A Secured Log Mining Approach to Collection, Monitoring, Rotation, and Analysis of Frequent and Heterogeneous Logs

Surendra Gaur[1], Nafisur Rahman[2]
{surendragour@gmail.com[1], nafiis@gmail.com[2]}

Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi[1, 2]

Abstract. Organizations tend to develop their IT infrastructure in a way that complies with Network and Information Security standards. Logs play an important role with respect to security. It is very difficult to collect the logs from various hosts in real-time and analyze these raw log messages without the help of analysis tools. In this paper, we propose a log mining approach employing Centralized log server and Analyzer tools for extracting information that matters from a huge amount of log messages and displays messages. We begin by introducing the challenges faced by the internet world and the role of logs in this connection. Then we describe the problem of extraction and making sense of logs. Then we discuss the architecture and the entire workflow of the proposed solution and the method of log collection, monitoring, rotation, and analysis. Finally, we discuss the result report sample that establishes the usefulness of our approach.

Keywords: Log, Security, Mining

1 Introduction

The world of Internet is growing at an unprecedented pace. It is a big challenge to maintain Confidentiality, Integrity, and Availability of services to authorized users. Technology is advancing day by day and cryptanalysts try to breach the security of the system to gain unauthorized access. The number of cybercrimes is increasing like never before.

Logs are one of the most fundamental resources for any security professional to rely upon. It is widely recognized by the government and industry that it is both beneficial and desirable to share logs for the purpose of security and research [1]. Logs are one of the most critical and beneficial information for security purposes. It is important to maintain a log management system to collect forensic evidence from various devices, like firewalls, Servers and Hard disks, etc. Log messages are also used to detect the violation of the security policies of an organization. Log management ensures that the security logs get stored in a specific format for a predefined period as per the available guidelines. An audit log is the simplest yet one of the most effective forms of tracking temporal information. The idea is that whenever something significant happens, we write some records indicating what happened and when it happened [2]. Log management is very much required for an organization or an enterprise to combat the attackers and intruders by protecting and securing log data from the intruders, who are trying

ICIDSSD 2020, February 27-28, New Delhi, India

DOI 10.4108/eai.27-2-2020.2303252

to modify or wipe out the log data to abolish evidence. For the purpose of the diagnosis of the reported crime, designated security agencies require the logs from the Internet Service provider (ISP) to use the logs as evidence.

2 Problem Description

Internet and Intranet-based applications produce very huge amounts of logs like access logs, database logs, system logs, policy violation logs, etc. It is a challenge for the administrators to extract the relevant information from the huge volume of logs which is generating on a daily basis.

Sources of Log Messages:

Commonly, centralized logs, stored in the log directory, can be classified as:

(a) Application logs: Almost all enterprises depend on a variety of commercial applications for their day to day activities like email services, web services, Supply chain management, ERP, etc. These applications generated a variety of logs.

(b) *Authentication Logs:* If we seek the information related to the user authorization, we can get this from the authentication log file. For example, the Diagnosis of brute-force attacks failed login attempts, etc.

(c) Security Logs: These logs are similar to the authentication logs. It stored the security-related logs including authentication failure. It also contains the details of successful logins and activities of authorized users.

(d) System Logs: All the boot-up messages are stored in this log file. This file stores booting related information messages logged during the system start-up process. With the help of this file, we investigate the issues related to the unplanned reboots, improper shutdown or boot failure, etc.

3 Proposed Solution

There are many vendor solutions available for log analysis and log management in the market. These do not generally yield efficacious outcomes. The log analysis solution should employ a Graphical User Interface (GUI), log monitoring and fetching in real-time, Log Analysis, and should fulfill the requirements of Network Security and audit.

An enterprise has various applications that are running on different platform servers and all these servers have antivirus, antimalware, VPN clients, FTP agents, etc. In this scenario, it is very tough to diagnosis the occurrence of the specific incident from the huge volume of the logs. Log analysis solution should consist of a Centralized log server, Log filtration mechanism, Database server, Log rotation program, and functionality to perform query and report generation.

