Proc. of The Fifth Intl. Conf. On Advances in Computing, Electronics and Electrical Technology - CEET 2016

ISBN: 978-1-63248-087-3 doi: 10.15224/ 978-1-63248-087-3-17

Design of Information Systems (Is) Real-Time and Historical Log Events Analyzer for System Administrator

Ayhan Akbal , Mohamed Mohamud

***Abstract*— The logs generated by the operating systems, and application programs, network devices, web browsers and all devices in the information systems during their normal course of operation are very important and allow system administrators to ensure that they have a reliable information system, and track what is going on in the IS, such as what web sites clients visited, whom they are sending e-mails to and receiving e-mails from and what applications are accessed. The main goal of this study is to develop software which can parse both logs generated by the information systems (IS) in real time and logs written to historical text files, and store logs in a central database in order to perform various analyses for reporting and detecting anomaly tasks on the system. Daily log entries belonging to the information systems department of Firat University Hospital are applied with the program and various analysis results were obtained. In general, information systems logs can be easily analyzed and converted into information.**

***Keywords*— information systems, log events, logging policy, log analysis.**

I. Introduction

Today Information Systems (IS) security is important more than ever for obtaining and improving quality of company's service. Nowadays ISs are essential in commerce because it facilitates information storage and analysis, assists with making decisions, increases business processes, improves the efficiency of the operations, builds confidence in customer/supplier and all in all develops the quality of company's service [1, 10].

One important factor that has a strong effect of improving the security of IS which then improves the service quality of whole company is management of log entries carefully and professionally. A log entry is a record of the event occurring within an organization's information systems and each entry contains information related to a specific incident that has occurred within a system or network [2].

Ayhan AKBAL
University of Firat, Department of Electrical and Electronics Engineering Turkey

Mohamed Mohamud*)*
University of Firat, Department of Electrical and Electronics Engineering Turkey

These entries contain information related to system workflow produced by several sources (like applications, authentication servers, firewalls, intrusion protection systems, network access control servers, network devices (routers, switches, etc.), operating systems, anti-malware software, remote-access software, web proxies and so on [3]. Most of log entries are created by security software, such as antivirus software, firewalls, intrusion detection systems, and intrusion prevention systems and then system administrators use this information to prevent security-related threats and secure entire system's functionality [3].

In any private or corporate Information System, there are plenty of activities that occurring at any given second. So that, systems must apply a logging tool and define a policy for logging to develop a solution that can be utilized by the system administrator to monitor the activity of the system [5]. Event logging is essential to an IS because it provides the ability to secure, troubleshoot, or investigate problems that arise in the IS and also to measure and improve the performance of the system [3, 4].

Watching log events gives the opportunity to monitor the activities of the devices and applications to ensure that they are working on the expected normal situation. On the other hand it can help you to discover the low performance holes in the system and take the necessary actions to improve them [2-4].

In addition, log events are important for identifying falsified activities and policy violations, supporting internal investigations, performing forensic analyses and applying national or international legislation and regulations like Federal Information Security Management Act (FISMA), Health Insurance Portability and Accountability Act (HIPAA), Sarbanes-Oxley Act (SOX), Gramm-Leach-Bliley Act (GLBA), Internet law No. 5651, etc. [3, 5,9].

All of these issues have a direct or indirect impact on the quality of the information systems.

This study examines the development of log management program which can analyze log events generated by the IS in real-time and can review logs stored in historical text files in order to perform analyses. In this paper, we give some of the results obtained when the program was used by the information systems department of Firat University Hospital for analyzing their daily log entries.

The rest of this paper is organized as follows: In part II, log event collection and retention laws are explained. In part III, things that should be considered to achieve log management goals are described. A proposed log analysis program is introduced in part IV. In part V the results achieved when implemented the program is showed. Finally, part VI concludes the paper..

Proc. of The Fifth Intl. Conf. On Advances in Computing, Electronics and Electrical Technology - CEET 2016

ISBN: 978-1-63248-087-3 doi: 10.15224/ 978-1-63248-087-3-17

II. Logging Policy and Laws

A logging policy shapes the kind of activities that the system should track and create log event from them [6]. For example, in a system database, the logging policy may watch and capture log event from any user who is connecting to it. Just some things that can be part of the log policy include:

- All administrative actions
- All invalid login attempts
- Account impersonations
- Failed login attempts
- Password changes, etc.

Essentially, the log policy should state what types of actions need to be logged in order to secure the system from potential frauds, inappropriate use of resources, low performance and etc. [6, 7].

