Log Management comprehensive architecture in Security Operation Center(SOC)

Afsaneh Madani, Saed Rezayi and Hossein Gharaee
Network Security Group, ICT Security Faculty, Iran Telecommunication Research Center (ITRC), Tehran, Iran
madani@itrc.ac.ir , saed.rezaei@aut.ac.ir, gharaee@itrc.ac.ir

***Abstract*—with the widespread use of information, variety of security logs have increased greatly, which due need for security log management. Organizations requirements have imposed to collect, store, and analyze tremendous volumes of log data across entire infrastructure for extended durations and at increasingly granular levels. It is the process of generating, transmitting, storing, analyzing, and disposing security log data from network to databases. Due to the wide variety of logs, storing comprises different methods. Recorded events in collection module are processed, normalized and classified. Logs are stored in storage module in order to use in forensic, reviewing, auditing and providing further necessities of correlation module. Routine log correlation analysis is beneficial for identifying security incidents, policy violations, fraudulent activities, troubleshooting and operational network problems. So log management is an important and efficient activity in network monitoring. Finding an effective log management functional architecture for network events analysis is the main goal of this paper. In this paper, we aim to suggest log management architecture with more common functions that are used by vendors. By studying logging architectures the main functions are administration of log collection, normalizing, categorization, queuing prioritizing and storing logged events/alarms by sensors. Log functions are different but the suitable architecture must justify the functions to send a normative, synchronized and prioritized log in an efficient way. The mentioned functions are gathered from SIEM products characteristics. Suggested architecture includes functions and activities in log collection server and storage server.**

Keywords:log management; SOC; normalizing;storage; event fiields.

I. INTRODUCTION

Security Operation Center is a centralized security organization which deals with distributed security attacks and is responsible to remove or block attacks. Treat management and incident response are done with analyzing of sensors logs. Log generation by NIDS, Firewall, OS's, Application programs and other software are out of our study scope. Security operation center comprises four main sub systems; Sensors to collect events from network traffic; Log Management to normalize, classify, prioritize and store collected logs in a storage area; Correlation Engine to analyze events; and at the end Response system to react properly against security threats, threat management and action alarm generation. This analysis is based on a knowledge base which includes attack signatures; attack behavior rules; network security policies and network configuration information. Log management is one of the main sub systems[1].

According to our study, log collection and log storage are the two main functions of log management sub system [2]. Log collection applies logging function on events and prepares the prioritized logs for storage. Besides, a copy of them is sent for analysis. Time synchronization, log queuing, conversion, normalizing, filtering, reduction and prioritization are the main functions that will be explained in the next sections. NIST 800-92 Standard explains all applying functions on logs; log generation, log storage, log protection, log transfer and at the end log analysis. Log management is implemented by different companies and each of which offered an SIEM tool, although all mentioned functions in NIST 800-92 are not implemented on these tools. The suitable architecture must justify the functions to send logs in an efficient way. These functionalities are so important on correlation results.

Logged events and alerts must be stored in a data base to retain all logs history for future statistic analyzing and forensics auditing usage. Data base have some tables to categorize stored event. Table fields contain all required information which is generated by sensors. Recorded information will be written into the data table, event table, system table, action table or log table. Formatted log data base security must be protected by security mechanisms including; confidentiality, integrity and access control. Only legal and authenticated users are allowed to have access data base information. Stored event, alerts and analysis results must be searchable by users.

The remainder of this paper is organized as follows:

In section II, challenges and main subsystem's technology are presented. After that, in section III, a broad survey of previous work relevant to log management by vendors is provided. These sections coach results to suggestion of a comprehensive architecture in section IV, and in the following, in section V, its components are introduced and explained thoroughly. Finally, we summarize this paper and conclude our work.

978-1-4577-1133-6/11/$26.00 ©2011 IEEE

II. LOG MANAGEMENT CHALLENGES

Log management architecture has some challenges that the most significant challenges would be discussed in this section [1]. The main challenge is categorizing and parsing the variety of stored logs generated by network instruments. Transferring, log filtering and reduction to save capacity are the other challenges.