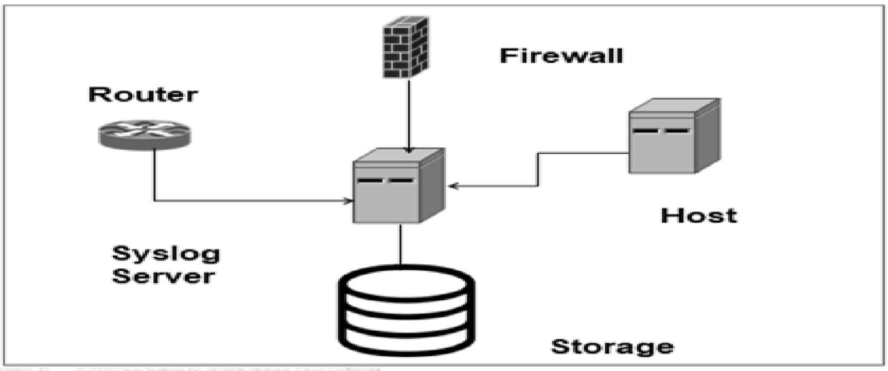


Fig.1. Schematic Representation of Syslog Architecture

The proposed model considers the storage and annotation of logs of an enterprise's servers and networking devices at a centralized location and protects this information from the intruders.

There are three phases of the proposed solution:

Phase-1: At Phase 1, we configure the servers/routers/firewalls for sending the logs to centralized log servers in real-time. Logs are fetched from various networking devices and transferred to a remote log server. Log monitoring is based on content filtering and message alerts. If the severity of the message is critical, an email script is generated and the System/Network Administrator is intimated.

Phase-2: At phase 2, Extracted log data consistency is checked using a log rotation script if there is a log rotation policy. If the policy for log rotation is defined, the logs are archived to the storage device. Now, the log data is encrypted and time-stamped based on the log messages received from various sources. This phase is primarily for Storage and line up of log data in the database.

Phase-3: Phase 3 is related to analysis, querying, and report generation. In this phase, the Log analyzer engine generates the reports as per the inputs provided by the user. For example, if the administrator wants to know the activity done by a particular user on a specific device, it will generate a query using the user name.

