LOG FILE ANALYSIS USING SIGNATURE DETECTION (LoFA-SD)

ABIGAIL KOAY MAY YEE

A report submitted in partial fulfilment of the requirements for the award of the degree of Bachelor of Computer Science (Computer Systems & Networking)

Faculty of Computer Systems & Software Engineering Universiti Malaysia Pahang

MAY, 2011

PERI	PUSTAKAAN G
UNIVERSITI	MALAYSIA PAHANG
No. Perolehan 069140 Tarikh	No. Panggilan QA 76.76 . A25 K63

ABSTRACT

The increasing popularity of network technology has brought convenience to human life. People have started to rely on network technologies more and more in their daily life. It has slowly becoming a very important part in the human life. Network technologies have involved in communication, medical, financial, business, education and so on. Although it brings many benefits for human, it also attracts hackers and attackers to attack servers and computers. This has created the need for network security to secure the network from being exposed to attacks. In order to solve the exposure of threats to the networks, organizations are therefore facing the challenge to implement adequate security method to secure the network from being exploited. The method they seek ought to be effective, reliable and persistence. The experts have come out with lots of methods in securing the network. It needs to depend on the situation whether which method is suitable for use. In each network devices, it contains log files which are a record of events occurring within their network. Using the log files from network devices is one of the ways to detect and analyze intrusion. Therefore this project, LoFA-SD proposes its approach of detecting and analyzing intrusion. The approach is by using signature detection and log files from the network devices to run the process. The system will execute a pattern matching mechanism between the network pattern and the reference intrusion patterns from database. The system will also create statistical reports on the intrusion attacks in the network and among the network devices involved. From the process data retrieval, signature detection, pattern matching until report generation will help the security administrator to identify vulnerable attacks and potential attacks which happen more frequently in a range of time.

ABSTRAK

Teknologi rangkaian dalam informasi maklumat yang semakin popular telah membawa manfaat dan membawa kemudahan kepada kehidupan manusia. Manusia telah semakin bergantung kepada teknologi rangkaian dalam kehidupan mereka. Ia telah menjadi sebahagian yang penting dalam kehidupan harian mereka. Teknologi rangkaian ini diperlukan dalam sektor telekomunikasi, perubatan, kewangan, perdagangan, pendidikan dan lain-lain lagi. Walau ia membawa banyak manfaat kepada manusia, ia telah menarik perhatian ramai pengodam ataupun penyerang untuk menyerang komputer serta rangkaian berkenaan. Ini telah menyebabkan keperluan untuk mewujudkan kaedah keselamatan untuk melindungi sistem rangkaian daripada diserang. Untuk menyelesaikan masalah pendedahan rangkaian serta maklumat daripada serangan, organisasi-organisasi kini menghadapi cabaran dalam mewujudkan kaedah yang sewajarnya untuk melindungi rangkaian masingmasing daripada dieksploitasi. Kaedah yang dicari hendaklah berkesan dan tepat. Terdapat banyak kaedah yang boleh digunakan untuk melindungi rangaian daripada serangan. Ia juga bergantung kepada keadaan dan situasi berkenaan untuk menentukan kaedah yang paling sesuai digunakan. Setiap mesin dan peralatan rangkaian mempunyai fail log yang mencatatkan semua aktiviti yang berlaku dalam rangkaian berkenaan. Penggunaan fail log daripada peralatan rangkaian adalah salah satu cara untuk mengesan dan menganalisis serangan. Dalam projek ini, LoFA-SD telah mencadangkan pendekatan menggunakan fail log untuk mengesan dan menganalisis serangan. Pendekatan ini menggunakan pengesanan corak dalam fail log peralatan rangkaian. Dalam proses pencarian data ke pengesanan kesan serangan ke perbandingan corak serangan sehingga penjanaan laporan akan dapat membantu pegawai rangkaian dalam mengenalpasti pendedahan serangan yang dihadapi serta serangan yang berpotensi.

