

Implementation of the Naive Bayes Classifier Method for Potential Network Port Selection

Rheo Malani, Arief Bramanto Wicaksono Putra, Muhammad Rifani

Department of Information Technology, Politeknik Negeri Samarinda, East Kalimantan, Indonesia

E-mail: anaogie@gmail.com*, ariefbram@gmail.com, cas.rifani@gmail.com

* corresponding author

Received: 28 November 2019; Accepted: 23 January 2020; Published: 08 April 2020

Abstract—The rapid development of information technology has also accompanied by an increase in activities classified as dangerous and irresponsible, such as information theft. In the field of network systems, this kind of activity is called intrusion. Intrusion Detection System (IDS) is a system that prevents intrusion and protecting both hosts and network assets. At present, the development of various techniques and methods for implementing IDS is a challenge, along with the increasing pattern of intrusion activities. The various methods used in IDS have generally classified into two types, namely Signature-Based Intrusion Detection System (SIDS) and the Anomaly-Based Intrusion Detection System (AIDS).

When a personal computer (PC) connected to the Internet, a malicious attacker tries to enter and exploit it. One of the most commonly used techniques in accessing open ports which are the door for applications and services that use connections in TCP/IP networks. Open ports indicate a particular process where the server provides certain services to clients and vice versa.

This study applies the Naive Bayes classifier to predict port numbers that have the potential to change activity status from "close" to "open" and vice versa. Predictable potential port numbers can be a special consideration for localizing monitoring activities in the future. The method applied is classified as AIDS because it based on historical data of port activity obtained through the port scan process, regardless of the type of attack. Naive Bayes classifier is determined to have two event conditions that predict the occurrence of specific port numbers when they occur in specified duration and activity status. The study results have shown a 70% performance after being applied to selected test data.

Index Terms—Intrusion, IDS, SIDS, AIDS, port scan, Naive Bayes classifier, potential port number.

I. INTRODUCTION

Data and information security is a crucial thing in the field of information technology. Nowadays, a lot of dangerous activities are carried out by someone for irresponsible things, such as information theft. In a network system, this dangerous activity classified as a

suspicious activity, which is commonly called intrusion. Cyber attacks are becoming increasingly sophisticated, which presents increasing challenges in accurate intrusion detection — failure to prevent intrusion results in decreased credibility of network system security services [1]. Intrusion Detection System (IDS) is widely used to protect both hosts and network assets. The key purpose of this technique is to help security administrators be able to recognize what IDS is doing. Many intrusion detection methods have proposed in various literature, which classified into the Signature-Based Intrusion Detection System (SIDS) and the Anomaly-Based Intrusion Detection System (AIDS) [2].

Signature-based intrusion detection techniques, also called misuse-based, have proven to be effective in detecting attacks without producing many false alarms. Attack detection is done by making a signature pattern of known attacks and storing them in the database as a priori information. However, this kind of approach cannot detect unknown attacks. Various applications of SIDS techniques have conducted in [3-12]. Anomaly-Based IDS (AIDS) is used to detect unknown and known attacks based on their profile or statistical model. These models use historical data on network usage to model and practice anomaly detection as a classification problem. This model tries to find anomalous than normal behavior. This model is more efficient and faster than SIDS, although many produce false-positive rates. Various studies related to various uses of algorithms in the AIDS approach have been carried out in [2, 13-23].

Deep Learning or often known as Deep Structured Learning or Hierarchical Learning, is one of the branches of machine learning that consists of high-level abstraction modeling algorithms in data using a set of functions non-linear transformations arranged in layers and depth. The techniques and algorithms can be used both for the needs of supervised learning, unsupervised learning, and semi-supervised learning in various applications. Deep Learning called "deep" because the structure and number of neural networks in the algorithm can reach up to hundreds of layers. Deep Learning has also widely used for IDS as in [24-32].

When a personal computer (PC) connected to the Internet, a malicious attacker tries to enter and exploit it. One of the most commonly used techniques is to access

open ports which are doors for applications and services that use connections in TCP/IP networks. Open ports indicate a particular process where the server provides certain services to clients and vice versa. The term "botnet" is a network consisting of infected end-hosts under the control of a human operator. SYN DDoS (Distributed Denial of Service) and Hypertext Transfer Protocol (HTTP) DDoS are the most common scenarios for botnet-assisted DDoS attacks [33]. DoS/DDoS attacks usually appear in the attached system. As a result, the service server that was attacked may crash or no longer be able to provide services to any client.

Transmission Control Protocol (TCP), as defined in Request For Comments (RFC) 793 [34], is a reliable, connection-oriented, end-to-end protocol. Initially, this was designed to fit in a layered hierarchy that supports multi-network applications. Some specifications of its capabilities are the establishment of connections, error recovery, flow control, and window size negotiations. In a typical TCP connection, each device maintains its status and the appropriate sequence number to trace the order of incoming packets. When the end-host device receives a new packet, it will send back an ACK (acknowledgment) packet, which contains an acknowledgment number. It indicates the device has successfully received data and is waiting for further incoming data under the numbers shown [35].

The status of the port that is the application door can be seen using the port scanning technique. Port scanning is a dangerous intrusion method for finding exploit loopholes. Port scanning works by checking the status of open ports on each host in a network. Port scanning can be called a form of information gathering that leads to services looking for potential targets. The collection of information on a computer network is not only used to find exploited loopholes but is also used to improve network system security. The concept of machine learning is one of the algorithmic approach techniques that can be used to study data patterns from information collected. Learning outcomes can be used as a reference to prevent exploitation from unwanted parties. The Naive Bayes classification method is one of the easiest data classification methods. The performance results are usually very dependent on the amount of training data used. In this study, the Naive Bayes classifier method was deliberately chosen as the method used because of its simplicity, to predict port numbers that could potentially change the activity status from "close" to "open" and vice versa. Predictable potential port numbers can be a special consideration for localizing future monitoring activities.

