SETSCI Conference Proceedings,
4 (1): 164-171, 2019

ISAS 2019
3rd International Symposium on Innovative Approaches in Scientific Studies
Section: Engineering and Natural Sciences
April 19– 21, 2019, Ankara, Turkey

LOG MANAGEMENT WITH OPEN SOURCE TOOLS

Emre GÜL[+1], Ercan Nurcan YILMAZ[*1]

[1]*Gazi University, Ankara, Turkey*
[*]*Corresponding author: enyilmaz@gazi.edu.tr*
[+]*Speaker: emre.gul@tubitak.gov.tr*
Presentation/Paper Type: Oral / Full Paper

Abstract: In all systems within the information technologies and in the applications on these systems, the transactions performed by users and system administrators are recorded in accordance with the legal regulations and corporate policies. These records, which are one of the basic and most important components of cyber security, are given the name "log". It has become imperative to know the log management processes which are critical to the confidentiality, integrity and accessibility of information. Systems and applications produce a continuous and high number of logs. To analyze these logs and to make sense, log management software should be used. Vulnerability and intrusion attempts can be detected using the capabilities of these softwares. The relevant system administrator can be informed with automatic warnings and measures against these attacks. The purpose of this study is to explain log management processes and exhibit how to use logs to pre-identify attacks against systems. As a log management tool, Graylog application is preferred for high performance, fast indexing and free of charge.

Keywords: Graylog, SIEM, Log Management, Log Analysis

1. Introduction

Today, information systems have become a structure that we use in almost all areas of our lives. In particular, the devices serving in the information systems of organizations and applications working on these devices operate interactively with other systems in both local and wide area networks.

This situation raises the requirement to control many areas for cyber security and to develop these controls separately. This rapid development and emerging needs in the management of information systems have made this field a scientific and academic research topic and numbers of different methods and standards have been established in this field.

Monitoring of systems used within certain standards in order to ensure information security has become a necessity. In spite of all measures taken, there may be many unexpected situations in the systems due to deliberate or non-intentional attempts, system errors, or incorrect encodings that may be present in the software. Besides these, there are many different types of cyber attacks that can be effective on systems for different purposes along with developing technology. For this reason, risk assessment studies should be carried out frequently in Information Systems Management. In these studies how a security breach event can be a threat for the organization and which log records to be proactively treated should be determined beforehand in order to reduce the risk. The state of the art in security attacks, the Advanced Persistent Threats (APT) are most of the times detected by combining and correlating log files from various sources [3]. If the evaluation of the log is not done adequately and correctly and is not reported to give meaningful results, the control of the information systems' components that produce a large number of logs cannot be achieved in a real sense. This situation may result in inefficient use of investments in IT infrastructures. Information Security Management covers an important part of the management processes of Information Systems. With information security management, procedures and instructions are prepared to reduce the vulnerability of information. ISO 27001 is an Information Security Management Standard, which is important for log management in accordance with the ISO 27001 Information Security Management Standard. As a result of a complete analysis of logs taken from different systems; a more comprehensive and different view of the security situation of the systems can be obtained. In past, the logs were only analyzed when there was an anomaly situation on the systems or when there was a control need for the past. However, today, it is important to analyze the logs in order to prevent any action or attack that would disrupt the integrity of the information.

2. Log Records, Analysis and Management

It has the ability to record transactions carried out on all structures such as server, client and network devices that exist in the infrastructure of IT systems. These records are analyzed to determine the existing security incidents on the systems and if there is an information security breach incident, measures are taken against this situation. This study is called "Log Analysis". With log analysis, it is possible to control the external or internal attacks on the systems, as well as the work done by the existing users in the systems, such as file saving, printer output, etc. In large companies, users can easily access information such as which site they

enter, which stage they leave, which page they spend more / less time on. Logs can also be useful for performing auditing and forensic analysis, supporting the organization's internal investigations, establishing baselines and identifying operational trends and long-term problems [1]. This section discusses the collection of log records from the systems, analysis of the collected logs and log management tools.

2.1. Collecting Log Records

Log records are collected as a result of the processes shown below.