TABLE OF CONTENTS

CHAPTER		TITLE	PAGE
		DENT'S DECLARATION	ii
	SUP	ERVISOR'S DECLARATION	iii
	DED	DICATION	iv
	ACK	KNOWLEDGEMENT	v
	ABS	TRACT	vi
	ABS	TRAK	vii
	TAB	BLE OF CONTENTS	viii
	LIST	Γ OF TABLES	xii
	LIST	r of figures	xiii
	LIST	T OF APPENDICES	xvi
1	INT	RODUCTION	1
	1.1	Introduction	1
	1.2	Problem Statement	3
	1.3	Objectives	3
	1.4	Scope	4
	1.5	Thesis Organization	4

2	LITI	ERATU	RE REVIEW	6
	2.1	Introduction		6
	2.2	Log F	ile	7
	2.3	Intrus	ion Detection	8
		2.3.1	Signature Detection	8
			2.3.1.1 False Positive	11
			2.3.1.2 False Negative	12
		2.3.2	Anomaly Detection	12
	2.4	Comn	non Attacks in the Network	13
	2.5	Study	of Current System	14
		2.5.1	KIT-1	15
,		2.5.2	Honeycomb	16
		2.5.3	Snort	19
		2.5.4	Signature Extraction Approach	21
	2.6	Softw	are Approach	24
	2.7	Progra	amming Language	25
		2.7.1	Java Programming Language	25
		2.7.2	Perl Programming Language	25
3	MET	'HODO	LOGY	27
	3.1	Introd	uction	27
	3.2	Projec	et Method	28
		3.2.1	Requirement Definitions and Analysis Phase	29
			3.2.1.1 Preliminary Investigations	30
			3.2.1.2 System Requirements	31
			3.2.1.3 System Flowchart	31
		3.2.2	System Design	34
			3.2.2.1 Data Flow Diagram (DFD)	34
			3.2.2.2 Use Case Diagram	37
			3.2.2.3 Sequence Diagram	38
		3.2.2.4 Interface Design		41

		3.2.3 Database Design	43
		3.2.3.1 Entity Relationship Diagran	n · 44
	•	3.2.3.2 Data Dictionary	45
		3.2.4 Implementation and Unit Testing	46
		3.2.5 Integration System Testing	46
	3.3	Software and Hardware Tools	47
4	IMP	LEMENTATION	49
	4.1	Introduction	49
	4.2	Topology	50
	4.3	Setting Up the Syslog Server	51
	4.4	Configure Network Devices to Send Syslog	g 54
		Messages to Syslog Server	
		4.4.1 Cisco Catalyst 3560 Series Switcher	s 55
		4.4.2 Windows 2003 Server	55
	4.5	Application System	56
		4.5.1 Main Interface	57
		4.5.2 Log Analysis Interface	58
		4.5.3 Report Interface	62
	4.6	System Database	65
		4.6.1 Signature Database	66
		4.6.2 Syslog Server Database	67
5	RES	ULT AND DISCUSSION	68
	- 5.1	Introduction	68
	5.2	Results	69
		5.2.1 Logs from the Syslog Server	70
		5.2.2 Application System	71
	5.3	Project Limitation	74
	5.4	Development Constraint	75
	5.5	Time Constraint	75
	5.6	Future Enhancement	76

6	CONCLUSION	77
	REFERENCES	78
	APPENDIX A-C	80

LIST OF TABLES

TABLE NO.	TITLE	PAGE	
2.1	Types of Attacks in Network Devices	13	
2.2	The Software Development Tool and Its	24	
	Specification		
3.1	Requirements Needed for LoFA-SD	31	
3.2	Data Dictionary for Signature Details	45	
3.3	Data Dictionary for Intrusion Details	45	
3.4	Software Tool Used	47	
3.5	Hardware Tool Used	48	

