Classification of Types of Computer Network Attacks Through IDS (Intrusion Detection System) using Naive Bayes Classifier

Tri Widodo
Department of Information Technology Education
Universitas Teknologi Yogyakarta
Yogyakarta, Indonesia

Adam Sekti Aji
Department of Informatics
Universitas Teknologi Yogyakarta
Yogyakarta, Indonesia

ABSTRACT

Computer network administrators use IDS (Intrusion Detection System) as part of a managed computer network protection system. IDS provides alerts or warnings to computer network administrators in the event of a computer network attack. All activities that pass through the computer network will be recorded in the IDS log or records. Computer network administrators need clearer information regarding what happens on the managed network such as the type of network attack, the number of attacks, and others. The most widely used classification algorithm is the Naïve Bayes Classifier. The use of Naïve Bayes Classifier is effective for grouping or classifying data based on existing data. This research is R&D research. This study aims to develop a website-based application that utilizes IDS log data classified using Naïve Bayes to identify computer network attacks. The website-based Naïve Bayes Classifier application developed can classify the types of network attacks recorded by the IDS. Network attacks can be identified by several variables, namely: Total incoming IP in range, packet length in range, time range, content, and destination port. Network administrators can improve computer network security by configuring the IDS rule using variable data processed by the Naïve Bayes Classifier application

Keywords

IDS (intrusion detection system), Network Attack, Naïve Bayes Classifier

1. INTRODUCTION

Many computer network administrators use IDS (Intrusion Detection System) as part of a managed computer network protection system. IDS acts as an early detection system or early detection system in the event of a computer network attack. IDS provides alerts or warnings to computer network administrators in the event of a computer network attack. Alerts or warnings given by IDS to network administrators are based on access to a computer network protocol, packet content, or several other variables. If access to a protocol is an access that violates the provisions set in the IDS rule, the computer network administrator will receive a warning of an attack. All activity through the computer network will be recorded in the IDS log or record. IDS records (logs) can be viewed using a log analyzer. The results of the log analyzer reading at this time only contain general readings which contain the number of login accesses, types of ports accessed, types of protocols most accessed, and others. This all-too-general reading of the logs does not provide sufficient information to computer network administrators. Computer network administrators need clearer information about what happened to the managed network, such as types of network attacks, number of attacks, and so on. The most widely used classification algorithm is the Naïve Bayes Classifier. The use of the Naïve Bayes Classifier is effective for grouping or classifying data based on previously existing data.

2. STUDY LITERATURE

2.1 Previous Study

Research related to the Naïve Bayes Classifier has been carried out by many researchers. The first research was conducted by Wirawan & Eksistyanto in 2017. The research conducted by Wirawan & Eksistyanto utilized the Naïve Bayes Classifier which uses discritized variables. In this study, discretization of variables can minimize classification errors from Naïve Bayes and can increase classification accuracy by up to 89% [1].

The second study was conducted by Fadlil, A., Riadi, I., & Aji, S in 2017. This study utilized Naïve Bayes to detect Distributed Denial of Service (DDoS) attacks. This research utilizes data taken from training and testing of network traffic on the core router at the Master of Information Technology Research Laboratory at Ahmad Dahlan University, Yogyakarta (MITRLADUY). This research tries to use a new approach in using Intrusion Detection System (IDS) to detect DDoS attacks [2]. Attack classification is based on the average and standard deviation of network packets according to Gaussian.

The third research was conducted by Tabash, M., Abd Allah, M., & Tawfik in 2019. This research is like the research conducted by Wirawan & Eksistyanto. Research conducted by Tabash et al tries to improve the accuracy of detecting network and computer system attacks with several techniques to find network anomalies using data mining and several other techniques such as genetic algorithms, Naïve Bayes, and several Deep Learning Techniques. The research conducted by Tabash et al was able to increase the accuracy of anomaly detection up to 99.9% [3].

The research that will be carried out is research that utilizes data which is then calculated and classified using Naïve Bayes. The research that will be carried out will not only classify using recorded attack data from IDS but will also develop applications that make it easier for network administrators to identify what types of attacks are recorded by IDS so that administrators can easily act and improve network security. In addition, the novelty aspect of this research lies in the search for new variables that have correlations and characterize certain types of network attacks, such as network ports, packet content, type of protocol and size of transmitted packets

2.2 Intrusion Detection System

Intrusion Detection System (IDS) is a device or software application that monitors data traffic on a computer network to detect malicious activity or policy violations [4]. There are several types of IDS [4], namely:

1. Network Intrusion Detection Systems (NIDS). Namely IDS which analyzes computer network data traffic.
2. Host-Based Intrusion Detection Systems (HIDS). Namely IDS which monitors operating system files.