A. Transferring log formats and protocols

Security appliances like NIDS creates logs in different formats including; *.evt, *.evtx, *.csv, *.xml, *.snmp and *.syslog[3]. All these formats should be transferred to log collection server. Log collection server receives logs from sensors by communicating protocols; IDMEF[4], Syslog , CEE [1] and CEF [2] for instance. Different modules in correlation system and log management system can communicate via communication protocols such as SNMP and Syslog. Main point in these connections is data transferring between modules not the language or structure of protocol, hence, simple parameter setting is enough to provide a proper connection between modules and sub systems. In addition, defining a unique format for data as log management output and correlation input is essential. Among all communication protocols, IDMEF[3] and syslog are the two most comprehensive in terms of log format which can specify event format and its components.

B. Normalizing and categorizing of information

Logged events have three types as it is categorized by NIST 800-92 [2]. In the next sections we compare various log types generated by Security software, OS's and Application software.

- Security appliances

There are many types of security software; Anti malware software, IDPS[4], RAS, Web proxies, authentication servers, Routers, Firewalls. Event log analyzing is the main feature of this category. In addition, security appliances append security information such as event ID, attack ID and rule ID which may be useful for correlation sub system. These logs are called alerts.

- Operating system

There are certain types of operating systems, but Linux is widely and typically used in SOC components. Alert and OS logs are correlated together but application logs are correlated separately.

- Application Software

All components have software agents called application programs. In correlation engine, alert correlation is analyzing alert. These alerts are generated by variety of security software like firewalls, NIDS, etc. Log correlation role is to detect multistep attacks, Distributed DOS attacks, polymorphic attacks and etc. some attacks could be detected by operating systems and security appliances. All security appliances analyze event logs separately.

C. Using logs for compliance reporting and analysis

Institutes and organizations have to monitor and audit the logs in forensic standard ways. Tracing of event source address and event detection equipment are forensic requirement. Statistical and legal issues become critical when prosecution is necessary. Compliance reporting standards examples are FISMA[5], HIPAA[6], SOX[7], etc. PCI DSS and SOX are the leading compliance drivers for log collection [1]. It is interesting to know that some vendors do not collect logs for compliance purposes!

These standards reports templates, security controlling mechanisms, policies, regularly report auditing and so on. Log storage duration time has been shown in table I.

TABLE I. DURATION TIME OF LOG STORAGE

Normative standard	Retention Requirement(Year)
SOX	7
PCI DSS	1
GLBA	6
UE Data Retention Directive	2
Base II	7
HIPAA	6 or 7
NERC	3
FISMA	3

D. Storing and archiving

1) Database Management System: (DBMS) controls creation, maintenance, and database usage in an organization and its end users[4], and allows organizations to place control of organization-wide database development in the hands of administrators and other specialists.

TABLE II. VARIETY OF DBMS SYSTEMS

Vendors	Maximum Size	Interface	Usage right
MySQL	Unlimited	SQL	GPL
PostgreSQL	Limited	GUI,SQL	BSD
MS SQL	Unlimited	GUI,SQL	Microsoft
Oracle	Unlimited	GUI,SQL,API	Sun

The leading DBMS softwares are MSSQL, MySQL, Oracle and postgreSQL which are compared considering their cost,

[1] Common Event Expression
[2] Common Event Format
[3] Intrusion Detection Message Exchange Format
[4] Intrusion Detection/Prevention System
[5] Federal Information Security Management Act
[6] Health Insurance Portability and Accountability Act
[7] Sarbanes–Oxley Act of 2002

flexibility, popularity, speed, complexity, security, maximum size, interface and usage right[6][7]. System Variaty is shown in Table II.