II. POTENTIAL PORT SELECTION METHODOLOGY

A. General Concept

In general, the methodology used in this study shown in Fig.1. Port scanning is a technique used to collect port status information from computers or devices connected to the network. For example, network administrators use port scans to recognize the open port status of a system so

that they can restrict access to that port, or turn it off completely. Port scanning grouped into three categories based on the various types of packets used. This study uses Full TCP that utilizes a *three-way handshake*, which is a full-duplex connection. When the *three-way handshake* process reached, a TCP connection will be established, and the port status is declared open.

Port scanning performed on a client that has a scenario of changing port status activities. This scenario in question is a client that has a port with an "open" status in one period, and the next period changes from "open" to "close". This scenario will be repeated until the scanning process is complete.

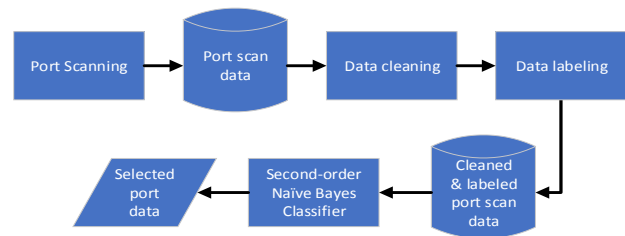


Fig.1. Research methodology

Some of the port scanning results that are of concern are as follows: (a). Port; (b). Duration; and (c). State. The "port" attribute contains the port number data obtained during the port scanning process for a certain period. The "duration" attribute contains data on the duration of "open/close" conditions of a port number. The "state" attribute contains the data condition class of a port number that is labeled "open" or "close."

The data cleaning stage is the process of ignoring scan data that does not change the port status activity from the initial scanning process until the scanning time is complete. Therefore, the remaining data is only the result of port scanning, which changes the port status activity. Changes in the status of port activity occur in a very varied duration. For simplicity, an evaluation of port status (change or not) done every 15 minutes. The port activity scan performed for one hour. The data labeling stage is the process of grouping the "duration" attribute into four classes, where each class is given a specific label, as follows:

- 1) 00:00:00 – 00:15:00 = T_1
- 2) 00:16:00 – 00:30:00 = T_2
- 3) 00:31:00 – 00:45:00 = T_3
- 4) 00:46:00 – 00:60:00 = T_4

B. Naive Bayes Classifier

Naive Bayes Classifier, based on the Bayes theorem, is the theorem used in statistics to calculate the probability of a hypothesis. Bayes calculates the probability of a class based on its attributes and determines which class has the highest probability. In machine learning, Naive Bayes classifies classes based on simple probabilities by assuming that each attribute in the data is mutually exclusive. The Naive Bayes method is one of the most widely used methods based on several simple properties. The Naive Bayes method classifies data based on data

attributes expressed as: $x = (x_1, x_2, \dots, x_n)$ on the probabilities model of each class k which can be written as follows:

$$P(y_k | x_1, x_2, \dots, x_n) \quad (1)$$

where n is the number of attributes in the data and k is the number of classes in the class y data set. Classification is a scheme of determining a particular data into a class that seen from the perspective of probability into Bayes rules written as follows:

$$P(y_k | x_n) = \frac{P(y_k) \cdot P(x_n | y_k)}{P(x_n)} \quad (2)$$

where $P(y_k | x_n)$ is the probability of the event y_k occurring when x_n occurs, $P(x_n | y_k)$ is the probability of the event x_n occurring when y_k occurs, $P(y_k)$ is the probability of the event y_k , and $P(x_n)$ is the probability of the event x_n .

The highest probability value of each possible class is chosen as the optimal class using the following formula:

$$\arg \max_{y_k \in y} = \frac{P(y_k) \cdot P(x_n | y_k)}{P(x_n)} \quad (3)$$

Because the value of $P(x_n)$ is always the same for each class, the equation can be written as:

$$\arg \max_{y_k \in y} P(y_k) \cdot P(x_n | y_k) \quad (4)$$

If A is the "duration" attribute that represents the duration class, B is the "port" attribute that represents the class of port number, C is the "state" attribute that represents the condition class of a port, the probability of events A , B , and C expressed by:

$$P(A_i) = \frac{n_{A_i}}{n} \quad P(B_i) = \frac{n_{B_i}}{n} \quad P(C_i) = \frac{n_{C_i}}{n} \quad (5)$$

where $n_{A_i}, n_{B_i}, n_{C_i}$ are the number of events $A_i, B_i,$ and C_i , respectively, i is the number of class for each attribute, and n is the number of total data.

The number of classes of "duration" attributes is four (T_1, \dots, T_4). The number of port numbers ranges from 0 - 65535. Port numbers divided into 3, namely *well-known ports* ranging from 0 - 1023, *registered ports* ranging from 1024 - 49151 and *private/dynamic ports* ranging from 49152 - 65535. In this study uses the *well-known ports*. Because observations only made on ports that have changed activity status during the scanning process, in this study, there were only ten ports with that condition. Therefore, the number of classes of "port" attributes is 10 ports. The number of classes of "state" attributes is 2

("open / close").

The probability of an occurrence of a specific port with a specific activity status is expressed by:

$$P(B_i | C_i) = \frac{n_{B_i C_i}}{n_{C_i}} \quad (6)$$

The probability of an occurrence of a specific port occurring in a specified duration is expressed by:

$$P(B_i | A_i) = \frac{n_{B_i A_i}}{n_{A_i}} \quad (7)$$

The probability of a specific duration occurrence having a specific activity status is expressed by:

$$P(A_i | C_i) = \frac{n_{A_i C_i}}{n_{C_i}} \quad (8)$$

The probability of a specific duration occurring when a specific port occurs is expressed by:

$$P(A_i | B_i) = \frac{P(A_i) \cdot P(B_i | A_i)}{P(B_i)} \quad (9)$$

The probability of a port activity occurring with a specific status when a particular port occurs is expressed by:

$$P(C_i | B_i) = \frac{P(C_i) \cdot P(B_i | C_i)}{P(B_i)} \quad (10)$$

Several training data are used to obtain $P(A_i | B_i)$ and $P(C_i | B_i)$. Suppose two events X and Y , with $P(Y) > 0$. The conditional probability of X given Y expressed by [36]:

$$P(X | Y) = \frac{P(X \cap Y)}{P(Y)} \quad (11)$$

This study attempts to select potential ports by predicting the occurring of specific ports with certain status activities within a specific duration $P(B_i | A_i \text{ and } C_i)$. By using Eq. (9), (10), and (11) can be obtained:

$$\begin{aligned} P(B_i | A_i \text{ and } C_i) &= P(B_i | A_i) \cap P(B_i | C_i) \\ &= \left(\frac{P(A_i | B_i) \cdot P(B_i)}{P(A_i)} \right) \cap \left(\frac{P(C_i | B_i) \cdot P(B_i)}{P(C_i)} \right) \end{aligned} \quad (12)$$

