4. Use keywords to search for failed authentication</td>
<td>Search the _audit index for events where the action of &quot;login attempt&quot; returned a &quot;failed&quot; info value for the username of admin over a period</td>
<td>(index=_audit action=&quot;login attempt&quot; info=failed user=admin)</td>
</tr>
</tbody>
</table>

Table 5-1 What Students Can Learn
CHAPTER 6. Conclusion and Future Recommendations

In this project, it has been assessed if the keyhole can be used as a relay forwarder while having 2 interfaces with Linux ubuntu installed. In addition, it has been tested if the keyhole can be used as a relay forwarder between a Splunk Enterprise server who has a public IP address and the forwarders who have private IP addresses.

For future deployments, I recommend using a medium enterprise deployment for up to 100 students as users, and to increase the number of indexers from one to four or nine indexers. I also recommend creating more search heads, for example, three search heads can do the work. This type of deployment would need to activate the deployment server within the Splunk Enterprise to have the best performance. Students may take Splunk Fundamentals 1 training first before practicing.
References


[2] Splunk Enterprise: 7.0.0 (wordpress.com)


[4] Splunk Fundamentals 1 Training, Splunk Fundamentals 1 | Splunk

