Host Based Detection Approach using Time Based Module for Fast Attack Detection Behavior

Faizal M. A.[1], Mohd Zaki Mas'ud[2], Shahrin S.[3], Asrul Hadi Y.[4], Robiah Y.[5], Siti Rahayu S.[6]

Faculty of Information and Communication Technology
Univeristi Teknikal Malaysia, Karung Berkunci 1200, 75450 Ayer Keroh, Melaka.
Tel : 06-2332510, Fax : 06-2332508

[4]*Faculty of Information and Science Technology*
Multimedia University ,Jalan Ayer Keroh Lama,75450 Ayer Keroh, Melaka

{[1]faizalabdollah,[2] zaki.masud, [3]shahrinsahib, [5]robiah, [6]sitirahayu}@utem.edu.my, {[4]asrulhadi.yaacob}@mmu.edu.my

Abstract-***Intrusion Detection System (IDS) is an important component in a network security infrastructure. IDS need to be accurate and reliable in order to detect the intrusive behaviour of a packet that travelling through the network. With the current technological advancement attack on network infrastructure has evolve to a new level and to make IDS sensitive enough to detect the new attack, the detection framework need to be frequently updated. Both the fast attack and slow attack mechanism has become the subset of phases inside the anatomy of attack. Each of the attack mechanism has their own criteria and fast attack is the important type of attack that need to be considered as any late detection of the fast attack can cause a major bad impact to the organization. Therefore, there is a need to identify a suitable technique to detect the fast attack and based on this, this paper introduce a static threshold using statistical and observation technique for detecting the fast attack intrusion that is within one second time interval. The Threshold selected was based on the real network traffic dataset and verified using classification table on a real network traffic.***

Keyword: **Fast Attack, Threshold, Time Based Module**

I. INTRODUCTION

In the last decade, there has been a revolution in the wired and the wireless networking. The revolution also changes the attack mechanism to exploit the network infrastructure. The exploitation of the network has becoming quite alarming especially with the help of the freely available attack tools on the Internet [1]. With the availability of the attack tools, novice attacker can launched a sophisticated attack with just a little bit knowledge and as a result, the growth of the incident reported due to the security breach by NISER [2] has roughly parallel with the evolution of the Internet. Thus, it is necessary to protect vulnerable machines from being compromised. One method to secure the machine is by implementing security mechanism such as IDS, where it can reduce the possibilities of security breach from happening inside the organization [3]. Consequently the confidentiality, integrity and availability of the organization properties can be protected.

Understanding the anatomy of an attack is important before developing an IDS. An attack can be disserted into 5 phases which are reconnaissance, scanning, gaining access, maintaining access and covering tracks [4]. The first two are initial stages for the attacker getting information from the potential vulnerable machines. These phases can be categorized into two, which are fast attack and slow attack. Fast attack can be defined as an attack that uses a large amount of packet or connection within a few seconds [5, 6, 7]. Meanwhile the slow attack can be defined as an attack that takes a few minutes or a few hours to complete [8]. Detecting the fast attack is very useful to prevent any early attack on the network and may help to reduce the possibilities of further actions such as gaining access, maintaining access and covering tracks. Zhang and Leckie [9] also stated that there is a strong need to detect the intrusion activity such as scanning which can be classified as fast attack as quickly as possible inside the network. Unfortunately, fast attack detection technique required a suitable threshold mechanism to increase the detection accuracy. Selecting suitable technique still become a major issues need to be tackle by the IDS developer since it is widely used by the current IDS development [10, 11].

To overcome this challenge, this paper focuses on selecting a threshold mechanism based on one second time interval in detecting the fast attack. By introducing this threshold, it helps reducing the false alarm generated by the IDS and at the same time increase the accuracy of detection. The rest of this paper is organized as follows. Section 2 discusses the related work in detecting the fast attack intrusion. Section 3 discusses the methodology used to select the suitable threshold based on one second time interval. Next, Section 4 presents the result and analysis. Finally, Section 5 presents the conclusion and possible future extension of the work.