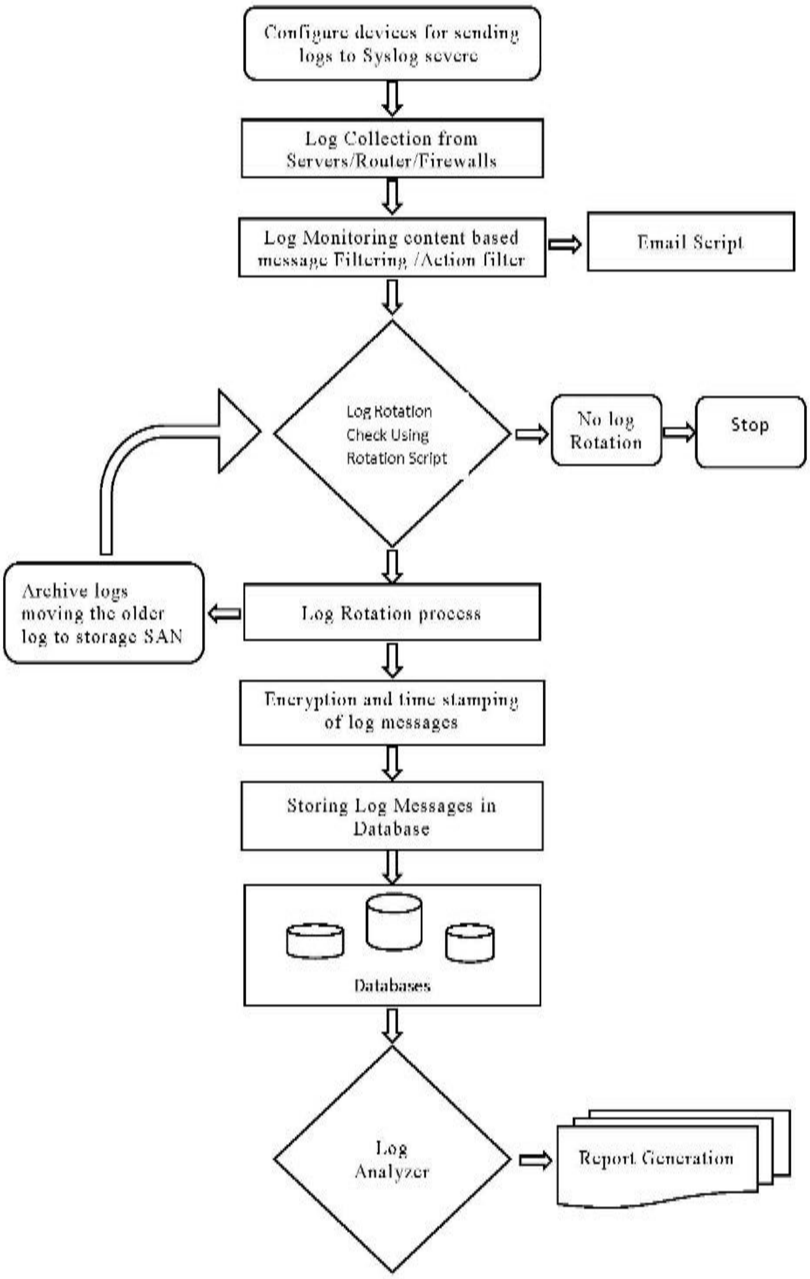

Configure devices for sending logs to Syslog severe
Log Collection from Servers/Router/Firewalls
Log Monitoring content based message Filtering /Action filter
Email Script
Log Rotation Check Using Rotation Script
No log Rotation
Stop
Archive logs moving the older log to storage SAN
Log Rotation process
Encryption and time stamping of log messages
Storing Log Messages in Database
Databases
Log Analyzer
Report Generation

Fig. 2. The workflow of the Proposed Solution

4 Methodology

Syslog - the service demon running on the Linux based operating system, in the background, is managed by Syslog service. This service deals with collecting and transforms the logs in a specific manner. Port 514 is used by the Syslog for the communication. As UTP is a connectionless protocol hence there is no need for acknowledgment. Syslog used TCP/UDP transport. Syslog protocol and message format are defined in RFC3164 and RFC 3195 which defines reliable delivery of Syslog over TCP [3,4].

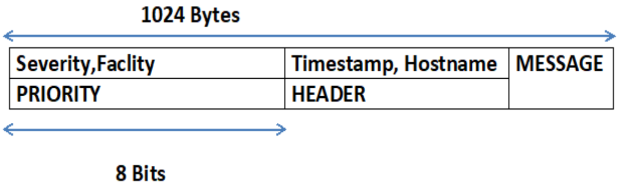


Fig. 3. Structure of Syslog Packet

Syslog has mainly three parts:

The first part is Severity and Facility, these are numerical values. Among 8 bits, three least significant bits represent the severity and other five bits related to the facility.

Severity: It specifies the severity of the log messages. For this, it uses the single-digit numbers as shown in the table.

Table 1. The severity of Log Messages

Numerical Code	Severity
0	Emergency- System is unusable
1	Alert: Action must be taken immediately
2	Critical: Critical Conditions
3	Error: Error Conditions
4	Warning: Warning Condition
5	Notice: Normal but significant Condition
6	Information: Information Messages
7	Debug: Debug-Level Message

Facility: It specifies which program is generating log messages. RSyslogd service configured to maintain logs from different sources differently.