2.3 Snort

Snort is an open source intrusion detection system (IDS) that is widely used to detect intrusions or suspicious activity in network traffic [5]. Snort is an example of a program from a Network-based Intrusion Detection System [6]. The way Snort works is similar to TcpDump, but focuses on security packet sniffing. The main feature of Snort that differentiates it from TcpDump is payload inspection, where Snort analyzes the payload rule set provided [7]. Snort has three components:

1. A sensor that can recognize security events.
2. Console that can monitor events and alerts and control sensors
3. Central Engine which is useful for storing logged events carried out by sensors into the database and uses security rules that are useful for handling events that occur.

According to Singh and Tomar, Snort works as a detection engine with IDS mode. When a packet comes through the switch, it will be detected by the detection engine on Snort, then Snort as an IDS will match the packet with the rules that have been set. When the packet does not comply with the rules (packet contains attack content) it will be stored in the Snort Log and will raise an alarm, but if the packet complies with the rules (does not contain attack content) then the packet is discarded (ignored) and forwarded directly [8].

2.4 Naive Bayes Classifier (NBC)

Classification is a process for determining a category from a set of objects whose category is unknown [9]. Naive Bayes Classifier (NBC) is a simple but efficient supervised document classification algorithm [10]. Naïve Bayes Classifier (NBC) is a classification method based on probability and the Bayesian Theorem with the assumption that each variable X is independent [9].

3. METHODOLOGY

3.1 Research Tools

The research tools used for this research are

1. Virtual Box
2. Ubuntu Operating System
3. Snort Intrusion Detection System
4. Snort Community rules

3.2 Research Stages

The research methodology used in writing this research is as follows:

3.2.1 Analysis

Analysis is used to find and collect material from various references related to IDS and network attacks. This analysis stage is also used to identify the network topology used to perform simulations and experiments.

3.2.2 Network design

Design at the installation and configuration stage of a computer network, two servers are used. The first server is a server that has the Ubuntu operating system installed and then IDS Snort installed. While the second server is a web server that has the Naïve Bayes application installed. This website-based naïve Bayes Classifier application will later be used to analyze data from IDS. The Figure 1 show the design of the network.

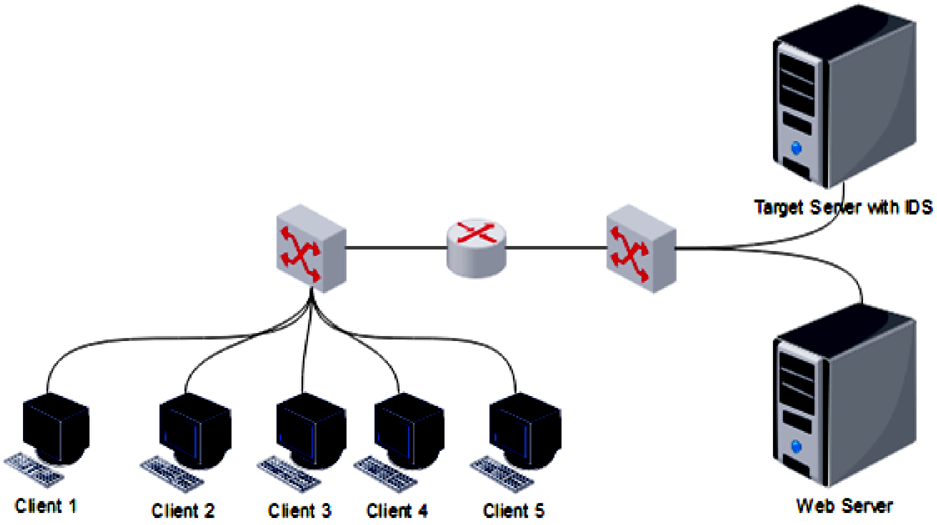


Fig 1: Network topology for research

3.2.3 IDS installation and configuration

The network topology design is implemented on a virtual box virtual machine using the Ubuntu operating system. then install IDS Snort on Ubuntu. Snort Community rules are then added to the installed Snort IDS. The community rules used are the community rules provided on the Snort IDS official website.

