Internet Forensics Framework Based-on Clustering

By Imam Riadi

Internet Forensics Framework Based-on Clustering

Imam Riadi
Information Systems Study Program, Faculty of Mathematics and Natural Sciences, Ahmad Dahlan University, Yogyakarta, Indonesia

Jazi Eko Istiyanto, Ahmad Ashari, Subanar
Computer Science Postgraduate Program, Faculty of Mathematics and Natural Sciences, Gadjahmada University, Yogyakarta, Indonesia

***Abstract*—Internet network attacks are complicated and worth studying. The attacks include Denial of Service (DoS). DoS attacks that exploit vulnerabilities found in operating systems, network services and applications. Indicators of DoS attacks, is when legitimate users cannot access the system. This paper proposes a framework for Internet based forensic logs that aims to assist in the investigation process to reveal DoS attacks. The framework in this study consists of several steps, among others : logging into the text file and database as well as identifying an attack based on the packet header length. After the identification process, logs are grouped using k-means clustering algorithm into three levels of attack (dangerous, rather dangerous and not dangerous) based on port numbers and tcpflags of the package. Based on the test results the proposed framework can be grouped into three level attacks and found the attacker with a success rate of 89,02%, so, it can be concluded that the proposed framework can meet the goals set in this research.**

***Keywords*—framework; forensics; Internet; log; clustering; Denial of Service**

I. Introduction

Background of this research starts from many attacks in the Internet. The attacks such as SYN Flood, IP Spoofing, DoS attacks (Denial of Service), UDP Flood attack, Ping Flood attack, Teardrop attacks, Land Attack, Smurf Attack, Fraggle Attack [1]. DoS attack is a type of computer network attacks that causes the operating system in that server running out of resources. This resulted in the server not being able to serve legitimate user demand and cause the network to be down. Based-on the attacks that often occurs in the Internet network, it is necessary to study forensics to help classifying the log so that attacker information can be immediately known.

Digital forensics is the science dealing with the process of recovery and investigation of material found in digital data, this is often done as part of a criminal investigation [2], [3], [4], in which the scope of digital data comprises a computer system, storage media, electronic documents, or even a sequence of data packets transmitted across computer networks. Network forensics is a part of digital forensics that monitor and analyzes data traffic on the network. Data that are handled in network forensic are dynamic. It is different from that of is digital forensics, where data is static [5].

Research on forensics is related to the data found in network traffic. Network forensics analyzes data traffic through a firewall or intruder detection system in network devices such as routers. The goal is to conduct the traceback to the source of the attack so that the identity of the attacker can be determined [6]. Besides, network forensics has a goal to collect, identify and analyze documents of some processing and transmitting digital data. This activity aims to obtain information or facts related to the attacker [7].

Today's network forensic process particularly by using Internet has increased rapidly. To help facilitate the Internet network forensic process it is needed an alternative solution in the form of a framework to facilitate the discovery of information about the attacker. Framework developed in this study includes overall stages that occur in the forensic process. The whole of forensic process starts from the input in the form of logs obtained from the capture and recording processes in the network traffic. Once the log information is obtained and stored in the database, the log processed using a clustering technique is able to generate the information about attacker needed by the user, in this cases is the investigator. Analysis and log management process requires an additional application that can help implement the framework. Furthermore, that additional application based web is reffered to NFAT (Network Forensic Analysis Tools).

The difference between this study and [8] is [8] focused on framework development, while this paper uses its framework to identify Denial of Service attacks that using NFAT machine. NFAT application is integrated within the framework proposed in this study. To help finding needed information about the attacker, NFAT application needs complex analysis process. To reduce the complexity of data processing, this study utilizes clustering techniques. Clustering technique is one of methods that can be used to facilitate identifying network attacks [9]. Clustering will divide the data into several clusters in which the data in one cluster have similar characteristics and essential equality. The reason of using clustering technique selection in this study is the data characteristic about information of attacker particularly hit access in the numerical form and log information in the network is very large. In addition, to facilitate the process of grouping the log information, it is necessary to facilitate knowing information about attacker in the Internet network. Clustering techniques can be implemented using k-means clustering algorithm.

K-means clustering algorithm is one of the most popular and widely used techniques in the industrial world. K-means clustering method classifies objects in a cluster, the cluster membership value of each cluster centroid is calculated by finding the distance between data and centroid. If the data has the shortest distance from the centroid of a cluster then the data will be the members of the cluster [10]. Web-based applications to detect attacks in the Internet is called NFAT machine (Network Forensic Analysis Tools) that will be used in this study as a proof of concept implementation of a framework for Internet forensic proposed.

II. Current Studies On Network Forensic

Recent research related to this study is divided into two parts, namely the study of forensics in the existing network security and research on clustering techniques often used in data processing.

A. Forensics in Network Security

Several previous studies have been done on digital forensics. In general, the purpose of digital forensic analysis is to identify digital evidence to assist in the investigation. In the mid 1990s the agency guidelines for best practice in the forensic examination of digital technology IOCE (International Organization on Computer Evidence) was established to build the standardization development of digital forensics. The main purpose of IOCE is to combine methods and practices to ensure the ability in using digital evidence. In addition, [11] also developed a scientific working group related to digital evidence for the purpose of forensic guidelines in dealing with digital evidence.