LIST OF FIGURES

FIGURE NO	TIGURE NO TITLE	
2.1	A Portion of One of UMP Router's Log File	8
2.2	An Example of Content in a Signature Database	10
2.3	An Example of Attack Signature	11
2.4	An Example of Attack Signature Pattern	11
2.5	The Protocol Used in KIT-1	15
2.6	KIT-1 System Architecture	16
2.7	High-level Overview of Honeycomb Signature	17
	Creation Algorithm	
2.8	The TCP Packet Exchange and the Way Honeycor	nb 18
	Traces the Connection	
2.9	The Standard Snort Rules that Pertain to Q Trojan,	19
	Found in the "Backdoor.Rules" Files	
2.10	A Portion of Snort Rule Chain Logical Structure	20
2.11	SNORT Pattern-Matching Code	21
2.12	General Architecture of the System	21
3.1	A Modified Waterfall Model for the Development	29
	of Log File Analysis Using Detection Mechanism	
3.2	Flow Chart for Log File Analysis	33
	Using Signature Detection (LoFA-SD)	

3.3	Context Diagram for Log File Analysis	35
	Using Signature Detection (LoFA-SD)	
3.4	Data Flow Chart for Log File Analysis	36
	Using Signature Detection (LoFA-SD)	
3.5	Use Diagram For LoFA-SD	37
3.6	Sequence Diagram for Analyze Log	38
3.7	Sequence Diagram for Enquire Log from Syslog	39
	Server	
3.8	Sequence Diagram for View Report	40
3.9	Front Page Interface Design for LoFA-SD	41
3.10	Log Analysis Interface Design for LoFA-SD	42
3.11	View Report Interface Design for LoFA-SD	43
3.12	ERD for Signature Database	44
4.1	LoFA-SD Implementation Topology	50
4.2	Kiwi Syslog Server Banner	51
4.3	Kiwi Syslog Server Setup (i)	52
4.4	Kiwi Syslog Server Setup (ii)	53
4.5	Kiwi Syslog Server Setup (iii)	54
4.6	Switch Configuration	55
4.7	Log Forwarder Configuration	56
4.8	Netbeans IDE 7.0 Beta 2 Startup	56
4.9	Main Menu of LoFA-SD	57
4.10	Log Analysis Interface	58
4.11	Variables for Database Connection	59
4.12	Connection to Database	59
4.13	Retrieve Logs Error Message	60
4.14	Calendar Window	60
4.15	Empty Data Warning	61
4.16	Signature Pattern Matching Code	61
4.17	Pie Chart Report	62
4.18	Part of Pie Chart Generating Coding	63
4.19	Bar Chart Report	64
4.20	Part of Bar Chart Generating Coding	65
4.21	Signature Details Table	66

4.22	Report Information Table	66
4.23	Log Files Table	67
5.1	Syslog Server Logs	70
5.2	Log Files Retrieval	71
5.3	Pie Chart Report	72
5.4	Bar Chart Report	73
5.5	Analysis Report Email	73
5.6	Analysis Report Email Notification	74

LIST OF APPENDICES

APPENDIX	TITLE	PAGI	
Α	Gantt Chart	80	
В	Source Code	82	
С	User Manual	108	

CHAPTER 1

INTRODUCTION

1.1 Introduction

Due to the widespread of networked servers, workstations, routers, switches, bridges and other network devices has also made the number of threats against the networks and systems increased. This has created a need for network security to secure the network. Each network devices contains log files which are a record of events occurring within their network. Logs are composed of log entries and each entry contains information related to a specific event that has occurred on a specific time. These logs are important for security purposes because it might contain records of data that can be used to track user authentication attempts and also possible attacks.

In this project, titled Log file Analysis Using Signature Detection (LoFA-SD) is using the log files from network devices such as routers and switches and sends to

computer for analysis. The log files are sent to the syslog server and then will be transferred to the analyzing system. Detection mechanism is applied in the analyzing system to detect attacks, policy violation, fraudulent activities, operational problems and also the abnormalities of the logs. There are two types of mechanisms which are statistical anomaly based detection and signature based detection.