2) Hardware interface: Hardware interface between storage and collection servers can be used from several existing technology such as SATA, SAS, SCSI and etc. in this section these methods are compared regarding three parameters; throughput, length and the number of devices each technology support.

3) Storage Methodology: There are several architecture for storing data in networks. How to choose the appropriate type of storage methodology is various according to access and use. Data should be available through the network and shared information shoud be accessible through an authorized policy. Two important model are known as SAN[8] and NAS[9].

TABLE III. STORATE METHODOLOGY COMPARISION

Method	SAN	NAS
Characteristics	• Use Fiber channel connection • Server class devices with SCSI fiber channel could be connected. • may not exist for all OS are being used	• Utilize TCP/IP based networks(Ethernet,FddI ,ATM • Any LAN based machine connects to NAS by CIFS[10],HTTP • Defferent OS could accesss to SAN
Advantage	• High Available • High scalable • low Latency	• Heterogeneous platform support • Standard web-based admin tool • Low cost
Disadvantages	• low distance • lack of managingstandard • High cost	• High app layer overhead • lower speed • Backup
Application	• Database System • Datawarehousing • Graphic • Video Editing,	• ISP • Search Engine • Email Server • Data Warehouse • Medical Libraries
Target	• Large block size • High throughput • Write Intensive	• Small block size • High IOPS • Burst Activity

A SAN connects multiple server systems to a centralized pool of disk storage. Compared to managing hundreds of servers (each with its own disk subsystem), SAN's simplify system administration tasks. By treating all the company's storage as a single resource, disk maintenance and backups are easier to schedule and control [8].Table III shows the comparison between these methodologies.

[8] Storage Area Network
[9] Network Attached Storage
[10] Common Internet File System

III. LOG MANAGEMENT OVERVIEW

Log Management functions are the SOC's primary functionalities, so we investigate SOC productions with special focus on log management sub system. There are special products named SIEM[11] that is made by vendors to Threat monitoring and responding, network log managing and so on. Architecture of logging forces us to study SIEM mechanisms products to find the main functionalities. The commercial product would be explained in the next step.

A. Commercial Systems

A centralized SOC production is called SIEM . SIEM architecture has all SOC sub systems mentioned in introduction. Security organizations often apply SIEM to improve capabilities for external and internal threat discovery and incident management [9].

Log management functions have become an expected and standard component of SIEM technology architecture. So in landscape study, we investigated SIEM products with special focus on log management functions.

There are a number of companies which develop network protection tools and SIEM productions including Cisco, RSA, Arc Sight, and IBM. The SIEM market is achieving a growth rate of 15% during 2010 [9]. Arc sight, Q1Labs and RSA are the most leading companies in SIEM productions. Compliance reporting, threat management, security event management are their other activities. Arc sight SIEM provides user monitoring capabilities with a Unix-like command line interface for IT operations searching and reporting. The strengths of Log features are explained in Table IV.

TABLE IV. SIEM PRODUCTION FEATURES COMPARISION

Vendors	Strengths of Log features
Arc Sight	• Collection and reporting • Command line interface for searching and reporting
IBM	• Monitoring and Compliance reporting • log management • SEM section focused external threat managment • real time analysing
LogLogic	• Monitoring capability • Security event management • Network security configuration management • Date Base management • Data Base activity monitoring • can implemente in virtual environment
NetIQ	• user activity monitoring • compliance reporting • Collect analyse logs • filtered data • User activity monitoring
CA	• compliance reporting • analytics for app., hosts, security devices… • Enterprise log manager not general purpose

IBM products have log management function with user activity monitoring and involve 100 or fewer servers. It has a

[11] Security Information and Event Management

real time analysis on logs. It does not have competitive evaluation.

Log Logic has its own log management sub system and provides new protocol for secure log transport. In addition, it has real time monitoring and correlation with data base activity monitoring.

Log Logic has a strength log management sub system in all mentioned SIEM products. It provides capabilities which can be integrated with a wide variety of third-party event managers.