Furthermore [12] presents a technique used in digital forensics to show the methods and tools used for digital forensics. In contrast to digital forensics, few studies done on network forensics has identified several important aspects that are used in network forensics, among others [5] state that network forensic is a part of digital forensic that recently has grown as a very important discipline that used to monitor, especially for the purposes of tracking disorders and attacks. This study suggests it is unlikely that a single tool will be enough for the investigation but it has to use a combination of several tools. While [13] describes a variety of techniques and measures in network security that have been developed to assist the process of digital forensic investigations. It also discusses the network security issues and vulnerabilities that have been exploited by hackers and network intruders. Network preventive measures have been identified through the use of various types of firewall and network architecture. In addition, [14] states the network architecture can have implications on network forensics while the network architecture design is perfect for improving the quality of information produced. According to a statement [15] threats to digital assets has increased so that it is necessary to eliminate the risk from various threats. Attackers have been using anti forensic techniques to hide evidence of Internet crime. Internet forensics equipment should increase the resilience in warding off an ongoing threat. In addition, [16] states that the network forensics solutions based intruder detection analysis need to be followed up in order to record the behavior and analyze the data network intruders in detail . The processing of this data is expected to ensure data integrity and authenticity of data, so the results of data analysis have a good level of credibility in the forensic system.

Forensic process is an activity that combines several disciplines. In contrast to the opinion [17] states that the network forensic analysis is not only a study of science but also need the art to do so. Forensic refers to the use of evidence after the attack to determine how the attack was carried out and what the attacker did. Data traffic on the network is very complicated to be studied. Role of network forensics is to detect abnormal traffic and identify intruders [18]. In addition, there are a few things to watch out where the problem also concerns law enforcement, some of the activities are the process of capturing and analyzing network traffic to get the keywords and information about attacker.

Some tools and techniques analysis used in forensic analysis of network can be seen in table 1 [5].

TABLE I. Some Tools Used To Support Network Forensics.

Tool	Web Site	Attributes
TCPDump Windump	www.tcpdump.org	F
Ngrep	http://ngrep.sourceforge.net	F
Wireshark	www.wireshark.org	F
Driftnet	www.backtrack-linux.org/backtrack-5-release [Release 3, August 2012]	F
NetworkMiner	www.netresec.com/?page=NetworkMiner	F
Airmon-ng, Airodump-ng, Aireplay-ng, Aircrack-ng	www.backtrack-linux.org/backtrack-5-release [Release 3, August 2012]	F L R C
Kismet	www.kismetwireless.net	F
NetStumbler	www.netstumbler.com	F
Xplico	http://packetstormsecurity.org/files/tags/forensics	F
DeepNines	www.deepnines.com	F
Sleuth Kit	www.sleuthkit.org	F R C
Argus	www.qosient.com/argus	F L
Fenris	http://lcamtuf.coredump.cx/fenris/whatis.shtml	F
Flow-Tools	www.splintered.net/sw/flow-tools	F L
EtherApe	http://etherape.sourceforge.net	F
Honeyd	www.citi.umich.edu/u/provos/honeyd	F
SNORT	www.snort.org	F
Omnipeek /Etherpeek	www.wildpackets.com	F L R
Savant	www.intrusion.com	F R
Forensic Log Analysis GUI	http://sourceforge.net/projects/pyflag	L
Analysis Console for Intrusion Detection	www.andrew.cmu.edu/user/rdanyliw/snort/snorta cid.html	L
Dragon IDS	www.enterasys.com	F R L C
Infinistream	www.netscout.com	F R C
RSA EnVision	www.emc.com/security/rsa-envision.htm	F L R C A
NetDetector	www.niksun.com	F R C A
NetIntercept	www.niksun.com/sandstorm.php	F R C A
NetWitness	www.netwitness.com [www.rsa.com]	F L R C A

Information : F : filter and collect;
L : log analysis;
R : reassembly of data stream;
C : correlation of data;
A : application-layer view.

Some software such as shown in table 1 requires follow-up to the next process could help investigate digital crimes. The investigation of digital crime is indispensable to help the investigation process.

Several attempts to detect such attacks have been carried out using the existing anomaly detection techniques in network traffic using statistical techniques (statistical anomaly detection). This detection technique involves a collection of data related to a legitimate user behavior during a certain time. Currently there are several methods for detecting DoS attacks.

The method is divided into three categories, namely the detection and defense based on the analysis of the protocol characteristics [19], the accumulation [20] and the statistical models on network traffic. Several methods of detection and prevention also have many obstacles, among others : an analysis of the detection and prevention based on the characteristics of the protocol can only be applied to the type of attack that the characteristics of the protocol abnormally occurs [21]. Many types of attacks that do not fit the protocol as well as the accumulation of network traffic statistical models cannot distinguish between normal traffic and large scale [22].

Based on the results of previous studies, carrying out a variety of mechanisms to detect Denial of Service (DoS) has advantages and disadvantages. However, techniques to detect DoS attacks are still complex.

B. *Clustering techniques using k-means clustering algorithms*

Fundamental issue on software development that supports forensic network is how to determine the appropriate method to facilitate the processing of log data into easily processed data to uncover digital crimes especially those using the Internet as a medium to conduct attacks.

Cluster analysis is the process of analyzing and interpreting a set of data based on similarity. It means that the data is grouped into one cluster due to the same pattern [23]. Clustering includes the type of learning that is unsupervised. Supervised learning and unsupervised learning have a different way of working which is very significant. To have some kinds of unsupervised learning algorithms, [10] has studied the comparison against some types of clustering algorithms. In doing this comparison [10] used several parameters such as the popularity, versatility and easily applied to the data in bulk. There are four kinds of clustering algorithms compared to performance, such as: k-means, hiearchical clustering, self organization map (SOM) and the expectation maximization (EM Clustering). Based on the test results can be concluded that the performance of the k-means algorithm and the EM is better than the hiearchical clustering algorithm. In general, partitioning algorithms such as k-means and EM are highly recommended to be used in the large size of data. It is different from a hierarchical clustering algorithms that have good performance when they are used on small data size.