Statistical based detection uses a baseline in order to detect the activity logs of the network devices are within the baseline parameter. If the activity is outside the baseline parameters, then there is a possibility that an attack has occurred.

On the other hand, signature detection uses preconfigured and predetermined attack patterns from the log files known as signatures. Many malicious attacks have distinct signatures. Therefore it is essential to have a bank of signatures and constantly updating it to mitigate emerging threats for a good security practice.

In this project, signature detection is used and pattern matching will be implemented to detect those harmful activities and abnormalities. Once it is detected, it will alert the administrator and also generate a report out to be used for security implementation. The report will consist of the details of the attacks and abnormalities to ease network administrator to manage the network. It can also be used to analyze more than one network devices at a certain time to generate out report for comparison used.

1.1 Problem Statement

In the world today, organizations need to know about their network devices such as routers and switches activities. The lack of knowledge in the devices activities has made it hard to be secure. Every network devices contain logs file that log in all the incoming data and also the outgoing data. The data are hard to read and interpret. Therefore, a tool is needed to translate all the data into readable format.

Besides, the poor notification system in the router and switches because it does not alarm network administrators when attacks or traffic originating from users is being used for ungainly purposes. Therefore, administrator could not make a timely modification to either the configuration or the software image of the network devices itself to decrease the threats and impact of an attack or potential attack.

1.2 Objectives

- i) To create a system that will analyze logs from different network devices.
- ii) To apply notification function for the purpose of alerting user when there is an attack found in the log file analysis.

1.3 Scope

- i) The user of this system will be network administrator.
- ii) The system can read log files from different vendor of network devices.
- iii) Only signature detection will be used to detect malicious attack or abnormality of the log file.
- iv) The network used is the UMP Gambang network.
- v) The system will only detect few types of attacks which are DoS attack, ping flooding attack, port scanning attack and UDP flooding.

1.4 Thesis Organization

This thesis contains six chapters. Chapter one gives an overview of the research conducted. Chapter two explains about research that is done regarding to this project. This chapter divided into two major parts namely, research on the existing system and about techniques and technologies that is related to this project. The research is based on the previous paper or research that had done by other scientist or any current systems that implements the techniques related to this projects. This chapter also explains about techniques or technologies relevant to this project. In Chapter three, the approach or overall framework about the development of project is discussed. This includes techniques, methods, or approaches that is used to develop and implemented throughout the project development. Chapter three also explains any justification of the techniques, hardware, software and methods that is used. Chapter four will document all the process that is involved in the development of the projects. It includes all the implementation and testing done for the project. Chapter five will discuss about the findings or result that is obtained and analysis of the data. This chapter also includes the result analysis,

project constraint and suggestion for improvement. Finally, in Chapter six will conclude overall projects that had developed. This includes the project summary, the summary of the data that is obtained and the effectiveness of data obtained with the objectives and problem statement. This chapter also discussed about the summary of methodology and implementation that is used throughout the project.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter will outline the general overview of any domain studies that is related to this project. This is purposely to increase the knowledge and understanding about the background of this project. This chapter also explains any research made that related to Log File Analysis Using Signature Detection (LoFA-SD) regarding on the approach that will be used. Besides, the chapter also includes the study on the meaning of log files, log file analysis and also detection techniques or methods that is currently used to deal with any log files in the network devices and discuss more specific about LoFA-SD. In addition, in this chapter also will have the brief information about the existing systems, the discussion of the software approaches and programming languages generally used.

2.2 Log File

Basically, log files contain logs that are records of events occurring within an organization's system or network (Kent and Souppaya 2006). Logs are also one of the most fundamental resources in the security field (Slagell, Wang and Yurcik 2004). Therefore, logs are very important as they contain important data that can be useful in securing the network. In the past, logs were generally used for troubleshooting problem but nowadays, log serve many functions within the organization such as optimizing the system and network performance, recording the action of a user and also providing information and data useful in investigating malicious activity.