Besides the logging policy there are also some remarkable universal logging laws or truths mentioned by Marcus Ranum (2012), and these laws are:

A. *Collection Law*

The first law concerns the log data collection policy and it says don't collect log data that you will never use it [7, 8]. This universal logging law directly states that for every event message logged and retained, there has to be a purpose for it. This law does not apply only the collection of log data, but also relates logging log events. It is another way to say, don't log what you are not planning to use.On the other hand, this is not mean to collect partial log data, for example, to collect only security logs and ignore non-security logs. conversely , it means collecting the appropriate log data considering all the sources of log data in the IS, those are required to alert suspicious activity, provide valuable forensic data, troubleshoot, or investigate problems, etc.

B. *Retention Law*

The second law states retaining log data as long as it is possible to be used or given by the governmental regulations [7, 8]. Log data is achieved and treated like any other organizational data according to the organization's log retention policy. This law suggests that organizations have to consider maintaining a very good retention policy which can store log data as long as it might be useful or required by the governmental regulations. Alternatively, it is rather unlikely that the ordinary debugging messages will be used for years. So that, it is not a recommend to store every single log message for more than 10 years.

C. *Monitoring Law*

The third laws says log all what you can as much as possible, but alert only on what you must respond for [7].

The first and second laws made the truth of not to log data that will never be used and retaining longer log retention policy as possible, but the opposite is true in the monitoring. Actually, this law becomes more useful when the information system easily stores terabytes of log data, i.e. billions of log messages, but the security response persons can hardly address a dozen issues a day. In other words the system can log billions (xxx, 000,000,000) of log messages but thousands (xxx, 000) of them can be monitored. For that reason organizations should follow the principle of log everything, store some and monitor what is necessary which is as little as possible.

D. *Availability Law*

The fourth law expresses that logging and monitoring effort has to be done like the remaining information systems activities, but not more than so [7]. Don't take care of your logging or monitoring system more than your business systems. No doubt that log data has to be collected and analyzed, to make sure that it is available for troubleshooting system problems, protecting security threads, future audit investigations or even court cases, but it is not worth than remaining business activities.

E. *Constant Changes Law*

Logging policy and log management methods are not constant all the time. Logs sources, log types, and log messages can change [7]. The information systems' logging sources and log analyzing systems would likely produce many different and new types of log messages, especially in today's era of cloud computing and virtualization age. Therefore, it is a valuable practice for any organization to regularly review the logging policies, procedures as well as operational tasks and system configurations.

III. Goals of Log Analysis

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

After setting up log data architectures, logging policy, log collection and retention laws; log analysis goals should also be planned and prepared [3]. The goals of log analysis are what are needed to accomplish via log analysis and that can be different from one company to another depending on company's particular needs. However, there are some common goals which can apply to mostly everyone such as finding bad things which have already happened or detecting evil things that have never seen before and be able to alert on them [7]. To achieve log management goals, log analysis system should be planned carefully [3, 7]. Some keystone things that should be considered when planning a log analysis system may include [7]:

IV. Developing Log Analysis

The new log analysis program is developed using Visual studio .NET framework programming language and SQL Server 2008 R2 database. The input of the system is supposed to be log messages generated by the ISs, and the output will be analyzed information of log data, that will be provided either in a graphical representation or in a tabular format. The process of analyzing log data passes through

Proc. of The Fifth Intl. Conf. On Advances in Computing, Electronics and Electrical Technology - CEET 2016

ISBN: 978-1-63248-087-3 doi: 10.15224/ 978-1-63248-087-3-17

number of stages, which correspond to the program procedures of log management methods. Figure 1: shows a high-level view of the system's architecture and its different stages, which include:

- Normalizing Information Systems (IS) log data.
- Transferring and storing normalized logs to the system database.
- Analyzing and reporting the contents of log events.