Furthermore [24] describes an intruder detection system that has been developed to achieve high efficiency and improve accuracy of detection and classification. The proposed system consists of two stages. The first stage is to detect the attack and the second stage is for classification of attacks. Data mining techniques can be used to improve the detection rate and reduce the false alarm rate. In addition [25] states that the k-means algorithm is needed to determine the final number of clusters (k) before. Based on the results obtained by k-means algorithm turns out better than the FCM algorithm. FCM produces result which is close to the k-means clustering, but it still takes longer than computing k-means. K-means algorithm appears to be superior compared with the Fuzzy C-Means algorithm (FCM).

III. Network Forensics

Network forensics is an attempt to find the attacker information to look for potential evidence after an attack or incident. These attacks include probing, DoS, user to root (U2R) and remote to local. Network forensics is the process of capturing, annotating and analyzing network activity in order to find digital evidence of an attack or crime committed using a computer network so that offenders can be prosecuted under applicable laws as illustrated in Fig.1 [26]. Digital evidence can be identified from the recognized attack patterns, deviations from normal behavior or deviations from the network security policy that is applied to the network. Network forensics has a variety of activities and techniques of analysis, such as: analysis of existing processes in the IDS , the analysis of network traffic [11] and the analysis of network device itself [27], all considered the part of a network forensics.

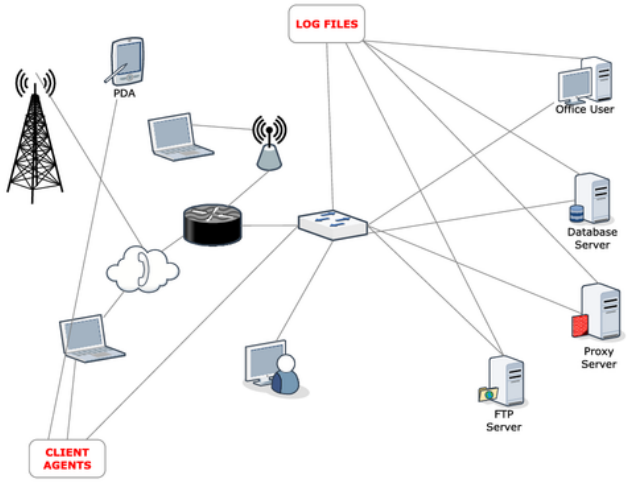

Fig. 1. Overview of network forensics process

Digital evidence can be gathered from various sources depending on the needs and changes in the investigation. Digital evidence can be collected at the server level, the level of proxy or some other sources. For example, at the server level digital evidence can be gathered from web server logs that store browsing activity behavior frequented. The log describes the user that accesses the website and what it does. Several sources includes the contents of the device and the network traffic through both wired and wireless networks. For example, digital evidence can be gathered from the data extracted by the packet sniffer such as tcpdump [28] to monitor incoming traffic in the network. Currently, the number of criminal evidence in the computer continues to increase, even most of the evidence is still used to represent the traditional or conventional crime.

A. *Level Attacks in Computer Networks*

Computer security often focuses on preventing attacks using authentication, filtering and encryption techniques. Nevertheles another important aspect is the act of detecting attacks after the violation occurs in a network attack.

There are two general approaches to determine whether there is an attack in a network such as digital signature detection, where the pattern of attacks signal will be searched as well as anomaly detection, where abnormal deviations in the network will be detected to determine whether there is an attack or not. Deviations will be divided into several attack types. Table 2 shows the grouping of several types of attacks based on the level of attacks [29].

TABLE II. LEVEL OF ATTACKS IN COMPUTER NETWORKS

No	Level of Attacks	Port / Protocol	TCP Flags	Information
1	Dangerous	80 / TCP	16,32	HTTP
		8080 / TCP	16,32	HTTP alternate
		443 / TCP	16,32	HTTPS (Hypertext Transfer Protocol over SSL/TLS)
		20 / TCP	16,32	FTP data transfer
		21 / TCP	16,32	FTP control (command)
		22 / TCP	16,32	SSH
		23 / TCP	16,32	Telnet protocol
		53 / UDP	-	DNS
2	Rather Dangerous	161 / TCP	20 - 24	SNMP
		143 / TCP	20 - 24	IMAP
		162 / TCP	20 - 24	SNMPTRAP
		110 / TCP	20 - 24	POP3
		993 / TCP	20 - 24	IMAPS
		137 / UDP	-	NetBIOS
		161 / UDP	-	SNMP
3	Not Dangerous	In addition to the above mentioned	TCP (20-27) UDP (-)	In addition to the above mentioned

The attack happens in the computer refers to the protocols and ports used. Based on the protocol and the port level, attacks will be grouped into three levels consisting of malicious port 80, 8080, 443, 20, 21, 22, 23, 53, and somewhat dangerous level consisting of ports 161, 143, 162, 110, 993, 137, 161 and harmless level consisting of a port in addition to those at the previous level.

B. Attack Detection Using IP Header

Threat of attacks in computer networks has grown rapidly. Monitoring and analysis of packet data traffic in the network is done by examining all packet headers and threats to each package. Normal pack behavior was analyzed according to each protocol and header. Each protocol has a header and function in accordance with the protocol TCP/IP [30].