According to Valdman (2001), the process of generating log file is easy and straightforward but it is often very large in size and has a complex structure. In addition, not everything in the log file can be used and this eats up quite a large space in the storage devices.

In the research paper of Hulshof (Hulshof 2001), log file analysis is the systematic approach to examine and interpret the content of behavioral data. We can be able to detect behavioral pattern and apply security counter measures to improve the network. Log file analysis also allows network administrators to detect if attacks are being orchestrated against their network and also detect what is considered normal traffic originating from the trusted user is being used for ungainly purposes.

```
15 04:21:53:878 2005 Black_A_S500G-I_24PartSFP FIB/S/PoRT LINK STATUS CHANGE:- 1 - abitEthernet1/0/2: is UP

16 00:00:06:430 2005 Black_A_S500G-I_24PartSFP FIB/S/log_fib:- 1 - synchro state: 2005/01/16 00:00:06 Regin FIB's periodical refreshing synchro state: 2005/01/16 00:00:06 Finished FIB's periodical refreshing synchro state: 2005/01/16 00:00:06 Finished FIB's periodical refreshing synchro state: 2005/01/16 00:00:06 Finished FIB's periodical refreshing synchro state: 2005/01/17 00:00:06 Begin FIB's periodical refreshing synchro state: 2005/01/17 00:00:06 Finished FIB's periodical refreshing synchro state: 2005/01/17 00:00:06 Finished FIB's periodical refreshing synchro state: 2005/01/17 00:00:06 Finished FIB's periodical refreshing synchro state: 2005/01/18 00:00:06 Finished FIB's periodical refreshing synchro state: 2005/01/20 00:00:06 Finished FIB's periodical refreshing sync
```

Figure 2.1: A Portion of Router's Log File

2.3 Intrusion Detection

There are two types of approach in intrusion detection. The approaches are signature detection and anomaly detection.

2.3.1 Signature Detection

Signature detection uses signatures to evaluate the network traffic for identifying intrusions and malicious attacks. This is because when a malicious attack is launched against network devices or any systems, the attack typically leaves evidence of the intrusion in the log files. Each intrusion leaves a kind of footprint behind (e.g., unauthorized software executions, failed logins, misuse of administrative privileges, file and directory access) that administrators can document and use to prevent the same attacks in the future. Due to the fact that each signature

is different, it is possible for network administrators to determine by looking at the intrusion signature what the intrusion was, how and when it was perpetrated, and even how skilled the intruder is.

The signature often consists of one or more specific binary patterns found in a given network packet (Thakar, Dagdee and Varma, July 2010). Continuous updates of the signature database are essential since as intrusion detection system is able to recognize an attack only when the signature for it is known. Also, continuous efforts are required to detect new attacks and determine appropriate signatures from them. Moreover, a slight change in the attack scenario may be enough to alter the attack signature and thus fool a signature filter. They are consequently vulnerable to polymorphic attacks and other evasion techniques which are expected to grow in the near future. The creation of these signatures is a tedious process that requires detailed knowledge of each software exploit that is to be captured and a large pool of ASCII-log data to analyze.

In a research paper by Neelakantan and Rao (2008) also stated that signatures are globally known vulnerable patterns based on operating systems, protocols and applications. Signature detection focuses on the detection of attacks in the network which is describe by a pattern known as signatures. The patterns are referred to as signatures which are specific activities/behaviours which are of interest to detect (Fooladvandi, et al. 2009). According to Thakar, Dagdee and Varma(2010), a signature can be a portion of code, a pattern of behavior, a sequence of system calls, and others. There is currently no common standard for defining these signatures. As a consequence, different systems provide signature languages of varying expressiveness. Content based signature generation (Kim and Karp, 2004) is process of extracting the attack signatures based on selection of the most frequently occurring byte sequences across the flows in the suspicious flow pool. In order to do so, various algorithms like (Longest Common Substring (LCS) are applied to extract the common patterns in it since malicious payload appears with increasing frequency as the malicious activity spreads.