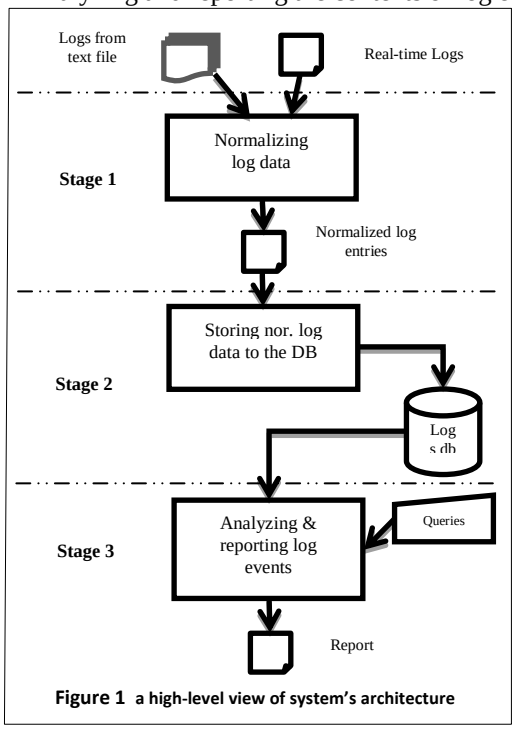

Figure 1 a high-level view of system's architecture

A. *Normalization*

Depending on the implementation platform of the system, the normalization phase reduces log entries to a minimal set of unique fields and transforms into Comma Separated Value (CSV) format.

In the first step, log records which contain similar information in terms of structure and layout are merged. Also, if the column headings are repeated in a row, they will be removed. This process focuses on reducing the size and the complexity of log entries by eliminating unnecessary data, in order to increase the efficiency of the system.

The second process of normalization parses log entries and prepares them for transmission to the application. This process convents log contents into the CSV format. CSV is a text file format in which each line represents a separate record and record is partitioned as a set of fields separated by a comma. This simplifies parsing log events to the database. An example of achieved resultant when the above processes are applied to the log messages is shown in figure 2.

B. *Storing logs to the database*

The second phase takes the normalized log entries to the database. For storing the parsed log data in the database, the program uses relational database (SQL) as its back-end database.

Figure 2 the database structure of the system

For analyzing and generating reports, the database can be queried from the program forms or an external third-party or script as well. The structure of the database used in the system is given in Figure 3.

15-03-2015 10:16:34Local0.Notice 10.81.1.2 fuhfw: NetScreen device_id=fuhfw [Root]system-notification-00257(traffic): start_time="2015-03-15 10:16:48" duration=0 policy_id=165 service=https proto=6 src zone=Trust dst zone=Untrust action=Permit sent=0 rcvd=0 src=10.8.12.118 dst=195.175.112.195 src_port=60415 dst_port=443 src-xlated ip=10.8.12.118 port=60415 dst-xlated ip=195.175.112.195 port=443 session_id=255089 reason=Creation<000>

15-03-2015,10:16:34,Local0.Notice,10.81.1.2,NetScreen,start_time="2015-03-15 10:16:48",0,165,https,6,Trust,Untrust,Permit,0,0,10.8.12.118,195.175.112.195, src_port=60415 dst_port=443

Figure 3 Example of before and after normalization a log message.

C. *Analyzing and reporting*

The final stage of the program's workflow is analyzing log contents to obtain information useful in different situations like tracking network traffics, troubleshooting information system failures, debugging application problems, getting the information needed by the government or law court, analyzing and understanding threats that could

Proc. of The Fifth Intl. Conf. On Advances in Computing, Electronics and Electrical Technology - CEET 2016

ISBN: 978-1-63248-087-3 doi: 10.15224/ 978-1-63248-087-3-17

happen in the system and more. The program uses SQL-T dialect statements like the one in figure 4 to search, filter

In general, the program has friendly interface and it is easy to use with the help of forms and prepared SQL commands for analyzing log events. Front-page screenshot of the program is shown in Figure 5.

```
SELECT DISTINCT (Field1), COUNT (Field1)
as Total FROM Table GROUP by Field1
HAVING COUNT (Field1) >= 100
```

Figure 4 Sample of transact-SQL command line.

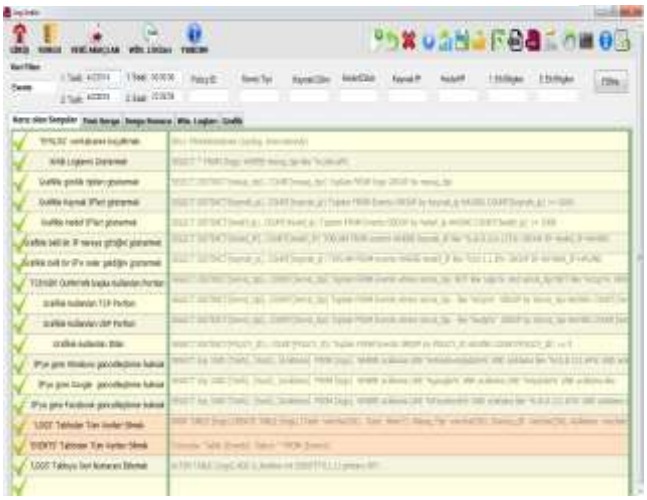

Figure 1 Front page of the program

Figure 5 Front Page of the Program

D. *Ins and Outs*

The network of the hospital can be accesses from outside world and vice versa. The internet firewall controls and monitors ins and outs of the network and keeps a log event for each. Figure 6 shows the percentage of ins and outs for the hospitals network via internet.