3.2.4 Simulation and testing

Simulating various attacks on the IDS-integrated server. After the computer network is configured and installed IDS snort. The next step is to carry out various network attack tests such as simulating DOS attacks, malware attacks, sending trojans, etc. This stage is carried out to find out whether IDS Snort can detect network attacks or not. IDS Snort detection results will be recorded in the log. This IDS Snort log will be processed by the application that will be developed to be classified using the Naive Bayes Classifier (NBC).

3.2.5 Classification using the Naive Bayes Classifier (NBC)

Analysis phase, at this stage an IDS log analysis is carried out and determines the criteria that will be used in the calculation of the Naïve Bayes Classifier. Classification is a process for determining a category from a set of objects whose category is unknown [9]. At this analysis stage, a needs analysis is also carried out in application development such as an analysis of functional system requirements and an analysis of non-functional requirements for the system to be developed.

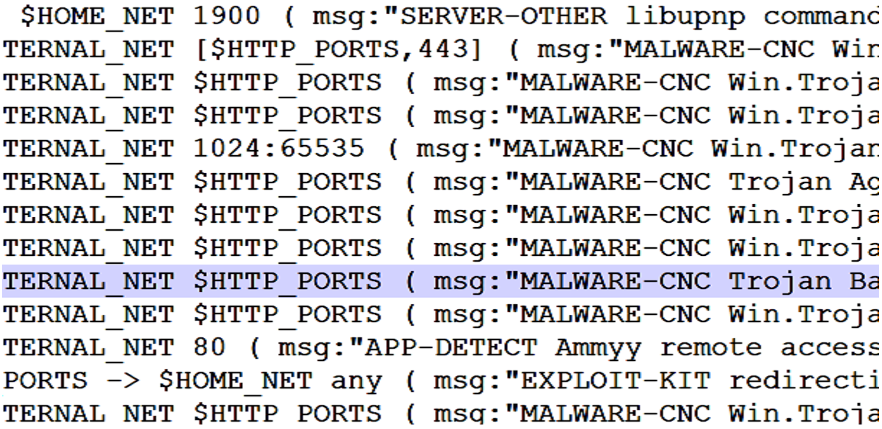
3.2.6 Analysis and validation of results

At this stage an analysis of the results of the classification carried out by the website-based application that has been developed is carried out. The results of the system classification are then validated to determine the accuracy of the data processing performed by the application.

4. RESULTS AND DISCUSSION

4.1 Implementation of IDS Snort

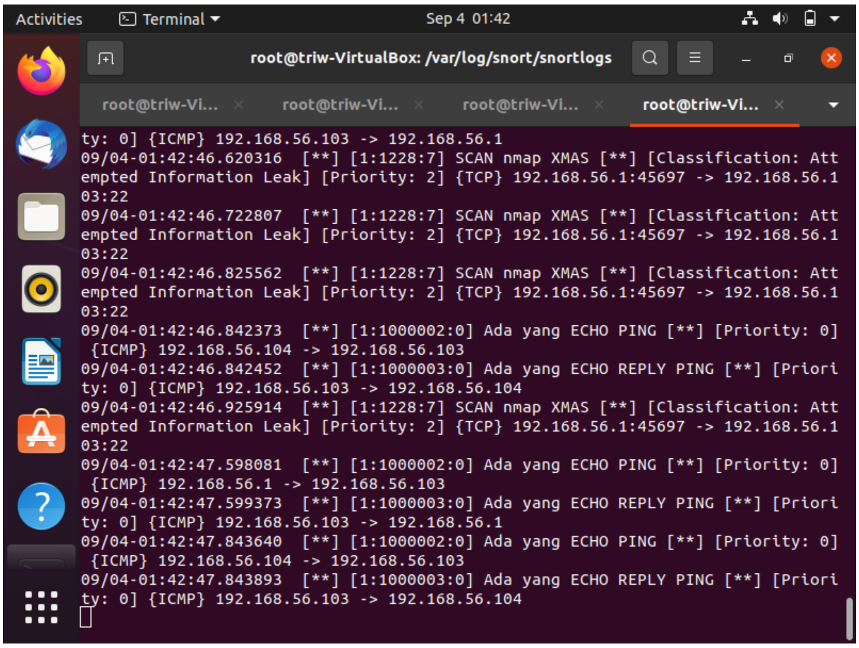
IDS Snort runs in several modes, sniffer mode, packet-logger mode, and IDS mode. Rule configuration in IDS can be done one rule at a time, or you can use rules that have been provided by the community. Determination of criteria in this study is based on the results of analysis and observations on IDS rules and log results obtained from IDS Snort. IDS snort is configured using community rules as shown in the picture. the Snort community rule itself has provided rules that allow IDS snort to detect network activity based on alert to take, protocol, origin from internal or external network (source address), port of origin (source port), destination post, destination protocol, message, content, metadata, service, references, and several other criteria. The Snort community rule itself has accommodated various rules for several types of attacks, both malware attacks, DOS attacks or other attacks. The results implementation and configuration of IDS Snort can be seen in Figure 2



```
 $HOME_NET 1900 ( msg:"SERVER-OTHER libupnp commanc
TERNAL_NET [$HTTP_PORTS,443] ( msg:"MALWARE-CNC Wir
TERNAL_NET $HTTP_PORTS ( msg:"MALWARE-CNC Win.Troja
TERNAL_NET $HTTP_PORTS ( msg:"MALWARE-CNC Win.Troja
TERNAL_NET 1024:65535 ( msg:"MALWARE-CNC Win.Trojar
TERNAL_NET $HTTP_PORTS ( msg:"MALWARE-CNC Trojan Ag
TERNAL_NET $HTTP_PORTS ( msg:"MALWARE-CNC Win.Troja
TERNAL_NET $HTTP_PORTS ( msg:"MALWARE-CNC Win.Troja
TERNAL_NET $HTTP_PORTS ( msg:"MALWARE-CNC Trojan Ba
TERNAL_NET $HTTP_PORTS ( msg:"MALWARE-CNC Win.Troja
TERNAL_NET 80 ( msg:"APP-DETECT Ammyy remote access
PORTS -> $HOME_NET any ( msg:"EXPLOIT-KIT redirecti
TERNAL_NET $HTTP_PORTS ( msg:"MALWARE-CNC Win.Troja
```