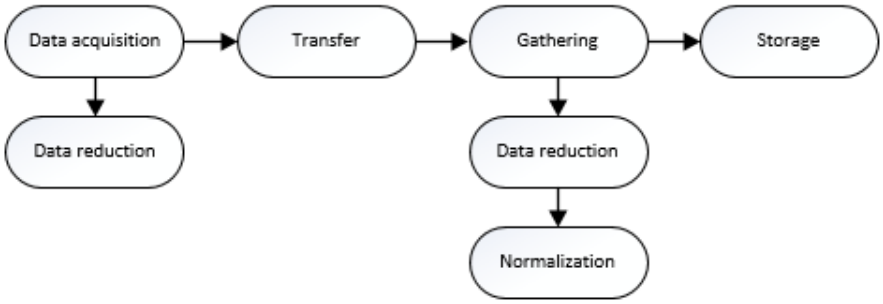

Figure 1: Log data collection life cycle

The log collection process is expressed in a formula shown below. In the equation, systems (domain) are represented by R, and all devices in this system are represented by D. Equation 2.1 sets of log records, 2.2 shows that the different log types of Information Systems Devices, 2.3 shows that each system device has a different set of event logs.

$R = \{D_1\ D_2,...., Dn\}$ (Equation 1)

Each system device has its own logs, B types of logs

$D_i = \{B_{i1}, B_{i2},....., B_{im}\}, i \in [1,n]$ (Equation 2)

If each index represents a device, each device creates different types of event records (e.g. windows systems, application, system, security logs, etc.), e event type; We can formulate log modeling.

$B_{ij} = \{e_{ij1}, e_{ij2},....., e_{ip}\}, i \in [1,n], j \in [1,m]$ (Equation 3)

Logs can be kept on the system itself or transferred to another system. The process of transferring the logs created in all systems to a single environment is called log storage. However, when the results are analyzed, all the computing incidents recorded in the form of a large number of piles have made the investigation of attempted criminal or error very complicated [7]. The management of log records is difficult for the following reasons:

- Logs produced from multiple systems in large numbers and sizes,
- Creating different types of logs from different systems,
- Log content is different from each other.

2.2. Analysis of Log File

The process of evaluation of log records will expose its employees to the difficulty of handling and managing the data in situations where a large number of event logs will take place [5, 6]. Not only the cyber attacks, but also the various other faults in the systems can be solved by log analysis. Existing log files provide a certain amount of visibility about the systems. When interpreting log files, it is important to realize the new events in different approaches to the actions that develop on systems and applications. An integrated information can be obtained through the complete analysis and comparison of the logs. If events are not considered as a whole and are not adequately analyzed, the impact and importance of these events may not be revealed. Many tools have been developed for log analysis.

Logs are collected from the systems in two different methods. The advantages and disadvantages of these two methods are stated below:

Agent Method:

• Advantages: Not data loss even if the log server is turned off. Log server can detect whether client/server is off or on.

• Disadvantages: All systems running the agent need preconfiguration, the agent can be stopped if the system is seizured.

Non-Agent Method

• Advantages: Very easy to install and configure, very flexible and scalable for very large systems.

• Disadvantages: Syslog can be a UDP-based protocol and data may be lost, Log server cannot track clients.

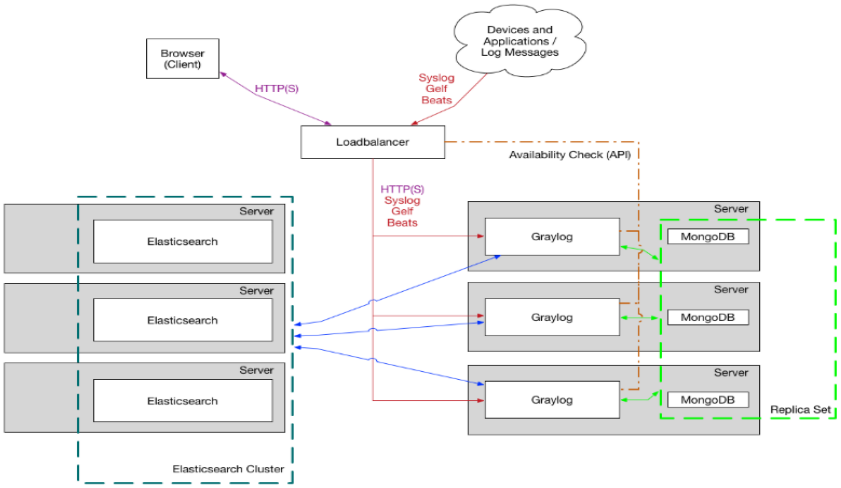

Figure 2 Log collection architecture [7]