Some Syslog Message Facilities with their numerical code are shown in the following table:

Table 2. Syslog Message Facilities

Numerical Code	Facility
0	Kernel messages
1	User-level messages
2	Mail system
3	System daemons
4	Security /authorization messages
5	Messages generated internally by Syslog
6	Line printer subsystem
7	Network news subsystem

8	UUCP system
9	Clock daemon
10	Security/authorization messages
11	FTP daemon
12	NTP subsystem
13	Log audit
14	Log alert
15	Clock daemon (note 2)
16	Local use 0 (local 0)
17	Local use 1 (local 1)
18	Local use 2 (local 2)
19	Local use 3 (local 3)
20	Local use 4 (local 4)
21	Local use 5 (local 5)
22	Local use 6 (local 6)
23	Local use 7 (local 7)

The second part constitutes the header.
Header: Header has two fields Timestamp and Hostname.
Timestamp: It is used for the local time, in MM/DD HH:MM:SS format when the logs generated by the device. Using the NTP server for synchronization is best practice.
Hostname: Hostname is the name of the host (Device name as configured) or IP address. In the case of multiple interface devices, Rsyslog considers the Interface IP address from which logs are transmitted.
The third part of Syslog contains the actual log message.

Syslog service is a methodology that collects the logs from various sources and store into a centralized location. A centralized log server generally having the following components:

(i) Rsyslog service: Rsyslog service is responsible to receive the logs from the various sources, transform it over a network. Devices configure as a client so that they transmit their logs to a Syslog server. Syslog services also configured in the dual-mode so that it can run as a client or server. By default, the Rsyslog client sends the log messages in plain text format.

Syslog configuration file has three sections

a. The first section is that of Modules. The modules section provides the local system support for logging via logger commands.
b. The second section is the Global directives. It is default templates for the Rsyslog file format.
c. The third section is the Rules. This section specifies the action taken when a particular match occurs.

(ii) Database: Servers and Networking equipment in an enterprise generate a large number of logs. There is a requirement of the Database server to store and retrieve Syslog. In the proposed solutions we used Maria dB database, which is open-source. We create a Syslog database under this database there are the following tables:

Table 3. Syslog Database Tables

Serial No.	**Table**
1	SystemEvents
2	SystemEventsProperties
3	Logon_charts

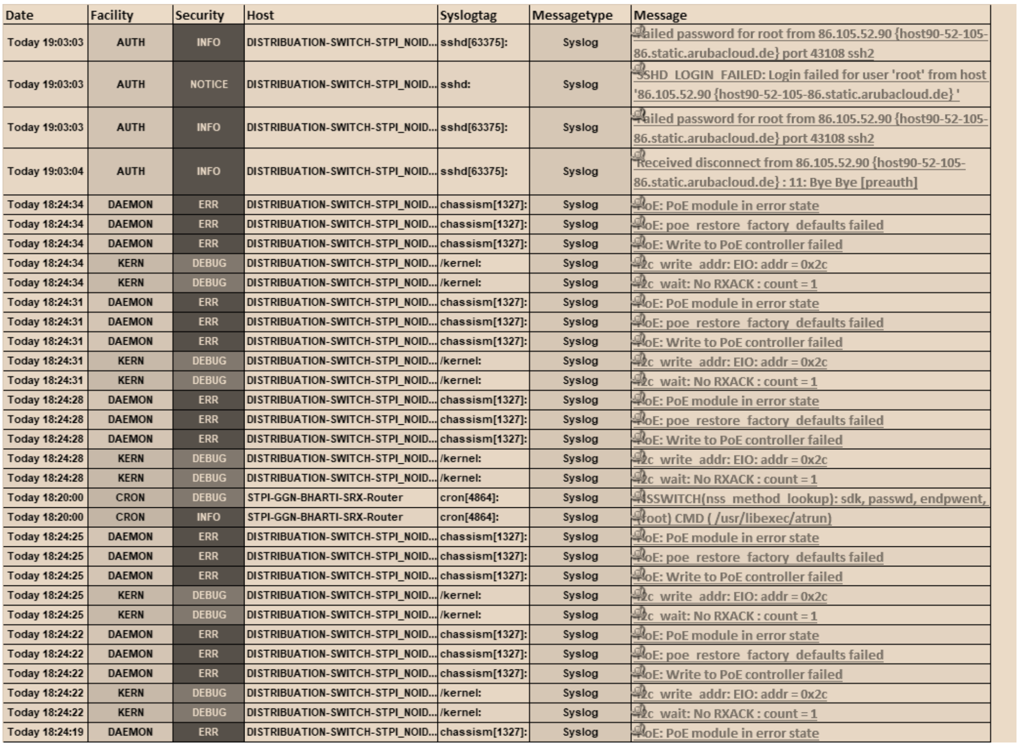
4	Logcon_config
5	Logcon_dbmappings
6	Logcon_fields
7	Logcon_groupmembers
8	Logcon_groups
10	Logcon_savedreports
11	Logcon_seraches
12	Logcon_sources
13	Logcon_users
14	Logcon_views