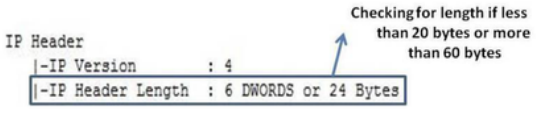

Fig. 2. Field packets that have been observed and analyzed

Fig. 2 shows that IP hearder lenght in the IPV4 must be equal or above 20 bytes and equal to or below 60 bytes. If the IP header length is less than 20 bytes or 60 bytes above, it can be suspected that there is an attack contained in a computer network.

This study has a case study about the types of attacks that often lead to network services disrupted. The disruption is caused by attackers who launched an action by sending and flooding data packets in the Internet network. All data traffic on the network can be saved into the log. This log is very important to help identifying the cause of the attack and can be used as evidence for the investigation. Logs can be obtained in one way to capture data traffic on an interface that is connected to a peripheral or router. Capture traffic activity on the network can be done using several tools, one of which can be used is tcpdump. Tcpdump application has the ability to capture data traffic on a particular interface that is specified. Results of the tcpdump application are then saved as a log that later can be used to reconstruct a digital crime using the Internet.

In this research, a case study used to process forensic is DoS attacks (Denial of Service). The DoS attacks include malicious attacks and often causes the system or network slow even down. The attacks examined in this study is a DoS attack that attacks port 80 (http), port 443 (https), port 21 (ftp) and port 22 (ssh). Reason for selecting the port for the case study because the service mentioned above is a public service which is often used by users to utilize the Internet network.

IV. FRAMEWORK FOR INTERNET FORENSICS

This section discusses the identification of critical needs in Internet forensics based on digital crime case studies as described above. NFAT engine development (Network Forensic Analysis Tools) proposed in this study requires a supporting infrastructure consisting of multiple hardware requirements (hardware) and some software (software) supports.

A. Proposed Framework for Internet Forensics

This section discusses a framework proposed in the study. This framework was developed based on the identification of the requirements needed in Internet forensics. Stages identified in Internet forensics are shown in Fig. 3 [8].

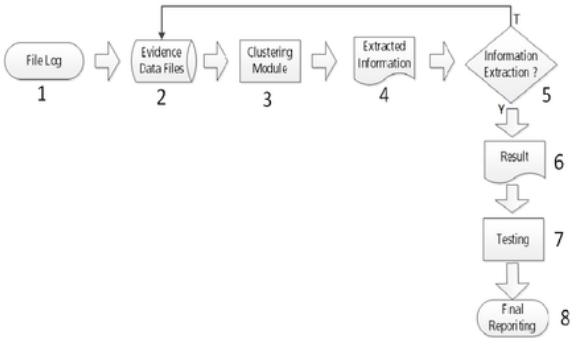

Fig. 3. Proposed framework for Internet forensics

Complete systematic of Internet forensics framework proposed in this research includes several stages. There are 8 stages of the process that must be performed sequentially. Details of each stage are shown in Table 3.

TABLE III. DETAILS OF INTERNET NETWORK FORENSICS PROCESS STAGES

No	Process	Information
1	File *Log*	This log is generated from tcpdump output, at this stage in realtime tcpdump application will capture all the data packets passing through the network interface specified.
2	Evidence data files	Evidence in question here is the original log, the output of the tcpdump application that is stored in a text file.
3	Clustering Module	NFAT module developed in this research is the application modules that can classify the level of these types of attacks into 3 groups (dangerous, rather dangerous and not dangerous). The concept is applied in this module uses clustering techniques using k-means algorithm.
4	Extracted Information	At this stage the log is already saved in the database to extract the data in accordance with the purpose of investigation. Log in information is stored in a database NFAT tools that apply the concept of partitioning MySQL database using horizontal partitioning techniques.
5	Confirmation Extract Information	At this stage, the investigator will conduct relevant confirmation log that contains the IP address is generated by the clustering module.
6	Result	At this stage, detailed information will be obtained IP addresses that have been identified through clustering module. With the help of application services that can be accessed via the URL: http://www.robtex.com the detail information of IP addresses that have been found will be clearer information relevant ASN (autonomous System Number) is used.
7	Testing	At this stage, the test will be held as a proof of concept dati form the framework for the proposed Internet forensics, testing is done using the software LOIC and DoSHTTP which will perform the simulation engine to NFAT and test results will be verified by the original logs in the form of a text file.
8	Final Reporting	At this stage the information is already known to the attacker can be molded according to the needs of the investigator.

B. Architectural Design Network

This section discusses the design of network architecture used to implement the NFAT machine. The approach used in the implementation of network topology uses hierarchy concept. In a hierarchical network concept the network is divided into several parts according to the functions and services provided at each proficiency level layer. Design of network architecture used in this study is shown in Fig. 4.

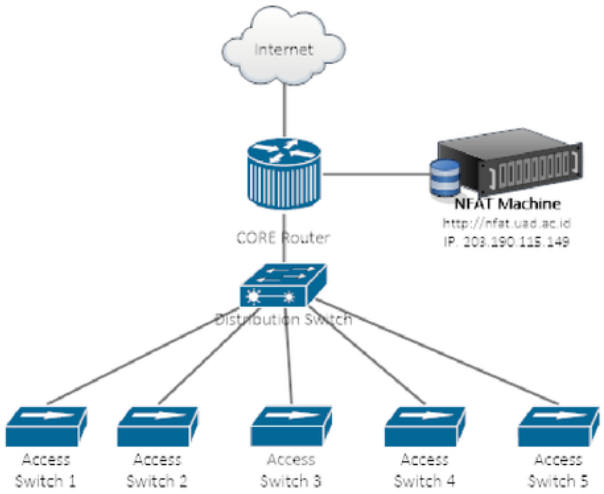

Fig. 4. NFAT machine network architecture design.