Fig 2: Snort IDS Rule display

The results of the IDS Snort log can be seen in the image below. records (logs) contain information on the origin of the IP packet, the destination IP of the packet, comments, the protocol used and the contents of the packet. The IDS log is also important information for all activities through computer networks, including computer network attack activities. The log if IDS Snort can be seen in Fig. 1.



```
ty: 0] {ICMP} 192.168.56.103 -> 192.168.56.1
09/04-01:42:46.620316  [**] [1:1228:7] SCAN nmap XMAS [**] [Classification: Att
empted Information Leak] [Priority: 2] {TCP} 192.168.56.1:45697 -> 192.168.56.1
03:22
09/04-01:42:46.722807  [**] [1:1228:7] SCAN nmap XMAS [**] [Classification: Att
empted Information Leak] [Priority: 2] {TCP} 192.168.56.1:45697 -> 192.168.56.1
03:22
09/04-01:42:46.825562  [**] [1:1228:7] SCAN nmap XMAS [**] [Classification: Att
empted Information Leak] [Priority: 2] {TCP} 192.168.56.1:45697 -> 192.168.56.1
03:22
09/04-01:42:46.842373  [**] [1:1000002:0] Ada yang ECHO PING [**] [Priority: 0]
 {ICMP} 192.168.56.104 -> 192.168.56.103
09/04-01:42:46.842452  [**] [1:1000003:0] Ada yang ECHO REPLY PING [**] [Priori
ty: 0] {ICMP} 192.168.56.103 -> 192.168.56.104
09/04-01:42:46.925914  [**] [1:1228:7] SCAN nmap XMAS [**] [Classification: Att
empted Information Leak] [Priority: 2] {TCP} 192.168.56.1:45697 -> 192.168.56.1
03:22
09/04-01:42:47.598081  [**] [1:1000002:0] Ada yang ECHO PING [**] [Priority: 0]
 {ICMP} 192.168.56.1 -> 192.168.56.103
09/04-01:42:47.599373  [**] [1:1000003:0] Ada yang ECHO REPLY PING [**] [Priori
ty: 0] {ICMP} 192.168.56.103 -> 192.168.56.1
09/04-01:42:47.843640  [**] [1:1000002:0] Ada yang ECHO PING [**] [Priority: 0]
 {ICMP} 192.168.56.104 -> 192.168.56.103
09/04-01:42:47.843893  [**] [1:1000003:0] Ada yang ECHO REPLY PING [**] [Priori
ty: 0] {ICMP} 192.168.56.103 -> 192.168.56.104
```

Fig 3: Record (log) Result of packet identification by IDS Snort

4.2 Variable of Network Attack

Based on the rules contained in IDS Snort and the results of the IDS log, several criteria are formulated that distinguish between normal network activity and conditions where there is an attack. The criteria used in this study are: Total incoming IP in range, packet length in range, time range, content, and destination port. This criterion was chosen because each computer network attack has certain characteristics contained in the five criteria, such as a DOS or DDOS attack that will originate from a computer that has an unreasonable number of incoming IPs and packet lengths in the range that are unreasonable and continuous. Malware or virus attacks will also have certain content characteristics. The data obtained from these criteria are then processed using the Naïve Bayes Classifier so that network activity can be easily identified and classified as shown in Figure 4.