NetIQ is a middle-sized SIEM production. It is designed to process filtered log data.

B. *Weakneses in commercial and open source systems*

In the following table there are some weaknesses in SIEM productions with special focus on log management functionality.

TABLE V. VENDPR COUTIONS OF LOG FEATURES

Vendors	cautions of Log features
Arc Sight	• ESM software is oriented to environment • Hard database tuning activities
IBM	• Log management appliance is not yet available from IBM, just its software implemented.
LogLogic	• More efforts to extend SEM knowlage to its sales forces, sales channels…
NetIQ	• Not optimized for event management for network & security devices.
CA	• Organization that require SEM capabilities should also evaluate SEM alternatives from other vendors

IV. LOG MANAGEMENT COMPREHENSIVE ARCHITECTURE

As mentioned above, there are several methods for log management implementation. The mentioned functions are gathered from SIEM products characteristics. Suggested architecture includes functions and activities in log collection server and storage server.

Collection server is the first module for collecting received logs from log generators like firewalls, NIDS, Os's, application systems, etc. log generators send logs by transmitting protocols like syslog [10], IDMEF, CEE, CEF[11] and SNMP. Thus, collection Server must be able to understand all log formats.

Collection server applies some functions on logs prior to correlation and storing. After study of SIEM vendor architectures on log management the most important functionalities are considered as follows: Normalization, filtering, reduction, rotation, time synchronization, aggregation and integrity check.

Storage server keeps logs for forensic, auditing and off line analysis.

In addition, log security needs to be considered in all architecture. Security mechanisms are done by access control mechanisms with authentication, authorization, confidentiality, integrity check and so on.

In this paper, a new comprehensive architecture for log management has been suggested and is illustrated in the following figure (figure 1). In the next sections, the log management components are introduced.