Finally, several test data are used to obtain $P(B_i | A, \text{ and } C_i)$.

III. IMPLEMENTATION

A. Implementation of the three-way handshake full TCP connection

Port scanning using a three-way handshake full TCP connection done in real-time as illustrated in Figure 2. The port status will be declared open when the three-way handshake process occurs. The three-way handshake process begins with the first host sending synchronize flags (*SYN*) to the second host, followed by the second host responding by sending synchronize (*SYN*) and acknowledgments (*ACK*). Finally, the first host will respond with the acknowledgment flag (*ACK*). As for the port, status declared closed, the second host will respond by sending a reset flag (*RST*) and Acknowledgment (*ACK*) when receiving a synchronize flag (*SYN*) from the first host.

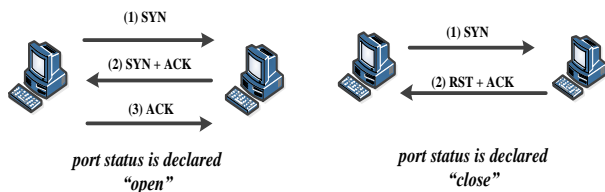


Fig.2. The three-way handshake full TCP connection

This study uses a research scenario as shown in Fig. 3. All work stations (client 1 ... client n) are randomly active. Observation of port status activities carried out for one hour. The results of port scanning have produced raw data for this research stored in data storage. From several raw data obtained, only 400 data were used and have been through the process of cleaning and labeling the data. Three hundred fifty data (350) used as training data, and 50 data used as test data. Training data have shown in Table 1, while test data have shown in Table 2, and graphically shown in Fig. 4 - 11.

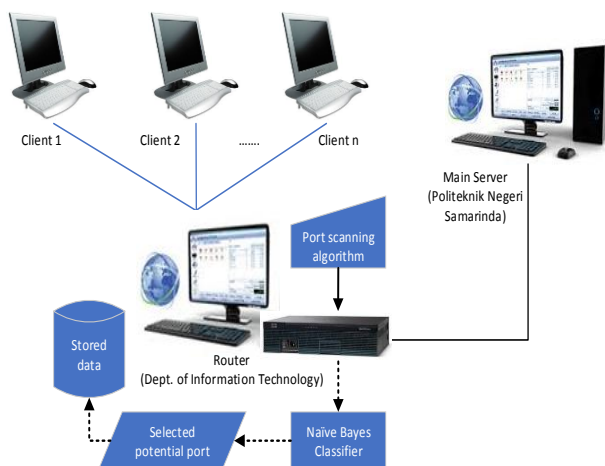


Fig.3. Research scenario

Table 1. Training data

| No. | Port (B) | Duration (A) | State (C) |
|-----|----------|--------------|-----------|
| 1 | 20 | T_2 | open |
| 2 | 21 | T_1 | close |
| 3 | 22 | T_3 | close |
| 4 | 53 | T_2 | close |
| 5 | 80 | T_2 | open |
| 6 | 110 | T_4 | open |
| 7 | 111 | T_3 | close |
| 8 | 143 | T_3 | open |
| 9 | 443 | T_2 | close |
| 10 | 995 | T_3 | open |
| 11 | 20 | T_1 | close |
| 12 | 21 | T_2 | open |
| 13 | 22 | T_3 | open |
| 14 | 53 | T_1 | open |
| 15 | 80 | T_1 | close |
| ... | ... | ... | ... |
| ... | ... | ... | ... |
| 101 | 20 | T_2 | open |
| 102 | 21 | T_1 | close |
| 103 | 22 | T_4 | close |
| 104 | 53 | T_2 | close |
| 105 | 80 | T_2 | open |
| ... | ... | ... | ... |
| ... | ... | ... | ... |
| 346 | 110 | T_4 | open |
| 347 | 111 | T_3 | close |
| 348 | 143 | T_1 | open |
| 349 | 443 | T_1 | close |
| 350 | 995 | T_3 | open |