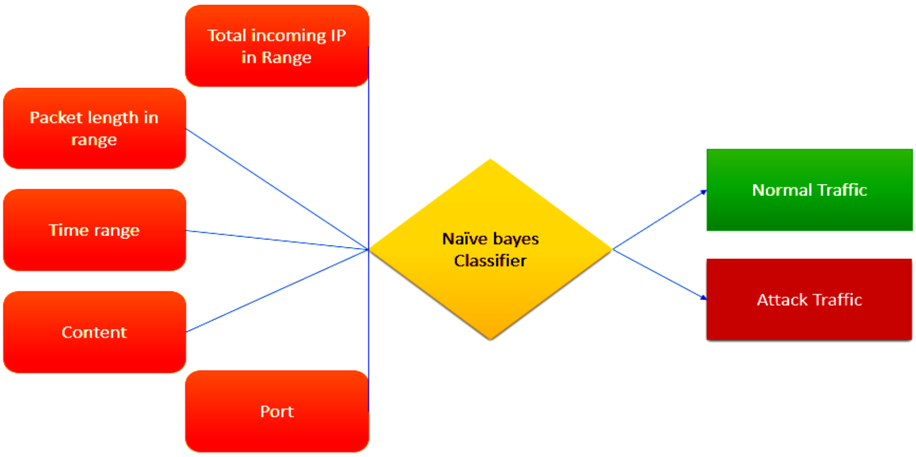


Fig 4: Example rule to detect DOS attack

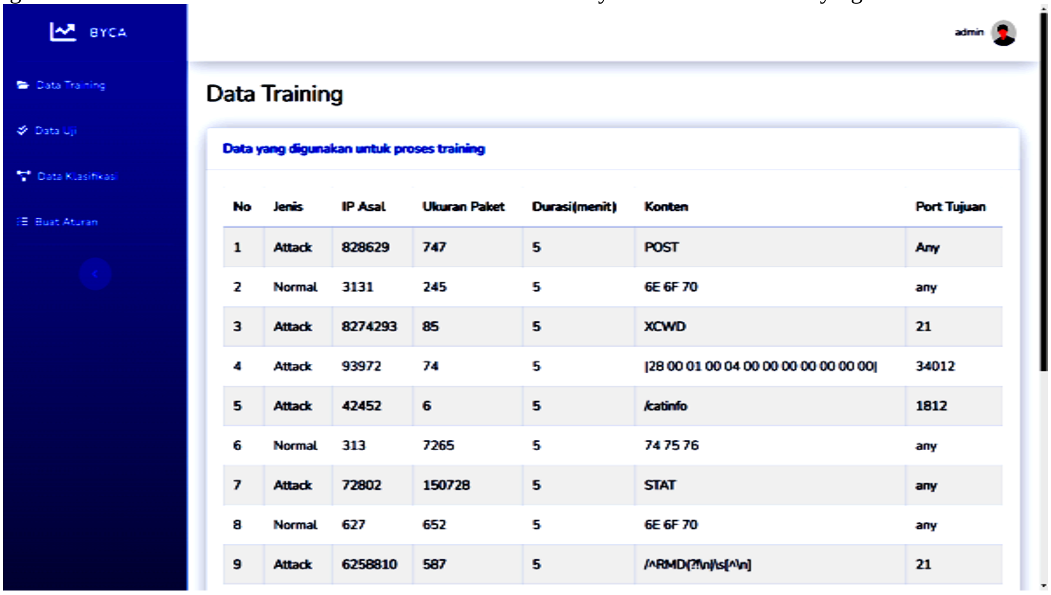
The results of the analysis of the IDS Snort community rule and network activity logs yield the criteria data shown in table 1. The data on these criteria (Figure 4) are then used to classify the type of attack using the Naïve Bayes Classifier. After obtaining the criteria that will be used for classification, the next step is to conduct data training.