II. RELATED WORK

Intrusion detection can be divided into three types which are host based intrusion detection system, network based intrusion detection system and hybrid based intrusion detection system. Although the intrusion detection can be divided into three types, the main goal for each of them is the same which is intrusion detection. Intruder detection system is a system to detect attacks, or to classify them as unwanted authorized login, regardless of their success [12]. The detection method used by intrusion detection system can be classified as anomaly based detection and signature based detection. Signature-based IDS is also known as misuse detection approach IDS [13]. Signature based system is a system which contains a number of attack description or signatures that are matched against a stream of audit data looking for evidence of modelled attack [6]. The audit data can be gathered from network traffic or an application log. Meanwhile, the anomaly based system identifies the intrusion by identifying traffic or application which is presumed to be different from normal activity on the network or host [14]. Both of these approaches have their own disadvantages. False alarms generated by both systems are a major concern and is identified as a key issue and the cause of delay of further implementation of reactive intrusion detection system. Although both system have their own drawbacks, anomaly based detection has capabilities to recognize new attack inside the network without a need to update new rules [15]. This capability requires appropriate value of threshold to distinguish between the normal and abnormal behavior of the fast attack activity. Thus, introducing a fast attack threshold based on the observation technique to select an appropriate threshold is required to reduce the false alarm generated by the anomaly based detection for the fast attack detection.

There are two techniques that can be used in selecting the appropriate threshold to distinguish between the normal network traffic and abnormal network traffic. The techniques are static threshold value and dynamic threshold value. This research will focus on static threshold value because selecting static threshold is very useful to prevent the intrusion activity before the attacker begins to launch the attack [16]. In identifying the static threshold, there are multiple techniques used by previous researcher. Hussain et al, [17] used 60 connections per second from source IP address as one of the criteria to identify the intrusion. The selection of the threshold which is 60 connections per second was purely based on the observation. The detailed process of the observation is not clearly stated in this research. Furthermore, we argue that the selection of 60 connections per second can be further delay for the fast attack detection technique. Therefore, this research insists that selecting the suitable threshold within one second time interval may help to detect the intrusion activity as soon as possible. Kanlayasiri et al, [18] also used static threshold mechanism in identifying the portscan activity. The researcher used rule-based approach combined with static threshold to identify the attacker who launched the portscan attack. The threshold was set to 20 connections and does not use time interval to calculate the threshold. Therefore, the threshold value used the slow attack approach to identify the attack. Furthermore, the selection of 20 connections per second as a threshold is based on the observation and the process of the observation is not stated clearly. The researcher also suggests that the threshold selected can be adjusted manually. KDDCUP99 [19] has introduce time based feature to detect the fast attack intrusion activity. Unfortunately, the feature construct in KDDCUP99 used 2 second time interval where this research used one second time interval to detect the fast attack intrusion activity. The difference of one second between KDDCUP and proposed threshold give a good contribution in fast attack detection system.

Gates and Damon also used static threshold mechanism in identifying the attacker [20]. The researcher used mean and standard deviation from a normal data of a host to distinguish between the normal and abnormal data. The mean and standard deviation was computed using observation data for one week. The system will raise an alarm if the packet exceeds two standard deviation. This technique had a weakness due to the higher threshold value selected in this research. Therefore this research will introduce a threshold value based on one second time interval to detect the fast attack intrusion activity. It will use observation technique and classification table approach to validate the result. Furthermore the threshold selected will be tested using a various attack tool for further validation process. The next section will discuss about the methodology used for the observation technique and classification approach.

III. METHODOLOGY

A new framework for fast attack detection is necessary prior to detection of the fast attack intrusion can be done. For that reason, the framework together with a suitable fast attack detection module which consists of time based module has been developed [6, 7]. The time based detection module was based on one second time interval which is the main objective of the research. This research is based on one set of data which is Darpa99[21] and it was used in this project as a reference to identify the normal behaviour of the network traffic especially on selecting the normal threshold of the network traffic.