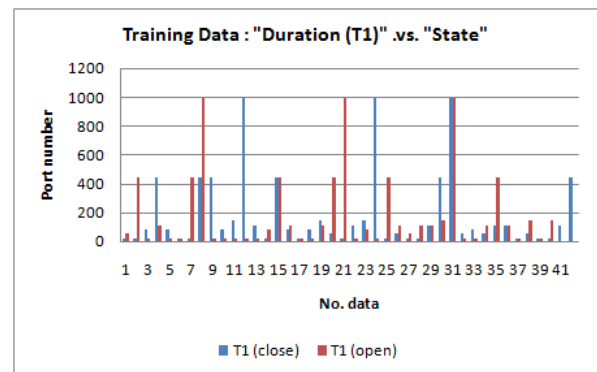


Fig.4. Training data: "Duration (T_1) .vs. State

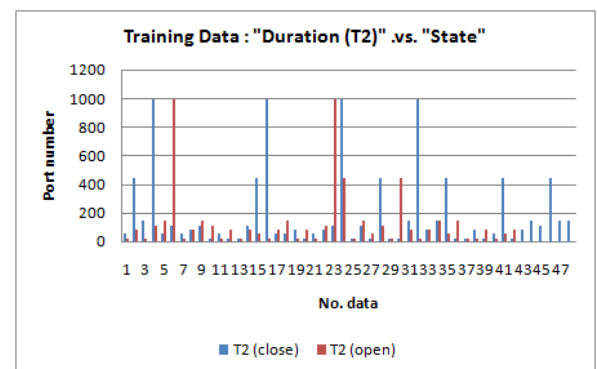


Fig.5. Training data: "Duration (T_2) .vs. State

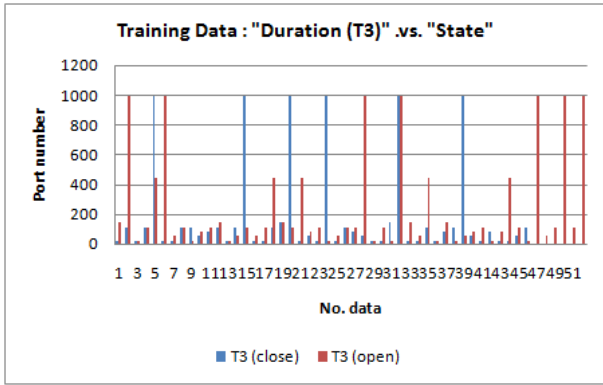


Fig.6. Training data: "Duration (T_3) .vs. State"

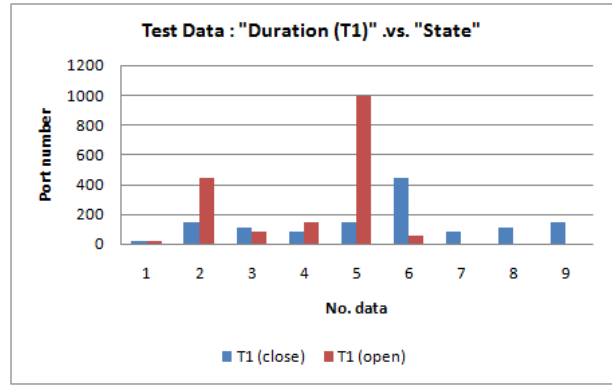


Fig.8. Test data: "Duration (T_1) .vs. State"

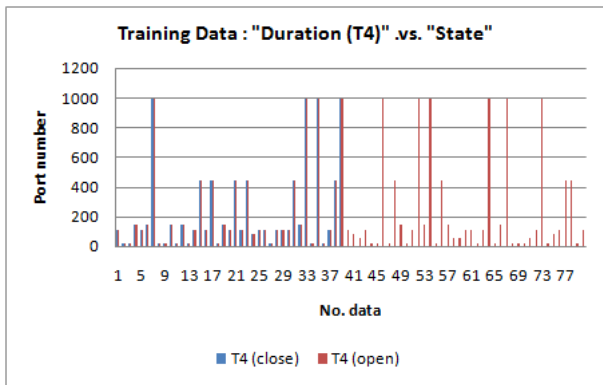


Fig.7. Training data: "Duration (T_4) .vs. State"

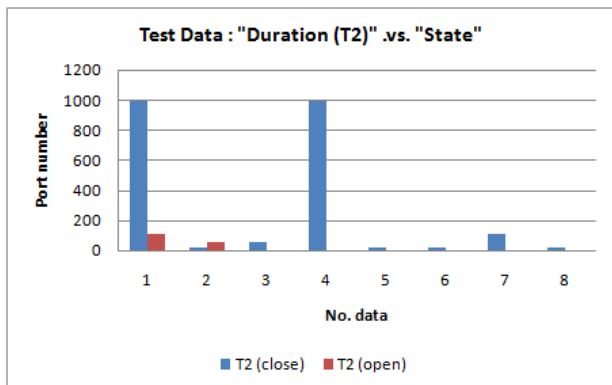


Fig.9. Test data: "Duration (T_2) .vs. State"

Table 2. Test data

| No. | Port (B) | Duration (A) | State (C) |
|-----|----------|--------------|-----------|
| 1 | 20 | T_1 | close |
| 2 | 21 | T_3 | open |
| 3 | 22 | T_1 | open |
| 4 | 53 | T_3 | open |
| 5 | 80 | T_4 | close |
| 6 | 110 | T_3 | close |
| 7 | 111 | T_3 | open |
| 8 | 143 | T_1 | close |
| 9 | 443 | T_4 | open |
| ... | ... | ... | ... |
| ... | ... | ... | ... |
| 46 | 110 | T_1 | close |
| 47 | 111 | T_4 | open |
| 48 | 143 | T_1 | close |
| 49 | 443 | T_4 | open |
| 50 | 995 | T_4 | close |

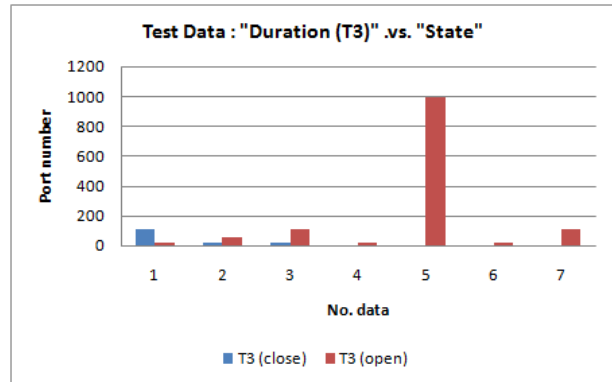


Fig.10. Test data: "Duration (T_3) .vs. State"

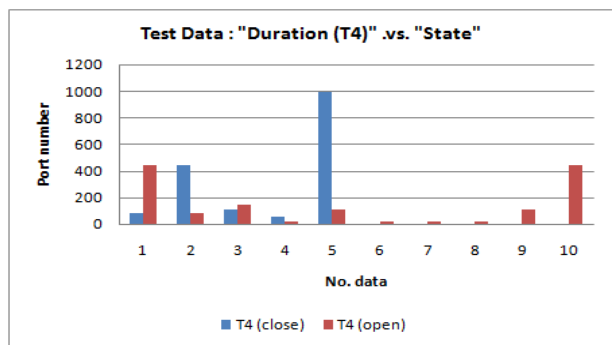


Fig.11. Test data: "Duration (T_4) .vs. State"

B. Implementation of Naive Bayes Classifier

The event probability of all classes in each attribute calculated by using Eq. (5) with results as shown in Tables 3, 4, and 5. While the probabilities were calculated using Eq. (6), (7), and (8) produce as shown in Tables 6, 7, and 8.