Network administrators and system administrators monitor the events and alarms that occur in Network and take immediate action on a daily basis.

(iii) Query and Management of Logs: As there are a large number of logs are generated in an Enterprise. It is inconvenient to extract the specific log data whenever required. For the said purpose various third party log analyzer tools are available in the market. These tools access the Syslog database and perform the various query operations and then present the result in the graphical form.

Logs Filtering and Report generation: Log data received from the various networking devices and servers are converted into the specified format as per the requirement of the organization. There are various criteria available for filtering these log data and generating reports as per the requirement of the enterprise.

Log Analyzer application has the capability to generate customized report for example Network address translation (NAT) report, Event severity, SSH Access Report, etc. Reports are displayed in table format as well as Graphical view.



Date	Facility	Security	Host	Syslogtag	Messagetype	Message
Today 19:03:03	AUTH	INFO	DISTRIBUATION-SWITCH-STPI_NOID...	sshd[63375]:	Syslog	ailed password for root from 86.105.52.90 {host90-52-105-86.static.arubacloud.de} port 43108 ssh2
Today 19:03:03	AUTH	NOTICE	DISTRIBUATION-SWITCH-STPI_NOID...	sshd:	Syslog	SSHD_LOGIN_FAILED: Login failed for user 'root' from host '86.105.52.90 {host90-52-105-86.static.arubacloud.de} '
Today 19:03:03	AUTH	INFO	DISTRIBUATION-SWITCH-STPI_NOID...	sshd[63375]:	Syslog	ailed password for root from 86.105.52.90 {host90-52-105-86.static.arubacloud.de} port 43108 ssh2
Today 19:03:04	AUTH	INFO	DISTRIBUATION-SWITCH-STPI_NOID...	sshd[63375]:	Syslog	Received disconnect from 86.105.52.90 {host90-52-105-86.static.arubacloud.de} : 11: Bye Bye [preauth]
Today 18:24:34	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: PoE module in error state
Today 18:24:34	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: poe_restore_factory_defaults failed
Today 18:24:34	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: Write to PoE controller failed
Today 18:24:34	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_write_addr: EIO: addr = 0x2c
Today 18:24:34	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_wait: No RXACK : count = 1
Today 18:24:31	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: PoE module in error state
Today 18:24:31	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: poe_restore_factory_defaults failed
Today 18:24:31	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: Write to PoE controller failed
Today 18:24:31	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_write_addr: EIO: addr = 0x2c
Today 18:24:31	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_wait: No RXACK : count = 1
Today 18:24:28	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: PoE module in error state
Today 18:24:28	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: poe_restore_factory_defaults failed
Today 18:24:28	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: Write to PoE controller failed
Today 18:24:28	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_write_addr: EIO: addr = 0x2c
Today 18:24:28	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_wait: No RXACK : count = 1
Today 18:20:00	CRON	DEBUG	STPI-GGN-BHARTI-SRX-Router	cron[4864]:	Syslog	SSWITCH(nss_method_lookup): sdk, passwd, endpwent,
Today 18:20:00	CRON	INFO	STPI-GGN-BHARTI-SRX-Router	cron[4864]:	Syslog	oot) CMD (/usr/libexec/atrun)
Today 18:24:25	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: PoE module in error state
Today 18:24:25	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: poe_restore_factory_defaults failed
Today 18:24:25	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: Write to PoE controller failed
Today 18:24:25	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_write_addr: EIO: addr = 0x2c
Today 18:24:25	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_wait: No RXACK : count = 1
Today 18:24:22	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: PoE module in error state
Today 18:24:22	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: poe_restore_factory_defaults failed
Today 18:24:22	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: Write to PoE controller failed
Today 18:24:22	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_write_addr: EIO: addr = 0x2c
Today 18:24:22	KERN	DEBUG	DISTRIBUATION-SWITCH-STPI_NOID...	/kernel:	Syslog	2c_wait: No RXACK : count = 1
Today 18:24:19	DAEMON	ERR	DISTRIBUATION-SWITCH-STPI_NOID...	chassism[1327]:	Syslog	oE: PoE module in error state