Fig. 1 illustrated the process flow of the threshold selection using the statistical and observation technique on real network traffic dataset. The threshold is computed based on the normal and abnormal network traffic captured from one

of the agencies and then verified using classification table. Based on the statistical and observation approach, we manage to conclude that maximum value that the normal behaviour of host connection is 3 connections per second. Therefore we select 4 connections per second as abnormal behaviour in detecting fast attack focusing on host based malicious activity. After the suitable threshold is choose, the result is verified using the classification table to prove that the threshold selected is suitable in detecting the fast attack intrusion activity. The next section will discuss about the result and analysis of the threshold selection.

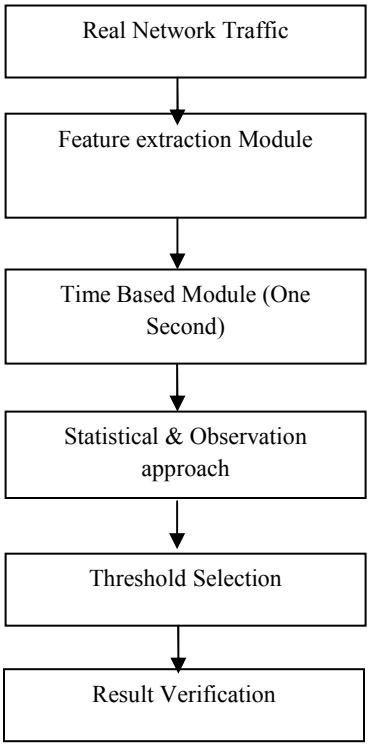

Figure 1: Threshold selection technique

Logistic Regression Model

Binary logistic regression is a form of regression which is used when the dependent variable is a dichotomy and the independents are any type [22]. The logistic regression can be used in order to achieve one of the two main objectives which are explanation or prediction. The explanation will reveal the significance of the variable in predicting the outcome variable. Meanwhile, in the prediction perspective, it will predict the normal or abnormal behavior of the event. In this research, we will adapt both the objectives which are explanation or prediction. Using the explanation objective, the influence of the feature in detecting the fast attack can be revealed. Meanwhile, the prediction objective will focus on identifying the appropriate threshold for detecting the fast attack.

Before selecting the suitable threshold from the logistic regression model, identifying the fitting of the model in necessary. The purpose of determining the fit of the logistic model is to assess how effectively the model describes the outcome variable [23]. If the model fits, then a conclusion can be made that the model is suitable and is good in predicting the outcome variable. Therefore the accuracy of the detection also becomes higher. There are two tests can be used to assess the fit of the model. The first test is called likelihood ratio test [22]. The likelihood ratio test can be used on two purposes which are to test the fit of the model and to test the contribution of the individual predictor. The likelihood ratio test is also called model chi-square test. The model chi-square test will test the difference between -2LL (-2 Log Likelihood) for the full model and -2LL for the initial chi-square in the null model. The null model is also called the initial model which involves only the constant.

Besides using the likelihood ratio test, the percentage of correct prediction is used to assess the model. The percentage of the correct prediction can be interpreted by using a classification table. The classification table is the most appropriate test if the test objective is based on the classification [23]. Therefore the classification table is chosen as one of the test used to assess the model. Using the classification table, the percentage of the detection attack rate and detection normal rate can be calculated. Furthermore, error rate can be calculated in the classification table also. The error rate of the classification table can be divided into two categories which are false positive and false negative. False positive means that the number of errors in which a normal event is considered as an attack event. Meanwhile, false negative means the number of errors in which the attack event is predicted to be normal, but is in fact an attack. Below is an example of the classification table. The calculation of the detection rate and error rate is also shown below.