Table 3. The event probability of the "duration" attribute

| | | | |
|------------|------------|------------|------------|
| $P(A=T_1)$ | $P(A=T_2)$ | $P(A=T_3)$ | $P(A=T_4)$ |
| 0.234286 | 0.257143 | 0.28 | 0.22857 |

Table 4. The event probability of the "state" attribute

| No. Port | $P(B)$ |
|----------|--------|
| 20 | 0.1 |
| 21 | 0.1 |
| 22 | 0.1 |
| 53 | 0.1 |
| 80 | 0.1 |
| 110 | 0.1 |
| 111 | 0.1 |
| 143 | 0.1 |
| 443 | 0.1 |
| 995 | 0.1 |

Table 5. The event probability of the "port" attribute

| | |
|---------------|----------------|
| $P(C="open")$ | $P(C="close")$ |
| 0.5 | 0.5 |

Table 6. The probability of the event "port" (B) occurring when "state" (C) occurs

| No. Port (B) | $P(B C="open")$ | $P(B C="close")$ |
|--------------|-----------------|------------------|
| 20 | 0.1029 | 0.0971 |
| 21 | 0.0971 | 0.1029 |
| 22 | 0.0971 | 0.1029 |
| 53 | 0.0971 | 0.1029 |
| 80 | 0.1029 | 0.0971 |
| 110 | 0.1029 | 0.0971 |
| 111 | 0.0971 | 0.1029 |
| 143 | 0.1029 | 0.0971 |
| 443 | 0.0971 | 0.1029 |
| 995 | 0.1029 | 0.0971 |

Table 7. the probability of the event "port" (B) occurring when "duration" (A) occurs

| No. Port (B) | $P(B A=T_1)$ | $P(B A=T_2)$ | $P(B A=T_3)$ | $P(B A=T_4)$ |
|--------------|--------------|--------------|--------------|--------------|
| 20 | 0.0976 | 0.1000 | 0.1225 | 0.0750 |
| 21 | 0.1342 | 0.1000 | 0.0918 | 0.0750 |
| 22 | 0.1220 | 0.0889 | 0.0714 | 0.1250 |
| 53 | 0.0854 | 0.1333 | 0.1225 | 0.0500 |
| 80 | 0.0976 | 0.1778 | 0.0816 | 0.0375 |
| 110 | 0.0976 | 0.0444 | 0.1327 | 0.1250 |
| 111 | 0.0732 | 0.0667 | 0.1225 | 0.1375 |
| 143 | 0.0732 | 0.1333 | 0.0714 | 0.1250 |
| 443 | 0.1463 | 0.0889 | 0.0510 | 0.1250 |
| 995 | 0.0732 | 0.0667 | 0.1327 | 0.1250 |

Table 8. the probability of the event "duration" (A) occurring when "state" (C) occurs

| Duration (A) | $P(A C="open")$ | $P(A C="close")$ |
|--------------|-----------------|------------------|
| T_1 | 0.2286 | 0.2400 |
| T_2 | 0.2400 | 0.2743 |
| T_3 | 0.2971 | 0.2629 |
| T_4 | 0.2343 | 0.2229 |

Table 9. The probability of a specific duration(A_i) occurring when a specific port (B_i) occurs

| No. Port (B) | $P(A=T_1 B_i)$ | $P(A=T_2 B_i)$ | $P(A=T_3 B_i)$ | $P(A=T_4 B_i)$ |
|--------------|----------------|----------------|----------------|----------------|
| 20 | 0.2286 | 0.2571 | 0.3429 | 0.1714 |
| 21 | 0.3143 | 0.2571 | 0.2571 | 0.1714 |
| 22 | 0.2857 | 0.2286 | 0.2000 | 0.2857 |
| 53 | 0.2000 | 0.3429 | 0.3429 | 0.1143 |
| 80 | 0.2286 | 0.4571 | 0.2286 | 0.0857 |
| 110 | 0.2286 | 0.1143 | 0.3714 | 0.2857 |
| 111 | 0.1714 | 0.1714 | 0.3429 | 0.3143 |
| 143 | 0.1714 | 0.3429 | 0.2000 | 0.2857 |
| 443 | 0.3429 | 0.2286 | 0.1429 | 0.2857 |
| 995 | 0.1714 | 0.1714 | 0.3714 | 0.2857 |

Table 10. The probability of a port activity occurring with a specific status (C_i) when a specific port (B_i) occurs

| No. Port (B) | $P(C="open" B_i)$ | $P(C="close" B_i)$ |
|--------------|-------------------|--------------------|
| 20 | 0.5143 | 0.4857 |
| 21 | 0.4857 | 0.5143 |
| 22 | 0.4857 | 0.5143 |
| 53 | 0.4857 | 0.5143 |
| 80 | 0.5143 | 0.4857 |
| 110 | 0.5143 | 0.4857 |
| 111 | 0.4857 | 0.5143 |
| 143 | 0.5143 | 0.4857 |
| 443 | 0.4857 | 0.5143 |
| 995 | 0.5143 | 0.4857 |

Table 11. Comparison between the predicted results and the actual data

| No. | No. Port | Predicted | | Actual | | True/False |
|-----|----------|-----------|-------|----------|-------|------------|
| | | duration | state | duration | State | |
| 1 | 20 | T_1 | close | T_1 | close | True |
| 2 | 21 | T_3 | open | T_3 | open | True |
| 3 | 22 | T_1 | open | T_1 | open | True |
| 4 | 53 | T_3 | open | T_3 | open | True |
| 5 | 80 | T_4 | close | T_4 | close | True |
| 6 | 110 | T_4 | open | T_3 | close | False |
| 7 | 111 | T_3 | open | T_3 | open | True |
| 8 | 143 | T_1 | close | T_1 | close | True |
| 9 | 443 | T_4 | open | T_4 | open | True |
| ... | ... | ... | ... | ... | ... | |
| ... | ... | ... | ... | ... | ... | |
| 46 | 110 | T_1 | close | T_1 | close | True |
| 47 | 111 | T_4 | open | T_4 | open | True |
| 48 | 143 | T_1 | close | T_1 | close | True |
| 49 | 443 | T_4 | open | T_4 | open | True |
| 50 | 995 | T_4 | close | T_4 | close | True |

The probabilities were calculated using Eq. (9) and (10) produce as shown in Tables 9 and 10.