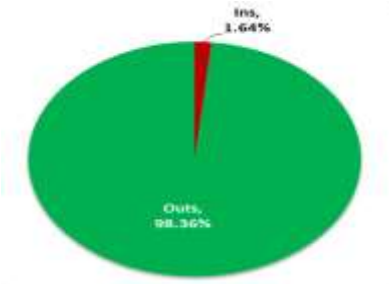

Figure 5 the ratio of ins and outs of

E. *Blocked/Permitted sites*

The firewall filters the internet traffic coming from untrusted world and network traffic going to outside world according to the network policy of the hospital. The accessibly of some websites may be blocked by the network administrators. Figure 7 indicates the ratio of blocked network traffic.

and generate reports from the database records based on the type of data stored in logs.

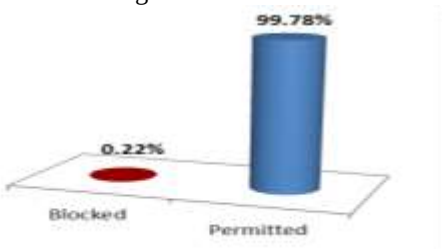

Figure 6 the ratio of blocked and permitted accesses

F. *Types of log messages*

Log events are in different types according to the priority and importance of the log message. The general categories of log messages include: notice, warring, alert, information, critical and debugging. An example of the categories ratio in log messages on daily basis is shown in figure 8.

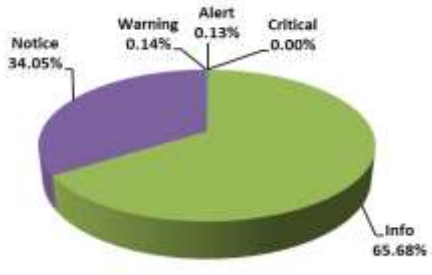

Figure 7 the percentage of log categories in a day

G. *The busiest IP addresses*

In this example we want to find 10 busiest IP addresses which are accessing the internet in a random day. We hided the IP numbers for confidentiality reasons. However, Figure 9 shows the IP addresses which are using the internet most.

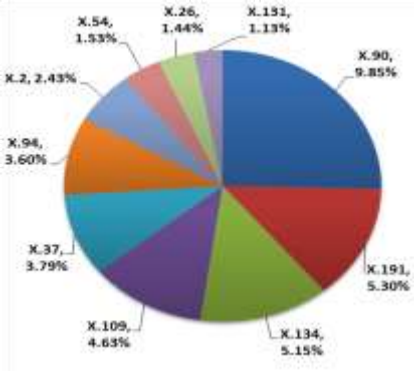

Figure 8 the IP addresses which are using the internet most

H. *The 5 most accessing IPs to the hospital network*

The network of the hospital can be accesses from the outside world through the internet in order to visit hospitals website or to manage other information. For example doctors can review and monitor patient records and network administrators can connect the system remotely to control overall workflow of the system. In this example we are showing the most 10 IP addresses which accessed our system in a random day.

Proc. of The Fifth Intl. Conf. On Advances in Computing, Electronics and Electrical Technology - CEET 2016

ISBN: 978-1-63248-087-3 doi: 10.15224/ 978-1-63248-087-3-17

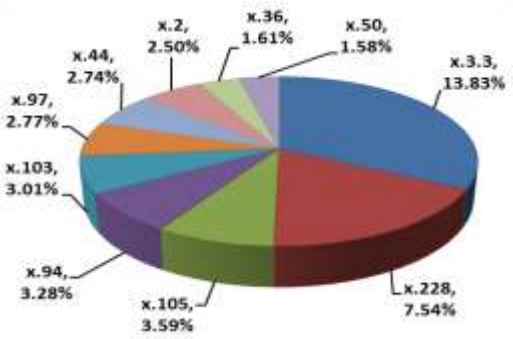

Figure 9 IPs which accessing out system most

I. *Number of logs in daily basis*

Syslog server saves logs coming from firewall to text file in daily basis. Figure 11 shows sample of the size in Gigabytes for log events generated daily.