Table 1: Data training into normal and attacks conditions

No	Classification	Variabel				
		Total incoming IP in Range	Total packet	Time range (Minute)	Content	Destination Port
1	Normal	21	41323	5	73 74 75	any
2	Normal	313	6521	5	64 65 66	any
3	Normal	313	7265	5	74 75 76	any
4	Normal	81681	5268	5	6C 6D 6E	any
5	Normal	71790	1244	5	6E 6F 70	any
6	Normal	3141	975	5	00 00 F2	any
7	Normal	14365	3218	5	69 6A 6B	any
8	Normal	3131	245	5	6E 6F 70	any
9	Normal	627	652	5	6E 6F 70	any
10	Normal	81689	1921	5	72 65 66	any
11	Attack	2863971	44	5	/^PASS(?!\n)	21
12	Attack	1672916	64	5	/^MKD(?!\n)	21
13	Attack	7265715	747	5	/^REST(?!\n)	21
14	Attack	6532781	457477	5	/^DELE(?!\n)	21
15	Attack	6258810	587	5	/^RMD(?!\n)\s[^\n]	21
16	Attack	1752	5858	5	/FtpSave.dll	2140
17	Attack	71651	58	5	/catinfo	3150
18	Attack	810	58	5	SITE	4120
19	Attack	18317	85856463	5	Ahhhh My Mouth Is Open	any
20	Attack	81980	150728	5	* Doly trojan v1.5 - Connected.	any
21	Attack	7345016	646	5	|00 01 86 A9|	any
22	Attack	2815811	74	5	CAPA	any
23	Attack	7432728	74	5	TOP	any
24	Attack	72802	150728	5	STAT	any
25	Attack	93972	74	5	|28 00 01 00 04 00 00 00 00 00 00 00|	34012

The training data process on the web uses data that has been uploaded to the system. The display of the training data menu is as seen in fig. 5.

This training data is used to provide information and knowledge base that will be used as a reference for the Naïve Bayes Classifier in classifying further data.



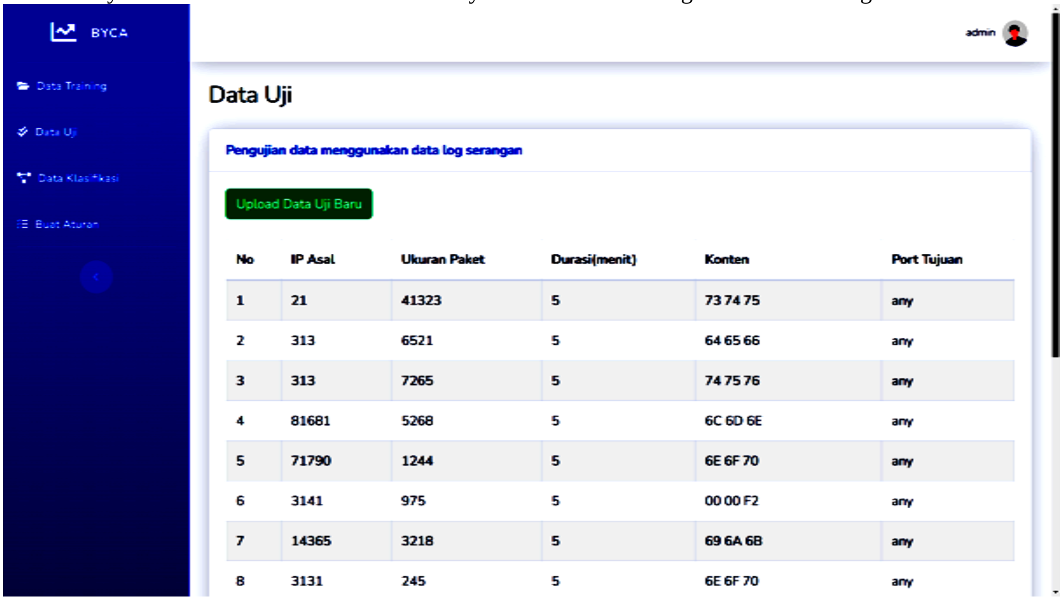
Data Training

Data yang digunakan untuk proses training

No	Jenis	IP Asal	Ukuran Paket	Durasi(menit)	Konten	Port Tujuan
1	Attack	828629	747	5	POST	Any
2	Normal	3131	245	5	6E 6F 70	any
3	Attack	8274293	85	5	XCWD	21
4	Attack	93972	74	5	|28 00 01 00 04 00 00 00 00 00 00 00|	34012
5	Attack	42452	6	5	/catinfo	1812
6	Normal	313	7265	5	74 75 76	any
7	Attack	72802	150728	5	STAT	any
8	Normal	627	652	5	6E 6F 70	any
9	Attack	6258810	587	5	/^RMD(?!\n)\s[^\n]	21

Fig 5: Process Data Training on web applications

In this study, 100 training data were used. The data is obtained based on a network simulation created and equipped with Snort community rule data to strengthen the grouping of types of network activity whether it is normal network activity or includes attack activity. After the data training process (data training) is carried out, data testing is then carried out to determine the accuracy of the Naïve Bayes Classifier method. Data testing can be seen in Figure 6.