C. Implementation of Data Traffic Arrest

The purpose of this Internet forensics is to help finding information about attacker for digital crime commited in the Internet network. It requires careful analysis process so that the goal of the forensic process can be achieved. The process of arrest data traffic in computer network is done using tcpdump assistive software. The function of this application is capturing data traffic in real time and saving it as a log. That tcpdump application arrests all data traffic passing through the interface eth0 and displays it in a time format that is suitable with timestamp in the form of IP Address and display information and protocol header of the data packet. then, The results of the tcpdump application is then stored in the form of a text file with the name of NFAT-eth0.log. Log in the form of a text file then stored and used as the original log verification if it is required by the investigator. In addition, log tcpdump output result is also stored into the database. Logs derived from the process of arrest data traffic is parsed using regex so that logs can be saved to the database by fi16 (timestamp, source mac address, mac destination address, source address, source port, destination address, destination port, and protocol length).

D. Implementation of Data Grouping

The process after storing log file in the databases is developing NFAT machine and grouping the data to find information about the attacker using Internet network. This study uses clustering techniques to group logs that have been stored in the database.

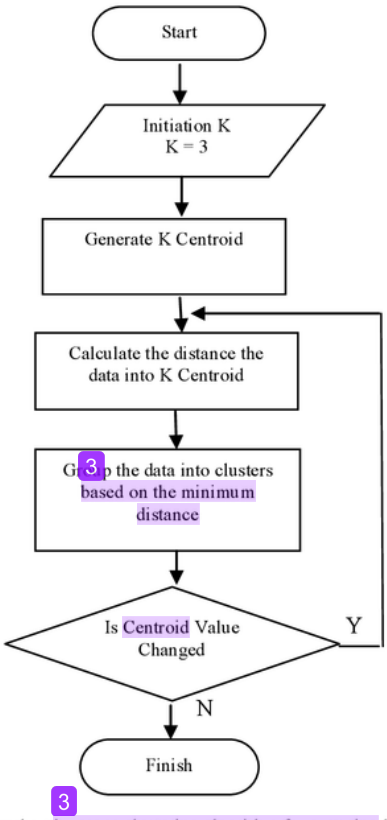

Fig. 5. Flowchart k-means clustering algorithm for grouping log.

NFAT engine that was developed in this study uses the k-means clustering algorithm. Clustering process is used in order to help finding information by classifying the attacker logs into three groups attack levels, namely: dangerous attack, rather dangerous attacks, and not dangerous attacks. Detail steps of process that occur in the process of grouping data using k-means clustering algorithm can be seen in the flowchart in Fig. 5.

E. Database Implementation

Result of the data traffic capture process on the network then is stored in NFAT machine database. Database server used to store logs has been previously filtered using MySQL. NFAT machine has 4 tables that serve to capture and process the results of data traffic logs to facilitate the process of storage and retrieval of information about the attacker as shown in Fig. 6.

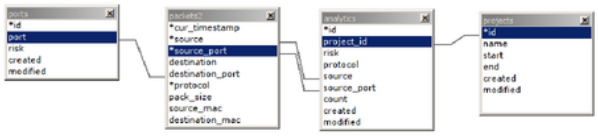

Fig. 6. NFAT engine database schema diagram

V. The Result Analysis NFAT Machine

The results are divided into two stages, The first stage, NFAT engine captures existing packet of data traffic in the network that next will be carried out to devide the clustering process into three levels of attack. The next stage, NFAT machine optimizes the search process and storage of logs into the database. Follows are details of the stages done by NFAT machine.

The first stage of the forensic process starts from collecting information related to the user reports to the investigators then followed by managing the information sought by the data and time attack events. In the analysis phase, the results of the data traffic on the network will be saved in the original logs in the form of a text file and also stored in the database. Incidents of attacks are captured and stored by the NFAT (Network Forensic Analysis Tools) machine. Information needed by investigators will be extracted from the clustering module, where the profile creation process and the analysis time are used as part of the incident investigation process. The resulting interim results clustering module will be verified by the investigator. If there is a verification process need to be clarified about the IP address information that has been generated by the clustering module, investigators can then re-check NFAT into the engine to make sure that the IP address is an IP Address of the suspected assailants who had attacked the system through the Internet network. It can be assisted and linked to the previous stage to repair information, whether the information was sufficient or not. In the final stages of reporting, information related to the attacker who has been found can be used to help uncover digital crimes committed using the Internet network.

Clustering module works using k-means algorithm, where the module can perform an attack level grouping into three groups:

1) dangerous attack,
2) rather dangerous attack,
3) not dangerous attack.

Based on the data stored in the database log, the clustering process will be carried out through the following steps.

1) Specify value of k as the number of clusters to be formed.
2) Generate initial k centroids randomly.
3) Calculate the distance of each data to each centroid.
4) Data will flock around the nearest centroid.
5) Determine the new centroid positions by calculating the average values of the data from the same centroid.
6) Go to step 3 if the new centroid position are not the same.

Results for data clustering, because of its random nature, is highly dependent on centroid generation, this is cause the result of attack detection on the data is always changing. After the attack data clustering process is done, then every cluster results do cluster labeling are included in the dangerous, rather dangerous or not dangerous level of attacks. After cluster labeling, the data entered are checked for the next process of grouping noted in the report. The process of clustering using k-means algorithm is shown in Fig. 7.