Probabilities $P(A_i|B_i)$ and $P(C_i|B_i)$ that have obtained from the training stage, are then used to predict potential ports by using test data through the application of Eq. (12). The results were presented in the form of a comparison between the predicted results and the actual data as shown in Table 11, and graphically shown in Fig. 12 - 15.

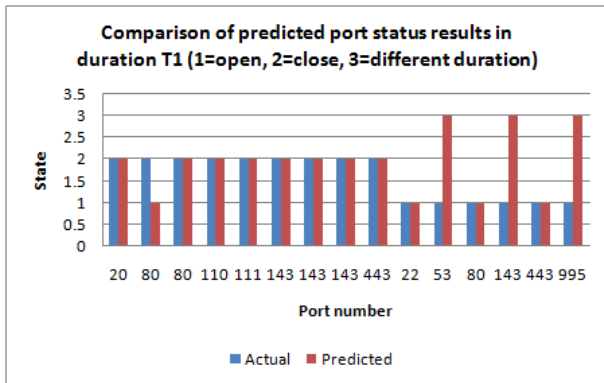


Fig.12. Comparison of predicted port status results in duration T₁

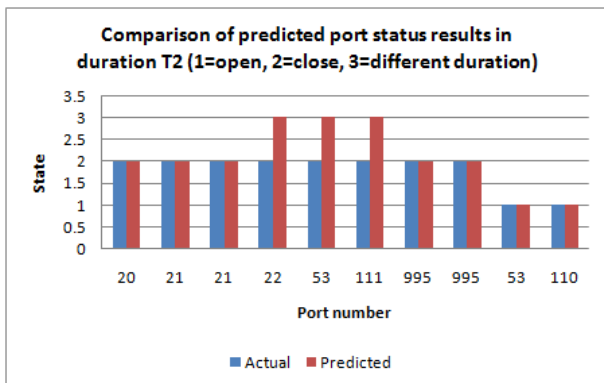


Fig.13. Comparison of predicted port status results in duration T₂

Prediction performance measured by the number of "true" to the total test data, obtained:

$$\frac{35}{50} \times 100\% = 70\%$$

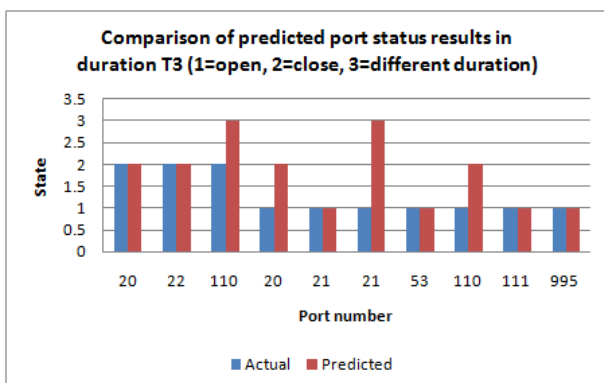


Fig.14. Comparison of predicted port status results in duration T₃

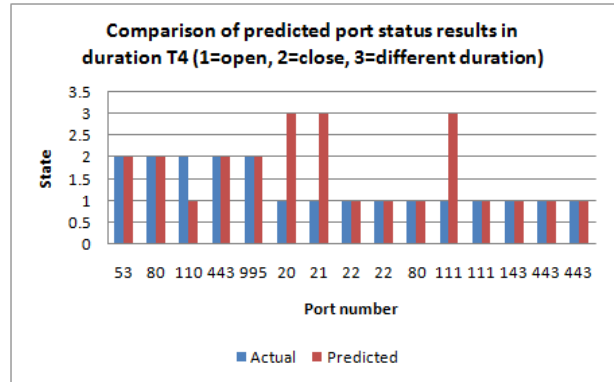


Fig.15. Comparison of predicted port status results in duration T₄

IV. CONCLUSION

This study has applied the Naive Bayes Classifier with two conditions for selecting potential ports. The method applied is classified as AIDS because it based on historical data of port activity obtained through the port scan process, regardless of the type of attack. The three attributes that have used are "port" which contains the number of ports scanned for a certain period, "duration" which contains the duration of the "open/close" condition of a port, and "state" which contains the condition of a port with the status "open" or "close". This study has chosen a potential port by predicting the occurrence of specific ports with certain status activities within a specified duration through the use of training data. The results have used to predict potential ports through the use of test data. The results of this study have shown a predictive performance of 70%. This result can be considered good enough to predict potential port numbers in the following occurrence. Prediction results for the occurrence of the next potential port number within a specific period will be compared with the results of the port status scan stage. Furthermore, the classification stage using the Naive Bayes method will be carried out to predict the potential port number at the next occurrence, where the scan results of the newly obtained port status will be involved as raw data. In this way, the amount of raw data will increase. The increasing number of raw data will further improve the performance of prediction results using the Naive Bayes method.

For further studies, the improvement of prediction performance using the Naive Bayes Classifier method will be carried out by making use of the renewal of the raw data that results from scanning port activity within a specified period. It will be done by modifying the research scenario, as shown in Fig.3.