Fig. 3. Generated Report Sample

5 Conclusion

Every Enterprise requires to monitor all the activity occurs within their network and implement the security policy related to the access of the sensitive information available within the organization. For Example web applications, Firewall rules, etc.

It is very difficult for a System Administrator / Network Administrator to monitor all the devices and immediate alert in the case of any security breach, system failure, Network Performance, etc.

An efficient centralized log server provides the solutions for the above-said problems. It provides all the details of the servers and networking equipment log details at a central place so that monitoring and analysis of the log are easy and immediate action taken. This will reduce the system failure, increase the performance of the network and increase the quality of the services.

In generated report Date field show the date and time of occurrence of the event, the Facility field describes the type of logs like kernel logs, System daemon logs or Authorization logs.

Severity Field describes the severity of the log message like error messages, For Information, Warning, Notice or Debug message.

The host field shows the name of the host from which these logs are receiving. This field is very important with respect to finding the health of a specific device. Syslog tag describes the types of logs whether it is related to ssh service, cron service, Kernel or Hardware chassis.

The message field is the narration of the log messages.

References

[1] Slagell A. Yurcik W,” Sharing computer network logs for security and privacy: a motivation for new methodologies of anonymization”, Proceedings of IEEE 1st International Conference on Security and Privacy for Emerging Areas in Communications Networks(SECCMW 2005), IEEE Press, September 2005, PP.80-89, DOI:10.1109/SECCMW.2005.1588299.

[2] Mihir Bellare, Bennet S.Yee,” Forward Integrity for Secure Audit Logs”, IEEE Transactions on Information and System Security(TISC 1997), November 23, 1997, DOI:10.1.28.7970.

[3] B. Block, D. Huemer, and A Min Tjoa. “Towards More Trustable log Files for Digital Forensics by means of Trusted Computing”. In Advance Information Networking and Applications (AINA), 2010 24th IEEE Internal Conference.

[4] P.K. Sahoo, Dr. R.K. Chottary, Dr.Gunamani Jena, Dr. S. Pattnaiak, Syslog a Promising Solution to Log Management”, International Journal of Advanced Research in Computer Science, 393), May-June, 2012, 584-588

[5] Ya-Ting Fan, Shiug-Jeng Wang, “Intrusion Investigation with Data-Hiding for Computer Log-File Forensics”, Proceeding of the IEEE 5th International Conference on Future InformationTechnology(FUTURETECH2010), IEEE, May 2010, pp.16, DOI: 10.1109/FUTURETECH2010.548

[6] Gu Zhaojun, Li Young, Niu Wenjing, China Tianjin.” Analysis and Implementation of PIX firewall Syslog “Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference, Chengdu, 16-18 April 2010.

[7] Anand Deveriya, “An overview of the Syslog Protocol. Cisco Press, 2005.

[8] Jian-huang, Man-qi Zhang and Yuanlong Jiang, “The design and implementation of the Centralized log gathering and analysis system” Published in 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), 25-27 May 2012.

[9] Bauer, Michael D., Chapter 10 (System Log Management and Monitoring) of Building Secure Servers with LINUX, O’ Reilly, 2002.

[10] Babbin, Jacob et al, Security Log Management: Identifying Patterns in the Chaos, Syngress, 2006.