N

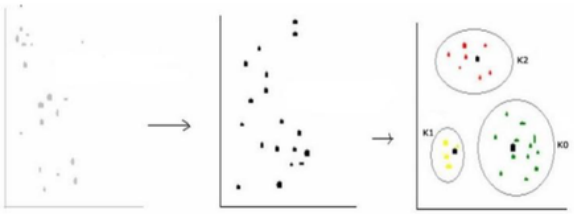

Fig. 7. The process of data clustering with k-means attack

Based on Fig. 7 clusters formed above are the best clusters obtained from the clusters that has the smallest value variants. From the above formed clusters, each cluster for the data is already formed but not yet labeled. At labelling process from the largest to the smallest variants, the result show that clusters K0, K1 clusters and clusters K2, K0 are not dangerous attack, cluster K1 is rather dangerous attack and cluster K2 is dangerous attack.

Logs generated by NFAT machine consist of several items, including IP address, port and hit. Example of a successful log data stored by NFAT machines is shown in Table 4.

TABLE IV. LOG DATA IS STORED BY NFAT MACHINES

No	IP Address	Port	Hit
1	103.19.183.67	80	90
2	202.67.40.24	80	50
3	192.168.100.15	80	30
4	202.67.40.25	80	70
5	203.190.115.149	80	80
6	202.67.40.11	80	40
7	103.19.180.2	80	20
8	203.190.112.231	80	30
9	202.67.40.5	80	50
10	198.24.130.167	80	60

The next stage is the clustering process which is based on the log data in Table 3. where the clustering parameter used is the number of IP Address that are grouped into three categories based on the number of hits. Those are : dangerous attacks, rather dangerous attacks and not dangerous attack. The algorithm used to perform the data clustering is k-means clustering using euclidean distance concept. Based on the results, the calculation is shown in Table 5.

TABLE V. DETAILS OF FINAL RESULTS FOR THE TENTH STAGE OF DATA CLUSTERING.

No	IP Address	Hit	Cluster	Category
1	103.19.183.67	90	1	dangerous attack
2	203.190.115.149	80	1	dangerous attack
3	202.67.40.25	70	1	dangerous attack
4	198.24.130.167	60	1	dangerous attack
5	202.67.40.24	50	2	rather dangerous attack
6	202.67.40.5	50	2	rather dangerous attack
7	202.67.40.11	40	2	rather dangerous attack
8	192.168.100.15	30	3	not dangerous attack
9	203.190.112.231	30	3	not dangerous attack
10	103.19.180.2	20	3	not dangerous attack

Implementation of the clustering process in NFAT engine was developed using the PHP programming language using CakePHP framework presented in the form of display. That kind of level of attack can be viewed in more detail by the port that will be analyzed. In addition, the clustering process is also displayed in graphical information by categorizing type level of attacks where dangerous attack is colored red, rather dangerous attack are colored yellow and not dangerous attack is colored green as presented in Fig. 8.

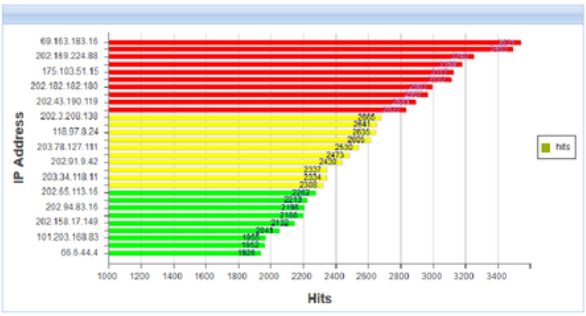

Fig. 8. The result of the clustering process step NFAT machine.

B. Scenario Testing

This section discusses what needs to be prepared prior to testing. Understanding the needs and design testing is important that the testing phase goes well according to plan. This study develops a framework forensics through the development of Internet and tests the system and field experiments in the form of test scenarios. Scenario testing will be done with the topology shown in Fig. 9.

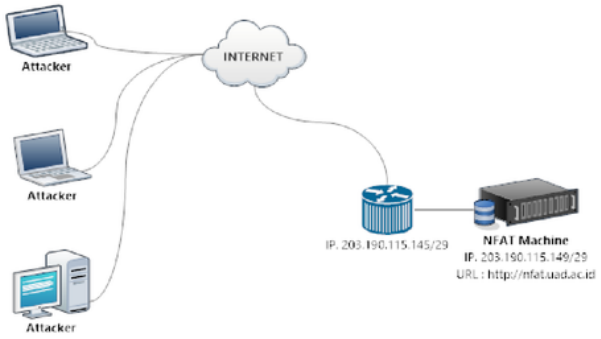

Fig. 9. NFAT machine design test scenarios

Fig. 9 explains that the machine has NFAT domain URL with an IP Address 203 190 115 149 nfat.uad.ac.id connected to the Internet via a router. In addition, there are some attackers who carried out the attack through the Internet network. This attack uses testing tools DoSHTTP. DoSHTTP software downloaded from the site (http://www.socketsoft.net/). It is a tool used to simulate an attack into NFAT machine. Tool DoSHTTP is a testing tool for HTTP flood DoS requests which sends packets to a NFAT machine. Ports that can be simulated using the tools DoSHTTP is port 80 (http) only. The testing process is done by inserting attacks IP address or domain of the NFAT machine nfat.uad.ac.id 203 190 115 149 or domains as shown in Fig. 10.

Fig. 10. Software testing machine DoSHTTP for NFAT

C. DoS Attack Scenario Port 80 (http) Using DoSHTTP

Attacker use this software to simulate an attack on a target machine that has an IP Address 203 190 115 149. IP addresses are included in the target URL or can be replaced with the domain name of the target machine (nfat.uad.ac.id) to be attacked. User Agent on DoSHTTP applications to select the type of browser will be used to simulate the attack. Sockets on DoSHTTP application show the magnitude of the package to be delivered to the target machine. Start Flood used to launch an attack on the target machine.