Intrusion detection signature is the description of the characteristic of attacks elements. In the research paper by Kreibich and Crowcroft (2003) stated that a good

signature should be narrow enough to capture precisely the characteristic aspects of exploit it attempts to address while it should also be flexible enough to capture variation of attacks. Schmerl, et al. (2010) also agrees that the effectiveness of this type of detection strongly depends on the conciseness and the topicality of the applied signatures. Imprecise signatures heavily limit the detection capabilities of the detection mechanism and failure in any way of it will cause either a large amounts of false positive or false negatives.

	T	
Field Name	Field Content 1	Field Content 2
Protocol ID	53	54
Protocol Name	SIP	RTP
Layer	Application Layer	Application Layer
Description	A protocol used for session initiation	A protocol used for real-time transmission
Field Name	Field Content 1	Field Content 2
Protocol ID	53	53
Field ID	1	5
Field Name	Start Line	From
Description	To distinguish requests from responses	The sender of the message
Type	String	String
Pattern	INVITE	sip:alice@domain.com
Stand-alone	False	False
Next Protocol ID	53	Null
Next Field ID	5	Null
Impact	INVITE requests from <i>Alice</i> should not be received for administrative reasons	Alice should not be

Figure 2.2: An Example of Content in a Signature Database

Field Name	Field Content
Protocol ID	53
State ID	30
State Name	BYE Received
Description	The system state after receiving BYE
Threshold	Null
Time Unit	Null
Timer	20 MSEC
Recommended Action	BYE_Procedure()
Impact	Such action causes Denial of Service (DoS) at the endpoint

Figure 2.3: An Example of Attack Signature

Attack	Attack signatures	False+	False-
CodeRed II	GET/*.ida?*\r\n??????\%u*\%u780*=*HTTP/1.0\r\n	0	2%
Apache-Knacker	GET~*~HTTP/1.1\r\n*:~*\r\nHost:~*\r\n*:~*\r\nHost:~ *\xFF\xBF????????\r\n	0	0
HSPrinter	GET?http://*:*\r\n??????????null.printer?*HTTP/1.0\r\n	0	5%
TSIG	\xFF\xBF*\x00???????????\x00\x00\xFA	0	6%

Figure 2.4: An Example of Attack Signature Pattern

2.3.1.1 False Positive

A false positive is when any detection system or firewall detects a specific vulnerability in a process or a program which in fact does not exist. There is sometimes in some conditions when scanners tend to make mistakes when attempting to find vulnerabilities in the targeted objects. We can use the term "false alarm". These are due to the errors and mistakes which exist in the signatures or the "check logic". Unfortunately, false positive will continue existing as it is hardly possible to have any signatures that are perfect in shape. However, this problem can be minimized and limited by the skill of the developers who write the signature and check logic. The problem does not exist due to the skills of these developers, but it is

hard to predict the custom environments different users are in (CgiSecurity.com 2000-2008).

2.3.1.2 False Negative

A false negative is basically the opposite of a false positive. While running a scanning process, the system may miss a vulnerability which in fact exists. The reason for this to happen is perhaps the check is not yet being written due to the vulnerability exists is too new, the error from the user who may not select the right policy and so on.

2.3.2 Anomaly Detection

This approach of intrusion detection detects computer intrusions or misuse by monitoring activities or traffics in the network which is classified as out of the normal. All the classification is based fully on heuristics or rules that result in identifying any type of intrusions which cause any abnormal activities in the normal operation activities. This approach will involve the recognition of normal traffic behavior in order to identify intrusion or attack traffic. Three broad categories of anomaly detection techniques exist, which are supervised, semi-supervised and unsupervised (Hans-Peter, Kroger and Zimek 2009).