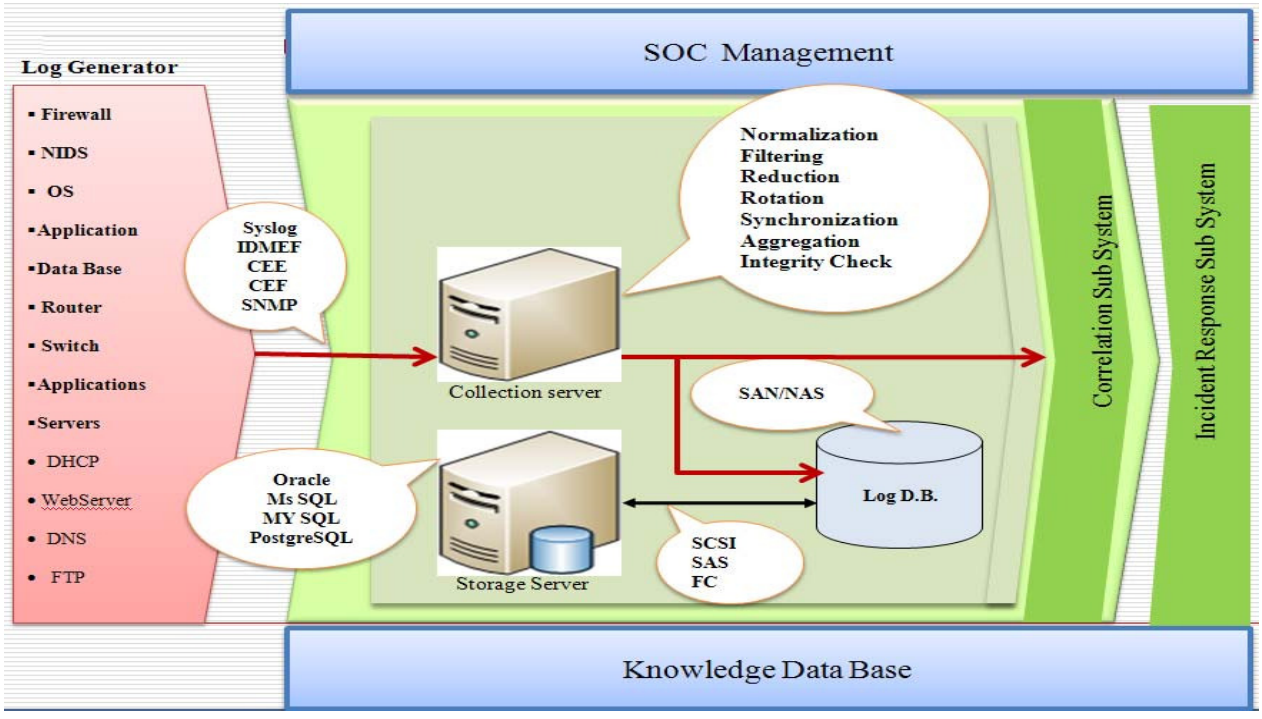

Figure 1. : Comprehensive Log Management Architecture

V. LOG MANAGEMENT COMPONENTS

Studying SIEM architecture illustrated us the main functions of logging in network. Not all of these functions are implemented in a log manager. But it is clear that log collection and log archiving must be done by any comprehensive tool. All mentioned functions could be implemented on these modules. Sometimes the functions are repeated on both collection and archive modules.

A. Collection Server

Log generation and storage is a complicated task due to several factors, including a high number of log sources; inconsistent log content, formats, and timestamps among sources; and increasingly large volumes of log data. Log server collects all received logs by transport protocols like syslog. A centralized logging infrastructure comprises central log servers with applications, servers, routers, switches and other system components acting as log clients. Syslog applies UDP in transport layer. Some syslog versions support event logging over TCP which provides guaranteed packet delivery and allows using tunneling software for message encryption [13]. Log Parsing is extracting data from a log, the parsed values can be used as input for another log process (Figure 1).

- Normalization should be able to take log events from all devices and present them in a common way for searching and analyzing in SOC correlation sub system.
- Event Filtering is the log entries suppression from analysis or storage because their information entropy is zero. NetIQ SIEM supports this functionality.
- Reduction: Omitting unnecessary entries and placing data fields of interest into log terms. This function is executed in storage phase too.
- Aggregation: similar entries are consolidated into a single entry containing the number of occurrences of the event.
- Rotation: Old log files closed and are archived when it is considered to be completed, and then a new log file is opened.
- Time Synchronization: changing log time sources in one common time source.
- Integrity check: is done in receiving and archiving logs to know that logs are not modified.
- Prioritization: in the almost time, alert has the highest priority in the received log. Prioritization of loges has been set in security policies. It can be configured by system manager.

B. Storage and archiving

The log management system provides precise and detailed log information [14]. There are different ways to save event logs. Logs will be saved for an exact duration of time as it is determined in forensic standards. Stored events are used in different activities such as forensic, monitoring and auditing, requisites in response and correlation units. Additional functions are implemented on logs; rotation, compression, encryption, conversion and archiving are the most important functions in log storing stage.

- Rotation: The default periodicity for log rotation is once a week and old logs are overwritten.
- Compression: An additional problem is the capacity of storage required for logs[15]. Compression is a storing function through which the amount of disk space needed for files is reduced, without modifying their contents.
- Archiving is retaining logs for an extended period of time, typically on removable media.
- Log conversion is parsing a log in one format and storing its entries in another format.

Storage module (DBMS) receives and responds queries. Moreover, it performs functions such as compression, encryption, authentication and integration. This sub system can be one of widely used technologies like Oracle, MS SQL, MySQL and Postgre SQL. MySQL is the most popular open source DBMS. Its development was initially based on speed and ease of use rather than functionality, although this speed is obtained at the expense of losing data durability. It has some privileges including high performance, low cost, easy learning and configuration and availability of source code and appropriate support.