2.3. Log Management Tool

Log analysis helps us to understand what is happening in IT processes and guide the monitoring, evaluation and solution of problems.Logging services provide protection to log data which contain valuable information about systems, networks, and applications [4]. Log management is a complicated process and organizations often make mistakes while evaluating it. Log management tools help us easily get information from log data of large sizes. These tools combine all the data and allow us to manage it with a central, accessible and easy-to-use interface. Thus, we can collect, store and manage data in one store. With log management tools, useful trends can be extracted from the current log data. The tool that is approciate for you will depend on the number of systems being monitored and your organization's compliance needs [2] .When the organization is faced with a difficult situation like cyber-attack, it is easier to use the log management tool instead of dealing with existing TXT files in the environment. With a single query, the root cause of problem of any application or software can be determined with the help of these tools. Log management can minimize the damage it will cause before a cyber-attack occurs.

Table 1 show the current advantages and disadvantages of popular log management tools.

Table 1 Advantages and disadvantages of log management tools

Log Management Tool	**Advantages**	**Disadvantages**
Splunk	Built-in alarm and reporting tool	Setup and adding new resources is not easy. Each new resource must be manually added
	Configurable charts and dashboards	High cost
	SaaS solution	
	Data transfer from a single server to multiple data centers	
	Real-time search, analysis, and visualization	
LogPacker	Provides support for more than 100 log types	Web interface not available
	Offers multiple storage providers	High cost
	e-Mail, Slack or SMS alert and reporting system	
	Easy installation via packages	
	Reliable cluster structure	
	Wide platform support: Unix, Windows, Mobile, JS	

Log Management Tool	Advantages	Disadvantages
	REST API to create custom solutions based on saved data	
	Collecting and analyzing security events	
	High performance	
	Disk-based cloud control panel	
LogRhythm	Real-time intelligent search	High cost
	Collects log from about 700 sources including applications and database	Unclear user manual and documentation
	Real-time monitoring and flexible, role-based alerts.	
	One click correlation from any search	
	The console allows users to quickly associate, search, and quickly return data.	
Logentries	Works with multiple PaaS and IaaS.	Manual installation and manual log resources management
	Real-time intelligent search	The source of errors in 3rd party libraries is not monitored.
	Custom labels for logs	Unsafe Web Client Logging
	SQL-like Query Language for Search (LEQL).	No specific reporting for JavaScript is available.
	e-Mail reports	
	Supports a variety of programming languages.	
	Understandable documentation	
Graylog	Free and open source	Support for a small number of log types
	Streams enable you to define events in real time and perform actions.	
	Easy setup	
	Server-side functionality can be extended through plug-ins	
	Rest API	
	Custom permission management for users and their roles.	
	Logs can be enriched and parsed using a comprehensive process algorithm.	
	Special dashboards for visual output of log data and queries.	
	Intuitive search interface	
Scalyr	Easy installation agent or API	Not free
	Import logs from Heroku, Amazon RDS or Amazon CloudTrail	No cloud solution
	Customizable graphics	
GoAccess	Updates the log data in milliseconds in the terminal environment.	Difficult installation
	Custom log strings	
	Follow the pages for response times; Ideal for applications	
	Effortless configuration; enough to select the log file and run it	
	Free and open source	
Fluentd	A merged log layer to parse data from multiple sources.	Restricted rights in the open source version
	Configuring structured and unstructured logs	
	Compatible with most modern data sources	
	Easy setup	
	Open source and free	
Flume	Multi-server support for data acquisition from multiple sources.	Insufficiency of add-ons
	Allows for the acquisition of large data sets from common social and e-commerce networks for real-time analysis.	
	Reliable rear end with durable storage and failover protection.	
	It can be scaled by adding more machines to transfer more events.	

Log Management Tool	Advantages	Disadvantages
	Free and open source	

In this study, Graylog was chosen as a log management tool due to the following features.
• It is an open source log management solution.
• It supports many inputs such as Syslog, GELF, TCP, UDP, AMQP. for logging process.
• Stores messages on ElasticSearch which enables fast search on archives.
• Uses MongoDB for statistical operations.
• It is possible to classify logs and perform graphical operations.
• Proactive alarms can easily be produced according to the desired conditions.

3. Graylog

Graylog is a fully integrated open source platform for collecting, indexing and analyzing structured and unstructured data from virtually any source. It has been developing since 2010. Xing has been supportive of its development.