Data Uji

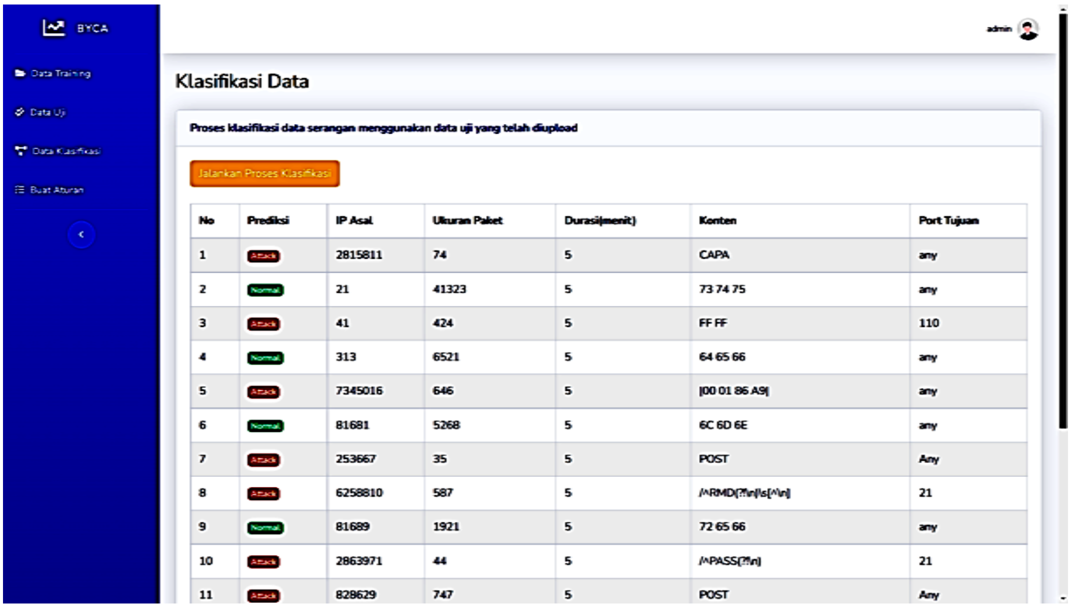
Pengujian data menggunakan data log serangan

No	IP Asal	Ukuran Paket	Durasi(menit)	Konten	Port Tujuan
1	21	41323	5	73 74 75	any
2	313	6521	5	64 65 66	any
3	313	7265	5	74 75 76	any
4	81681	5268	5	6C 6D 6E	any
5	71790	1244	5	6E 6F 70	any
6	3141	975	5	00 00 F2	any
7	14365	3218	5	69 6A 6B	any
8	3131	245	5	6E 6F 70	any

Fig 6: Data testing process

At the data classification stage, the naïve Bayes classifier application developed can classify network activity into 2 classifications, namely normal activity and attack (Figure 7).

The classification is based on previous training and testing processes because the Naïve Bayes Classifier is a supervised learning algorithm.



Klasifikasi Data

Proses klasifikasi data serangan menggunakan data uji yang telah diupload

No	Prediksi	IP Asal	Ukuran Paket	Durasi(menit)	Konten	Port Tujuan
1	Attack	2815811	74	5	CAPA	any
2	Normal	21	41323	5	73 74 75	any
3	Attack	41	424	5	FF FF	110
4	Normal	313	6521	5	64 65 66	any
5	Attack	7345016	646	5	|00 01 86 A9|	any
6	Normal	81681	5268	5	6C 6D 6E	any
7	Attack	253667	35	5	POST	Any
8	Attack	6258810	587	5	/^RMD(?!\n|\s[^\n])	21
9	Normal	81689	1921	5	72 65 66	any
10	Attack	2863971	44	5	/^PASS(?!\n)	21
11	Attack	828629	747	5	POST	Any

Fig 7: The results of the classification of types of network activity using the Naïve Bayes Classifier

Implementation of IDS Snort and development of classification applications using the Naïve Bayes Classifier can improve computer network security. IDS Snort monitors network activity in real time and prevents activity that violates IDS rules. The website-based Naïve Bayes Classifier classification application helps network administrators know the types of network activity that include attacks and understand the characteristics of types of attacks by looking at the criteria for types of computer network attacks. Administrators can increase the level of network security by setting IDS rules based on the criteria and conditions displayed by the Naïve Bayes Classifier application.