REFERENCES

- [1] A. Khraisat, I. Gondal, P. Vamplew *et al.*, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, 2019.
- [2] W. Alhakami, "Alerts Clustering for Intrusion Detection Systems: Overview and Machine Learning Perspectives," *International Journal of Advanced Computer Science and*

- Applications (IJACSA)*, vol. 10, no. 5, 2019.
- [3] R. Bogdan, "Detecting Malicious Codes: A Signature-Based Solution," *International Conference on Computer and Software Modeling, IACSIT Press, Singapore*, 2011.
 - [4] P. H. A. A, J. M *et al.*, "Signature-Based IDS for Software-Defined Networking," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 7, no. 9, 2018.
 - [5] P. P. Ioulianou, V. G. Vassilakis, I. D. Moscholios *et al.*, "A Signature-based Intrusion Detection System for the Internet of Things," *White Rose Research Online, University of York*, 2018.
 - [6] V. Kumar, and D. O. P. Sangwan, "Signature Based Intrusion Detection System Using SNORT," *International Journal of Computer Applications & Information Technology*, vol. I, no. III, 2012.
 - [7] N. Mastorakis, A. Andreatos, V. Moussas *et al.*, "A Novel Intrusion Detection System Based on Neural Networks," *MATEC Web of Conferences*, vol. 292, pp. 03017, 2019.
 - [8] W. Meng, W. Li, C. Su *et al.*, "Enhancing Trust Management for Wireless Intrusion Detection via Traffic Sampling in the Era of Big Data," *IEEE Access*, vol. 6, pp. 7234-7243, 2018.
 - [9] Prof.D.P.Gaikwad, P. Pabshettiwar, P. Musale *et al.*, "A Proposal for Implementation of Signature Based Intrusion Detection System Using Multithreading Technique," *International Journal Of Computational Engineering Research*, vol. 2, no. 7, 2012.
 - [10] N. Sameera, and M. Shashi, "Transfer Learning Based Prototype for Zero-Day Attack Detection," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 4, 2019.
 - [11] S. N. Shah, and M. P. Singh, "Signature-Based Network Intrusion Detection System Using SNORT And WINPCAP," *International Journal of Engineering Research & Technology (IJERT)*, vol. 1, no. 10, 2012.
 - [12] Thamizharasi.E, and P.Salini, "Survey on Fuzzy Based Extreme Learning Machine for Intrusion Detection," *IOSR Journal of Engineering (IOSR JEN)*, pp. 69-76, 2019.
 - [13] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *Journal of Computational Science*, vol. 25, pp. 152-160, 2018.
 - [14] S. K. Amrita, "Machine Learning and Feature Selection Approach for Anomaly based Intrusion Detection: A Systematic Novice Approach," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 6S, 2019.
 - [15] S. Jose, D. Malathi, B. Reddy *et al.*, "A Survey on Anomaly Based Host Intrusion Detection System," *Journal of Physics: Conference Series*, vol. 1000, pp. 012049, 2018.
 - [16] S. Khonde, and U. Venugopal, "Hybrid Architecture for Distributed Intrusion Detection System," *Ingénierie des systèmes d'information*, vol. 24, no. 1, pp. 19-28, 2019.
 - [17] H. Li, F. Wei, and H. Hu, "Enabling Dynamic Network Access Control with Anomaly-based IDS and SDN," *SDN/NFV Security Architecture, Association for Computing Machinery (ACM)*, <https://doi.org/10.1145/3309194.3309199>, pp. 13-16, 2019.
 - [18] E. Nikolova, and V. Jecheva, "Applications of Clustering Methods to Anomaly-Based Intrusion Detection Systems," *8th International Conference on Database Theory and Application, Jeju, South Korea*, pp. 37-41, 2015.
 - [19] Z. Rustam, and A. S. Talita, "Fuzzy Kernel Robust Clustering for Anomaly based Intrusion Detection," *Third International Conference on Informatics and Computing (ICIC), Palembang, Indonesia, Indonesia, IEEE*, 2018.
 - [20] B. A. Tama, M. Comuzzi, and K.-H. Rhee, "TSE-IDS: A Two-Stage Classifier Ensemble for Intelligent Anomaly-based Intrusion Detection System," *IEEE Access*, vol. 7, 2019.
 - [21] N. T. T. Van, and T. N. Thinh, "Accelerating Anomaly-Based IDS Using Neural Network on GPU," *2015 International Conference on Advanced Computing and Applications, Ho Chi Minh City, Vietnam*, pp. 67-74, 2015.
 - [22] F. Wang, H. Zhu, B. Tian *et al.*, "A HMM-based method for Anomaly Detection," *2011 4th IEEE International Conference on Broadband Network and Multimedia Technology, Shenzhen, China*, 2011.
 - [23] C. Young, H. Olufowobi, G. Bloom *et al.*, "Automotive Intrusion Detection Based on Constant CAN Message Frequencies Across Vehicle Driving Modes," *Association for Computing Machinery (ACM)*, <https://doi.org/10.1145/3309171.3309179>, pp. 9-14, 2019.
 - [24] D. AKSU, and M. A. AYDIN, "Detecting Port Scan Attempts with Comparative Analysis of Deep Learning and Support Vector Machine Algorithms," *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), ANKARA, Turkey, Turkey, IEEE*, 2018.
 - [25] M. Al-Qatf, Y. Lasheng, M. Al-Habib *et al.*, "Deep Learning Approach Combining Sparse Autoencoder With SVM for Network Intrusion Detection," *IEEE Access*, vol. 6, pp. 52843-52856, 2018.
 - [26] G. Karatas, O. Demir, and O. K. Sahingoz, "Deep Learning in Intrusion Detection Systems," *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), ANKARA, Turkey, Turkey, IEEE*, 2018.
 - [27] S. M. Kasongo, and Y. Sun, "A Deep Learning Method With Filter Based Feature Engineering for Wireless Intrusion Detection System," *IEEE Access*, vol. 7, pp. 38597-38607, 2019.
 - [28] Navaporn, Chockwanich, Vasaka *et al.*, "Intrusion Detection by Deep Learning with TensorFlow," *2019 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang Kwangwoon_Do, Korea (South), IEEE*, 2019.
 - [29] S. Otoum, B. Kantarci, and H. T. Mouftah, "On the Feasibility of Deep Learning in Sensor Network Intrusion Detection," *IEEE Networking Letters*, vol. 1, no. 2, pp. 68-71, 2019.
 - [30] S. ustebay, Z. Turgut, and M. A. Aydin, "Intrusion Detection System with Recursive Feature Elimination by using Random Forest and Deep Learning Classifier," *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), ANKARA, Turkey, Turkey, IEEE*, 2018.
 - [31] R. Vinayakumar, M. Alazab, K. P. Soman *et al.*, "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE Access*, vol. 7, pp. 41525-41550, 2019.
 - [32] W. Wang, Y. Sheng, J. Wang *et al.*, "HAST-IDS: Learning Hierarchical Spatial-Temporal Features Using Deep Neural Networks to Improve Intrusion Detection," *IEEE Access*, vol. 6, pp. 1792-1806, 2018.
 - [33] G. Spathoulas, N. Giachoudis, G.-P. Damiris *et al.*, "Collaborative Blockchain-Based Detection of Distributed Denial of Service Attacks Based on Internet of Things Botnets," *Future Internet*, vol. 11, no. 11, pp. 226, 2019.
 - [34] F. Gont, *Security Assessment of the Transmission Control*