Results of attack scenarios using software DoSHTTP is then recorded especially in the request packet that will be checked with result of log that is stored in the text file form in the database. Based on the results of testing scenarios conducted a number of attacks ten times using a different IP address, obtained the results as shown in Table 6.

TABLE VI. RESULTS OF TESTING SCENARIOS USING PORT 80 ATTACKS DOSHTTP

No	IP Address Attacker	Port	Request Issued (Hit)	Request Received (Hit)	Detect (%)
1	103.19.183.67	80	100489	89434	88,99
2	202.67.40.24	80	100213	89217	89,03
3	192.168.100.15	80	100119	89118	89,01
4	202.67.40.25	80	100247	89232	89,01
5	203.190.115.149	80	100320	89290	89,01
6	202.67.40.11	80	100393	89360	89,01
7	103.19.180.2	80	100474	89469	89,05
8	203.190.112.231	80	100297	89286	89,02
9	202.67.40.5	80	100317	89304	89,02
10	198.24.130.167	80	100424	89415	89,04

Based on the test results carried out attacks on port 80 using the software DoSHTTP obtained the results that the attack scenario can be processed by NFAT machine with an average success rate of 89,02%.

Results of testing of port 80 scenario attacks using DoSHTTP software can also be presented in graphical form as shown in Fig. 11.

Fig. 11. Graph the results of attack scenario using port 80 DoSHTTP

VI. CONCLUSION

Framework for forensic Internet generated in this study allows users, in this case inverstigators to know the attacks level-related to the attacker's information and resources that is going on in the Internet. This Framework was developed using two stages, namely the clustering process stages and phases of database storage and search logs improved performance. Clustering techniques used in this research was able to classify the level of attacks and shows the attacker information occurs in a network the Internet. Clustering algorithms used machine NFAT (Network Forensic Analysis Tools) using the k-means algorithm. Results of traffic data that is captured in the network is stored in a database for later will be processed using the k-means algorithm to classify the level of attacks into three categories, namely a dangerous attack, rather dangerous attack, and not dangerous attack. The result of testing NFAT machine demonstrate and inform attacks level as well as the information about the attacker that happen in the Internet network with 89,02% success rate to ease the verification process of the source of the attack.

NFAT machine can eventually expanded to be able to detect all the protocols that are commonly used in communications networks, especially the Internet. Additionally, NFAT machine can be installed in some portable devices or embedded so that it has smaller size dimensions, light weight and efficient in the use of power.

ACKNOWLEDGMENT

The authors would like to thank Ahmad Dahlan University (http://www.uad.ac.id) to the research funding.

REFERENCES

[1] Jingna.L, An Analysis on DoS Attack and Defense Technology, The 7th International Conference on Computer Science & Education (ICCSE) July 14-17, Melbourne, Australia, 2012.

[2] Anstee.D, *Worldwide Infrastructure Security Report*, vol. 7," Arbor Networks, Feb. 2012, www.arbornetworks.com/report

[3] NIST-a, Information Testing Laboratory, *Computer Forensics Tool Testing Program*, 2012, www.cftt.nist.gov

[4] NIST-b, *Guide to Integrating Forensic Techniques into Incident Response*, 2012, http://csrc.nist.gov/publications/nist-pubs/800-86/SP800-86.pdf

[5] Hunt R. *New Developments In Network Forensics*–Tools and Techniques, Proceedings of the IEEE, 2012, pp. 376-381.

[6] Pilli, *A Generic Framework for Network Forensics*, International Journal of Computer Applications (0975-8887), 2012, vol. 1, pp. 1-6

[7] Palmer, G, A Road Map for Digital Forensic Research, 1st Digital Forensic Research Workshop, New York, 2001, pp.15-30.

[8] Riadi.I, Istiyanto.J.E, Ashari.A, Subanar, Log Analysis Techniques using Clustering in Network Forensics, *International Journal of Computer Science and Information Security (IJCSIS)*, 2012, vol. 10, pp. 23-30.

[9] Liao.SH., Chu.P.H., Hsiao.PY., *Data Mining Techniques and Application - A Decade review from 2000 to 2011*. Expert Systems with Application, 2012, pp. 11303-11311.

[10] Abbas.O.A, *Comparisons Between Data Clustering Algorithms*, The International Arab Journal of Information Technology, 2008, vol 5, 320-325.

[11] Casey, E. *Handbook of computer crime investigation: forensic tools and technology*. 2004, Academic Pre

[12] Mabuto. E.K H. S Venter2, *State of the art of Digital Forensic Techniques*, roceeding of Information Security South Africa Conference, Department of Computer Science, University of Pretoria, Preto 011.

[13] Aichi.H, A.Hellany & M.Nagrial, Network Security Approach for Digital Forensic Analysi 08.

[14] Strauss. T, Martin S. Olivier, *Network Forensics in a Clean-Slate Internet Architecture*, Proceedings of the IEEE, 2011.

[15] Sridhar N, Dr.D.Lalitha Bhaskari, Dr.P.S.Avadhani, *Plethora of Cyber Forensics*, (IJACSA) International Journal of Advanced Computer Science and Applic ns, 2011, vol. 2, pp. 110-114

[16] Liu. J, Guiyan.T, *Design and Implementation of Network Forensic System Based on Intrusion Detection Analysis*, International Conference on Control Engineering and Communication Technology, 2012.

[17] Raftopoulos E, Matthias E, and Xenofontas D, *Shedding Light on Log Correlation in Network Forensics Analysis,* 2012.