5. CONCLUSION

Based on the research conducted, the results of the discussion and analysis of application development, several conclusions can be drawn, namely:

1. Network attacks can be identified by several variables, namely: Total incoming IP in range, packet length in range, time range, content, and destination port.
2. The developed website-based Naïve Bayes Classifier application can classify the types of network attacks recorded by IDS.
3. Network administrators can improve computer network security by configuring IDS rules using variable data processed by the Naïve Bayes Classifier application
4. Network administrators can improve computer network security by configuring IDS rules using variable data processed by the Naïve Bayes Classifier application
5. Further research is expected to be able to collaborate on network security applications with artificial intelligence or machine learning applications. Research that combines computer network security applications and artificial intelligence or machine learning can improve computer network security because it is able to analyze computer network attacks or malware based on certain patterns.

6. REFERENCES

[1] Wirawan, I., & Eksistyanto, I. (2015). Penerapan Naive Bayes pada Intrusion Detection System Dengan Diskritisasi Variabel. JUTI: Jurnal Ilmiah Teknologi Informasi, 13(2), 182. https://doi.org/10.12962/j24068535.v13i2.a487

[2] Fadlil, A., Riadi, I., & Aji, S. (2017). Review of Detection DDOS Attack Detection Using Naive Bayes Classifier for Network Forensics. Bulletin Of Electrical Engineering and Informatics, 6(2), 140-148. https://doi.org/10.11591/eei.v6i2.605

[3] Tabash, M., Abd Allah, M., & Tawfik, B. (2019). Intrusion Detection Model Using Naive Bayes and Deep Learning Technique. The International Arab Journal Of Information Technology, 17(2), 215-224. https://doi.org/10.34028/iajit/17/2/9

[4] Barracuda.com. 2020. What Is an Intrusion Detection System? | Barracuda Networks. [online] Available at: https://www.barracuda.com/glossary/intrusion-detection-system#:~:text=An%20intrusion%20detection%20system%20(IDS,information%20and%20event%20management%20system. [Accessed 27 October 2020].

[5] Paramitha, I., Sasmita, G. and Raharja, I., 2020. Analisis Data Log IDS Snort dengan Algoritma Clustering Fuzzy C-Means. Majalah Ilmiah Teknologi Elektro, [online] 19(1), pp.95-99. Available at: https://ojs.unud.ac.id/index.php/JTE/article/view/58376/36819 [Accessed 27 October 2020]. -4

[6] Sandi, D. and Arrofiq, M., 2018. Implementasi Analisis NIDS Berbasis Snort Dengan Metode Fuzy Untuk Mengatasi Serangan LoRaWAN. JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi), [online] 2(3), pp.685-696. Available at: http://jurnal.iaii.or.id/index.php/RESTI/article/view/504 [Accessed 27 October 2020]. -5

[7] Dewi, E. and Kasih, P., 2017. Analisis Log Snort Menggunakan Network Forensic. JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), [online] 2(2), pp.72-79. Available at: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/370 [Accessed 27 October 2020]. -6

[8] Singh, R. and Tomar, D., 2015. Network Forensics:

Detection and Analysis of Stealth Port Scanning Attack. International Journal of Computer Networks and Communications Security, [online] 3(2), pp.33-42. Available at: http://www.ijcncs.org/published/volume3/issue2/p2_3-2.pdf [Accessed 27 October 2020]. -7

[9] hardianti, A.T., Manga, A. R., & Darwis, H. (2018). Penerapan Metode Naïve Bayes pada Klasifikasi Judul Jurnal. Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi 3(2). -2

[10] Pujianto, U., Widiyaningtyas, T., Prasetya, D., & Romadhon, B. (2019). Penerapan algoritma naïve bayes classifier untuk klasifikasi judul skripsi dan tugas akhir berdasarkan Kelompok Bidang Keahlian. TEKNO, 27(1), 79. https://doi.org/10.17977/um034v27i1p79-92