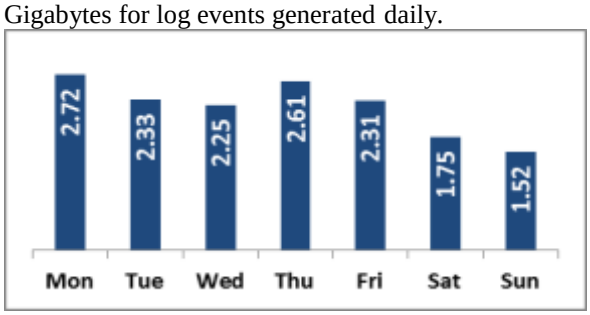

Figure 10 Sample of the logs size in a daily operation

J. *Filtering a specific website*

Using flexible filter options of the program we can search who accessed a specific website such as Facebook, YouTube, etc. This time we are displaying results in a tabular format as in Figure 12.

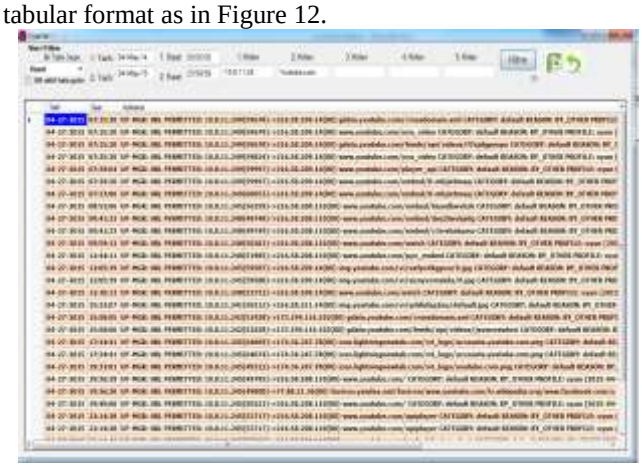

Figure 11 Example of finding who accessed to a specific website

v. Results of Implementation

Log data that is generated from different sources like firewalls, network devices and servers, etc. is quite large and it is difficult to open with text editors to analyze it. Therefore, other log analysis techniques that can handle the complexity of logs must be used. There are a number of log analyses available in the market over the years. This study also, contributes to design and development of a generic, flexible and adaptive system log analyzer, which collects logs in real time and stores in database for analyzing and reporting. It also parses text files containing historical logs and stores to the database for analysis as well.

This new log analysis program has the following features:

- Collects logs in real time and stores to a database.
- Parses text files containing log messages and imports to the database.
- Uses SQL server database to handle huge log data.
- It has flexible filter to analyze log events.
- Presents results clearly either in graphical or in tabular way.
- It has friendly graphical user interface.
- Displays Windows application and system event logs as well.

References

[1] Narasimhaiah Gorla, Toni M. Somers and Betty Wong (2010), "Organizational impact of system quality, information quality, and service quality", Journal of Strategic Information Systems 19 (2010) 207–228.

[2] Karen Kent and Murugiah Souppaya (2006), "Guide to Computer Security Log Management", NIST Special Publication 800-92, p. 2-1.

[3] Alert logic, Inc. (2011), "Log management best practices: the benefits of automated log management"

[4] Radack, S. (2006), "log management: using computer and network records to improve information security", Information Technology Laboratory.

[5] ÖNAL, H. (2009), "Standartlar ve Güvenlik Açısından Log Yönetimi", Beyazsapka dergisi

[6] Chuvakin, A. (2011), "The Complete Guide to Log and Event Management", [White Paper]

[7] Anton A. Chuvakin, Kevin J. Schmidt, Christopher Phillips "Logging and Log Management: The Authoritative Guide to Understanding the concepts Surrounding Logging and Log Management", 2013

[8] Marcus Ranum Logging Laws available at http://ranum.com/security/computer_security/archives/logging-notes.pdf (retrieved May 2012).

[9] Wawak, S. "The Importance of Information Security Management in Crisis Prevention in the Company", 6th International Symposium on Business Administration, 2010.

[10] W.B. Adeoti-Adekeye (1997), "The importance of management information systems", Library Review, Vol. 46 No. 5, 1997, pp. 318-327. © MCB University Press, 0024-2535

About Author (s):

Ayhan Akbal was born in 1977 in Elazig, Turkey. He is currently working as an assistant professor at Electrical-Electronics Engineering Department of Firat University, He received his Ph. D. degree in electrical and electronics engineering in 2008, M.S. degree in computer engineering in 2001, from Firat University, Turkey . His research interests include Communication, Network, wireless, FPGA security, wireless sensor network.