Components of the Graylog application:

- Graylog server
- Graylog web interface
- Mango DB (Statistics and graphs)
- ElasticSearch (messages and search)

Graylog supports many input types out of the box. More inputs are available in the Graylog Marketplace. At the time of writing, Graylog supports the following;

- Syslog (TCO, UDP, AMQP, Kafka)
- GELF (TCP, UDP, AMQP, Kafka, HTTP)
- AWS (AWS Logs, FlowLogs, CloudTrail)
- Beats/Logstash
- CEF (TCP,UDP,AMQP,Kafka)
- JSON Path from HTTP API
- Netflow (UDP)
- Plain/Raw Text (TCP, UDP, AMQP, Kafka) [7]

Stores messages on ElasticSearch. It stores statistics and graphics on MongoDB. Internal message queuing system is available. Delayed or batch indexing is also supported.

3.1. Log Management with Graylog

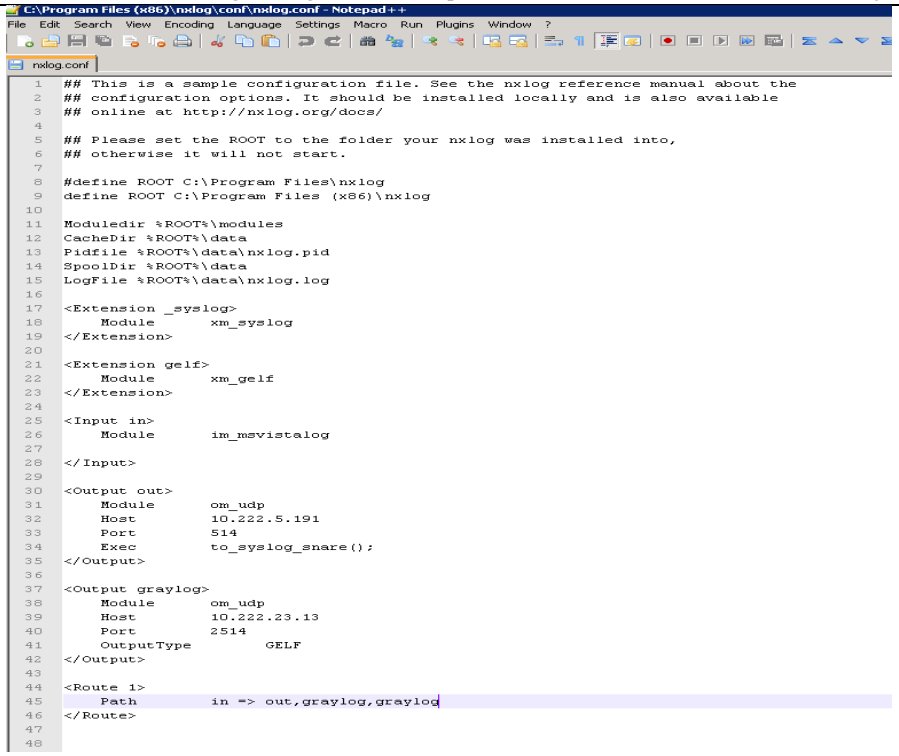

Figure 3 Nxlog Conf File Example

In the study, Graylog application was integrated with Active Directory and File Server systems in TUBITAK BILGEM Software Technologies Research Institute and the logs of these systems were collected and stored. By analyzing these logs, users who want to gain unauthorized access to the systems have been identified so that the confidentiality of the data is ensured. In addition, the integrity of the data is ensured by keeping logs of the changes made to the systems. In the study, the usage and results of Graylog are explained below.

Search, Stream, Alerts, Dashboards, Sources and System tabs are available on Graylog's web interface. To send the logs of existing systems to Graylog, Nxlog application must be installed on Windows machines and necessary definitions should be made in Conf file of this program. A sample conf file is available on the DC machine as shown in Figure 3.