Protocol (TCP), United Kingdom: Centre for the Protection of National Infrastructure (CPNI), 2009.

- [35] F.-H. Hsu, Y.-L. Hwang, C.-Y. Tsai *et al.*, "TRAP: A Three-Way Handshake Server for TCP Connection Establishment," *Applied Sciences*, vol. 6, no. 11, pp. 358, 2016.
- [36] M. J. Evans, and J. r. S. Rosenthal, *Probability and Statistics, The Science of Uncertainty - Second Edition*, Toronto, 2009.

Authors' Profiles



Rheo Malani. Born in Samarinda, August 23, 1978. Completed undergraduate (S1) majoring in Computer Science at STIMIK WidyaCipthaDarma of Samarinda in 2003. Completed postgraduate study of Information System Department at Diponegoro University Semarang in 2013. Beginning in 2005 working as a lecturer in the Department of Information Technology, State Polytechnic of Samarinda until now.

In 2003 he was a programmer at the company IntSys Tech, and in 2008 worked as the head of IT in the organization of "PON 2008" East Kalimantan.

My Scopus ID : 57202205135

SINTA ID: 6024712

Research that has been published in SCOPUS is 2018:

- Image mosaicing by using random seeds generation based on fuzzy membership function
- Total asset prediction of the large Indonesian bank using adaptive artificial neural network back-propagation
- Modelling of contractor selection using fuzzy-TOPSIS
- Rainfall prediction using fuzzy inference system for preliminary micro-hydro power plant planning
- Comparison of Canny and Centroid on Face Recognition Process using Gray Level Cooccurrence Matrix and Probabilistic Neural Network
- Secured Data Transmission using Metadata Logger Manipulation Approach

Research that has been published in SCOPUS is 2019:

- Prediction of the Topographic Shape of the Ground Surface Using IDW Method through the Rectangular-Neighborhood Approach
- Optimization of the spatial interpolation based on the sliding neighborhood operation method by using K-mean clustering for predicting the topographic shape of the ground surface
- Assessment of the average level of TOEFL score by using SOM (Self organizing map) and K-mean clustering techniques

Areas of interest:

Computer Science, Computer Networks, Robotics & Artificial Intelligent.



Arief Bramanto Wicaksono Putra. Born in Balikpapan, January 20, 1983. Completed undergraduate (D4) majoring in Information Technology at Electronic Engineering Polytechnic Institute of Surabaya in 2006. Completed postgraduate study of Electrical Engineering Department at Brawijaya

University Malang in 2014. Beginning in 2008 working as a lecturer in the Department of Information Technology, State Polytechnic of Samarinda until now. His representative published articles list as follow: A Gray-Level Dynamic Range Modification Technique for Image Feature Extraction Using Fuzzy Membership Function (2018).Texture Feature Extraction based on Local Weighting Pattern (LWP) using Fuzzy Logic Approach (2018).The New Proposed Method For Texture Modification Of Close Up Face Image Based On Image Processing Using Local Weighting Pattern (LWP) With Enhancement Technique, Modeling of Time Series Data for Forecasting The Number of Foreign Tourists In East Kalimantan Using Fuzzy Inference System Based On ARX Model (2018, with SCOPUS Indexing). In 2019 with IEEE conference published as Prediction of The Topographic Shape of The Ground Surface Using IDW Method through The Rectangular-Neighborhood Approach, Feature-Based Video Frame Compression Using Adaptive Fuzzy Inference System.Steganography for Data Hiding in Digital Audio Data using Combined Least Significant Bit and 4-Wrap Length Method and Measurement of Electrical Power Usage Performance using Density Based Clustering Approach.

Areas of interest:

Computer Vision, Computer Networks, Robotics & Artificial Intelligent



Muhammad Rifani, Born in Samarinda, May 09, 1996. Completed diploma (D3) majoring in Computer Engineering at Information Technology, Samarinda State Polytechnic in 2017. Completed undergraduate (D4) Majoring Multimedia Informatics Engineering, at Information Technology Samarinda State Polytechnic in 2019. Beginning in 2018 working as a Laboratory Technician in the Department of Information Technology, Samarinda State Polytechnic until now.

Some Competency Certifications as follow : Network Tecnician, ECITB International Health and Safety Passport, Mikrotik Certified Network Associate (MTCNA), Certified Secure Computer User (CSCU).

Areas of interest:

Computer Networks, Data & Network Security, Robotics & Artificial Intelligent

How to cite this paper: Rheo Malani, Arief Bramanto Wicaksono Putra, Muhammad Rifani, "Implementation of the Naive Bayes Classifier Method for Potential Network Port Selection", *International Journal of Computer Network and Information Security(IJCNIS)*, Vol.12, No.2, pp.32-40, 2020. DOI: 10.5815/ijcnis.2020.02.04