VI. CONCLUSION and Future Works

Log management is one of the most major sub systems in an SOC. This module is responsible for performing functions on received logs in order to facilitate correlation process to find complex and multi-step attacks against network. Stretching first generation log management tools imposes significant trade-offs between log collection rates, log analysis speed and log storage efficiency. A next-generation, universal log management solution must eliminate the classic trade-off between performance and efficiency, and provide enterprise and infrastructure-wide visibility into log data. Unlike point solutions, it should be flexible enough so that it can be either used by individual teams or expanded into an enterprise-wide log management solution when needed.

Log data can also provide visibility into network, system and application health and availability; support security operations; and streamline network troubleshooting. It is essential to ensure that security records are stored in sufficient detail for an appropriate period of time. The study of SIEM architecture from different vendors enabled us to recognize the most significant functionalities which are usually implemented in log management systems. The proposed architecture in this paper is obtained through comparison of different methods as well as regarding high efficiency in log management. The presented model classifies not only alerts but also all kinds of received logs to prepare them for correlation. In our proposed scheme collection and storing log data are done in a distributed

manner in order to focus on different parts individually. Centralized architecture suffers from traffic processing increase and becoming a bottleneck, because there is a large amount of logs and it needs a high processing power. By separating collection and storage units some functions such as integrity check, encryption, compression, rotation and archiving are performed only in storing process.

For the future work, it is recommended that architecture efficiency is evaluated. Architecture evaluation would be done by log management implementation on real network traffic. Log capturing by sensors and normalizing will be done. Log management performance and conformance tests could be done with analyzing of logs. Number of detected attacks and the false positive, true negative ration could be show the system performance.

ACKNOWLEDGMENT

Special thanks to our coworkers in Iran Telecommunication Research Center (ITRC), ICT Security group Mr. Enayati & Mr. Sozani because of their helps for finishing of this paper.

REFERENCES

[1] J. Shenk, SANS Seventh Annual Log Management Survey Report, A SANS Whitepaper-April-2011

[2] K. Kent, M. Souppaya, NIST Special Publication 800- 92, "Guide to Computer Security Log management", 2006

[3] C.Lonvik," RFC3164- The BSD Syslog Protocol", IETF Request For Comments, 2001

[4] H. Debar, D. C. Guardian, B. Feinstein, RFC4765, " The Intrusion Detection Message Exchange Format (IDMEF) ", 2007

[5] EnterpriseDB White Paper, "PostgreSQL vs. MySQL: A Comparison of Enterprise Suitability," June 2009

[6] C. Reason, Tometa SoftwareInc. white paper, "MySQL vs. SQL Server," 2010.

[7] C. Rindal, Tometa SoftwareInc. white paper, "MySQL vs. PostgreSQL," 2010.

[8] D. A. Heger, Fortuitous Technology,"SAN and NAS Solutions - Introduction, Topology, and Terminology," 2006

[9] M. Nicolett, K. M. Kavanagh, Magic Quardrant for Security Information and Event Management, Gartner Inc. RAS Core Research Note,2011

[10] R.Gerhards Adiscon GmbH, " RFC 5424: The syslog Protocol", 2009

[11] E. Fitzgerald, A. Chuvakin, B. Heinbockle, D. Karg, and R. Marty, "Common Event Expression," November 2010, [Online]. Available: http://cee.mitre.org/docs/CEE_Architecture_Overview-v0.5.pdf

[12] ArcSight Inc. CEF "common Event Format" 2006, [Online]. Available: http://www.arcsight.com/collateral/CEFstandards.pdf

[13] R. Vaarandi, Tools and Techniques for Event Log Analysis, A PHD thesis, Tallinn university of thechnology,USA,2005

[14] T. Yue, L. Xiaobin, Y. Zhengqiu, The research and Design of Log Management System Based On Struts Frame, ISCSCT ieee conference,2008

[15] A. Tomono, M. Uehara, M. Murakami, M. Yamagiwa, A Log Management System for Internal Control, International Conference on Network-Based Information System, 2009.