[18] Chennaka. A, *Network Forensics : A Survey*, Electrical and Computer Engineering, Iowa State University, 2013.

[19] Chen J C, Jiang M C Liu, *Wireless LAN security and IEEE 802.11i*, IEEE Wireless Communications 005, pp. 27-36

[20] Xing. X.Y, Shakshukie B., *Security analysis and authentication improvement for IEEE 802.11i specification*, Proc of IEEE GLOBECOM, 2008, pp.

[21] Sheng. Y. Tank Chen G, *Detecting 802.11 MAC layer spoofing using received signal strength* C Proc of IEEE, 2008, pp. 1768-1776.

[22] Bagus A., Ali S, Ardelia H., *The design of a mazesolving system for a micromouse by using a potential value algorithm*, Journal World Transactions on Engineering and Technology Education, 2006, pp. 509-512

[23] Han J. and Kamber M., *Data Mining : Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.

[24] Kumaravel. A, *Multi-Classification Approach for Detecting Network Attacks*, Proceedings IEEE Conference on Information and Communication Technologies, 2013.

[25] Ghosh.S., Sanjay Kumar Dubey, *Comparative Analysis of K-Means and Fuzzy C-Means Algorithms*, (IJACSA) International Journal of Advanced Computer Science and Applications, 2013, vol. 4, pp. 35-39

[26] Mukkamala, S. and Sung, A.H. *Identifying significant features for network forensic analysis using artificial techniques*. International Journal of Digital Evidence, 2003, vol. 1, pp. 1-17

[27] Petersen, J.P. *Forensic examination of log files*. MSc thesis, Informatics and Mathematical Modelling, Technical University of Denmark, mark, 2005.

[28] Jacobson, Van, Craig Leres, and Steven McCanne, tcpdump - dump traffic on a network, UNIX man pages, 1998.

[29] Fauziah L, *Computer Network Attack Detection Based on Snort IDS with K-means Clustering Algorithm*, ITS L ry, 2009.

[30] Haris,S.H.C, Shadoon, M.G, Ahmag, Ghani, *Anomaly Detection of IP Header Threats,* International Journal of Computer Science and Security (IJCSS), 2011, vol. 4, pp. 497-504.

Internet Forensics Framework Based-on Clustering

ORIGINALITY REPORT

5%

SIMILARITY INDEX

PRIMARY SOURCES

1	Jianfeng Ma. "Security Architecture Framework", Security Access in Wireless Local Area Networks, 2009 Crossref	18 words — < 1%
2	Yan, Gongjun, Wu He, Hui Shi, and Danda B. Rawat. "Applying a bilingual model to mine e-commerce satisfaction sentiment", Journal of Management Analytics, 2015. Crossref	18 words — < 1%
3	Bostani, Hamid, and Mansour Sheikhan. "Modification of supervised OPF-based intrusion detection systems using unsupervised learning and social network concept", Pattern Recognition, 2017. Crossref	18 words — < 1%
4	airccse.org Internet	16 words — < 1%
5	Nick Nikiforakis. "HProxy: Client-Side Detection of SSL Stripping Attacks", Lecture Notes in Computer Science, 2010 Crossref	16 words — < 1%
6	vpacademic.acadiau.ca Internet	16 words — < 1%
7	Sun, Jia-Rong, Mao-Lin Shih, and Min-Shiang Hwang. "Cases study and analysis of the court judgement of cybercrimes in Taiwan", International Journal of Law Crime and Justice, 2015. Crossref	16 words — < 1%

8	www.cl.cam.ac.uk Internet	15 words — < 1%
9	www.isfs.org.hk Internet	15 words — < 1%
10	ir.canterbury.ac.nz Internet	14 words — < 1%
11	www.rnp.br Internet	13 words — < 1%
12	www.corpus-delicti.com Internet	13 words — < 1%
13	lrd.yahooapis.com Internet	13 words — < 1%
14	docs.com Internet	13 words — < 1%
15	mo.co.za Internet	12 words — < 1%
16	bks1.books.google.it Internet	12 words — < 1%
17	Jiang, Liu, Guiyan Tian, and Shidong Zhu. "Design and Implementation of Network Forensic System Based on Intrusion Detection Analysis", 2012 International Conference on Control Engineering and Communication Technology, 2012. Crossref	12 words — < 1%
18	Anand, Vijay. "Intrusion Detection : Tools, Techniques and Strategies", Proceedings of the 2014 ACM SIGUCCS Annual Conference on User Services Conference - SIGUCCS 14, 2014. Crossref	11 words — < 1%
19	196.21.83.35 Internet	

10 words — < 1 %

20	uad.academia.edu Internet	9 words — < 1%
21	Maji, Pradipta, and Ekta Shah. "Significance and Functional Similarity for Identification of Disease Genes", IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2016. Crossref	9 words — < 1%
22	Reddy, K., and H.S. Venter. "The architecture of a digital forensic readiness management system", Computers & Security, 2013. Crossref	8 words — < 1%
23	dblp.dagstuhl.de Internet	8 words — < 1%
24	Program: electronic library and information systems, Volume 50, Issue 1 (2016) Publications	8 words — < 1%
25	www.ijcst.org Internet	8 words — < 1%
26	Elwakil, Emad Zayed, Tarek. "Construction knowledge discovery system using fuzzy approach. (Report)", Canadian Journal of Civil Engineering, Jan 2015 Issue Publications	8 words — < 1%
27	Mousa, R.. "Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural", Expert Systems With Applications, 200505 Crossref	7 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	OFF		