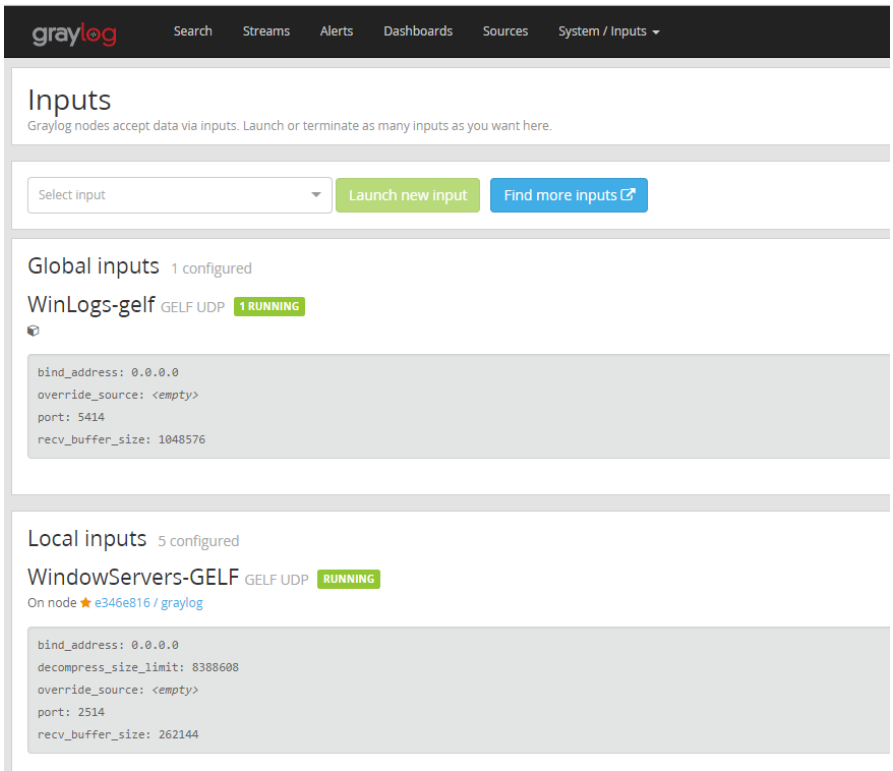

Figure 4 Graylog Input Tab

After making the necessary changes to the conf file, the server operating system's services must be restarted to enable nxlog. After the necessary operation is done on nxlog, the system will start sending log to Graylog. In order to see the sent logs on the Graylog screen, some necessary definitions must be made in the Graylog web interface. On the Input tab, as seen in Figure 4, type of Port and log files should be specified. In our application, the logs are taken with the Gelf protocol.

In Figure 5, Sources tab is shown. You can see the systems that send log files to Graylog in Sources tab.

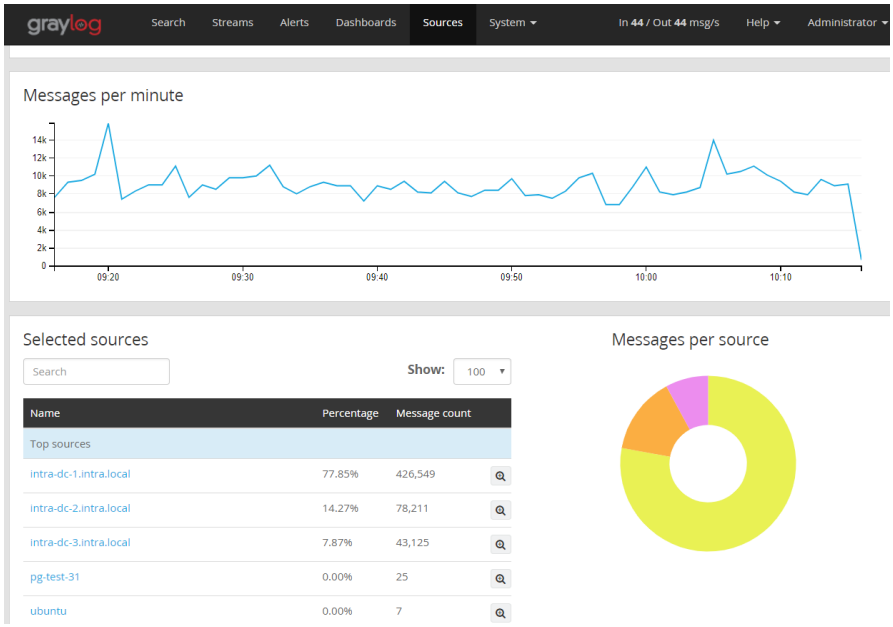

Figure 5 Graylog Sources tab

In the study, the logs of three DC machines and file servers are sent to the system. In the Graylog web interface, the logs can be filtered on the Stream tab for the desired systems. A specified event can be followed more easily by filtering the logs with an ID. In the study, the following streams were created.

- Active Directory - A computer account was created
- Active Directory - A computer account was deleted
- Active Directory - A user account was disabled.
- Active Directory - A user account was locked out.
- Active Directory - An account failed to log on.
- Active Directory - Domain Policy was changed.
- Active Directory - The domain controller failed to validate the credentials for an account File Server – Unauthorized access

In the Dashboard screen, shown in Figure 6, summaries of events defined in Stream take place.

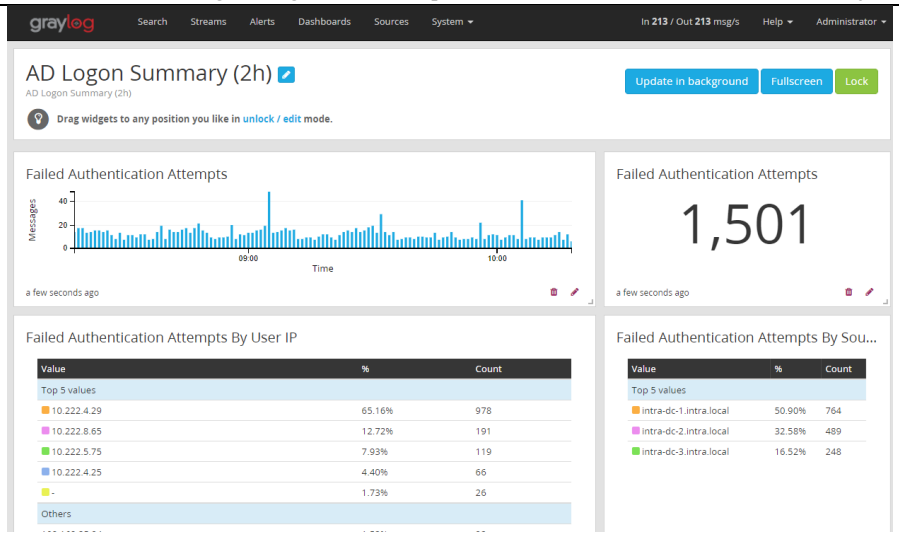

Figure 3 Graylog Dashboard Screen

Additionally, alarms can be triggered for desired logs by using created streams and desired conditions and can be sent as e-mail notifications. In tests, Graylog reported all such activities as an access to an unauthorized folder in the organization, access to a different person's account, or modification of data, to previously defined responsible persons through the generated alarms.

4. Result

The logs produced by the system are a great importance to protect the information in the system. Log records can be considered as a security camera. Information such as who has logged in and what has been done can be accessed from the security cameras, and this information is also available in log records for systems and applications. Data protection is the most important risk assets. It is very important to obtain log records of access to this data. However, keeping logs only about access does not provide information about maintaining the integrity of the data. Therefore, to ensure that the data is changed, the databases must also be kept in the logs. In other words, it is important that log analysis is performed throughout the system. Log management tools help to make the analysis conveniently. With the Graylog application we used in our study, the necessary and important logs can be obtained easily from the thousands of log files created by the systems. If the application is configured so, an alarm also can be generated to prevent an information security incident.

5. References

1. Souppaya, M. and K. Kent, (2006). Guide to computer security log management. White Paper, NIST Special Publication 800-92, Computer Security, http://permanent.access.gpo.gov/lps69969/LPS69969.pdf
2. Chuvakin, A., Schmidt, K., Phillips, C., (2013). Logging and Log Management. Retrieved from https://doi.org/10.1016/B978-1-59-749635-3.00015-4
3. SANS. Log and event management survey results (SANS eighth annual) May; 2012. [Online] https://www.sans.org/reading-room/analysts-porgram/SortingThruNoise
4. Chuvakin, A., Peterson, G. Building Security In. 1540-7993/09/ © 2009 IEEE, may/June 2009
5. Pouget. F., Dacier. M., (2003), White Paper: Alert Correlation: Review of the state of the art 1, France Institut Eurecom
6. Forte Dario V., The “ART” Of Log Correlation: Part 1: Tools And Techniques For Correlating Events And Log Files, Computer Fraud & Security, Volume 2004, Issue 6, June 2004, Pages 7-11, Science Direct
7. Anonim, https://www.graylog.org/ (Last Access Date:: 12.05.2018)
8. Graylog Documantation (2018, December 19) Retrieved from http://docs.graylog.org/en/2.5/pages/architecture.html