TABLE I
Example of the Classification Table

Classified		**Predicted**	
		Normal	**Attack**
Observed	**Normal**	a	b
	Attack	c	d

Detection Attack rate = d / (c + d)

False Positive = b / (b + d)

Detection Normal rate = a / (a + b)

False Negative = c / (a + c)

Overall Detection rate = (a + d) / (a + b + c + d)

The above calculation will be used in the next subsection for assessing the model based on the classification table. The detailed explanation on assessing the model is presented below.

IV. RESULT AND ANALYSIS

The approach in this research is verified using real time network traffic captured from one of an agency in Malaysia. From the network traffics there are 105 connections has been declare as normal connection while 108 connection is an abnormal connection. The result analysis is based on the classification table generated by the logistic regression model. The model also suggested the 3 connection per second as a suitable threshold for detecting fast attack intrusion activity focusing on host based attack. For the Classification table, there are two model involve which are null model and full model. The Null Model is a model which has only a constant value without any mechanism to distinguish between the attack and normal connection. Meanwhile the Full Model is model generated after the predictor is involved in the detection. The predictors include focus on the number of connection based on one second time interval. Table I and II show the result of the Null Model and Full Model.

Table II, shows that the null model managed to predict 50.7% correctly in classifying the overall percentage but the false positive was also very high at 49.3%. Thus, the model was useless in detecting the normal network traffic and will affect the network security in the organization in terms of massive log generated from the model. The massive log gave extra burden to security administrator in verifying the intrusion activity.

TABLE II

Null Model

Classified		Predicted	
		Normal	**Attack**
Observed	**Normal**	0	105
	Attack	0	108

Detection Attack rate = 0%, False Negative = 6.4%, Detection Normal rate = 100%, False Positive = 49.3%, Overall Detection rate = 50.7%

TABLE III

Full Model

Classified		Predicted	
		Normal	**Attack**
Observed	**Normal**	104	1
	Attack	2	106

Detection Attack rate = 98.1%, False Negative = 1.9%, Detection Normal rate = 99%, False Positive = 0.9%, Overall Detection rate = 98.6%

Moreover, the model did not have capabilities to detect the normal event because using constant, the model assumed most of the data were attack. After the predictor was included inside the model, the detection accuracy was high and reduced the false alarm as depicted in table III. The model had capabilities to predict 98.1% correctly in classifying the attack and only 1.9% false negative generated from the full model. Meanwhile for the normal data, the full model was able to predict 99% correctly and false positive is 0.9%. The model had better prediction and had capabilities to distinguish the difference between the attack and normal traffic. Furthermore, the overall percentage of the classification table for the null model was 50.7 %. The result of the overall percentage increased to 98.6 % after the full logistic regression model was applied to the data. As a conclusion, the increase in of the correct percentage for the classification between the attack and normal indicate that the model is suitable, fits and good in predicting the normal and abnormal behavior. Therefore, it validate the result that the model generated from the logistic regression suggested that 3 connection per second can be used to detect the intrusion behaviour especially for fast attack detection system.

V. CONCLUSION AND FUTURE WORK

In this research we manage to select the suitable threshold for detection host based attack for fast attack intrusion activity. Although the selection of the threshold was based on the observation technique but the classification table approach have proven the threshold can be used to detect the intrusion behaviour. Additionally the detail process of the observation technique also has been reveal which most of the previous research do not insist to do so.

For future work, the researcher would like to identify the new technique to select the suitable threshold for fast attack detection system. The future technique will be implemented on real network traffic from various sites to validate the result. In addition, the timeliness parameter also will be assessed to identify the accurate threshold for fast attack detection.

ACKNOWLEDGMENT

This research has been supported by Universiti Teknikal Malaysia (UTeM) Melaka and Malaysia Government under FRGS Fund.

REFERENCES

[1] McHugh J., Christie A., Allen J. (2000). *"Defending Yourself: he Role of Intrusion Detection System*". In Proceeding of IEEE, Software, 2000.

[2] Niser,(2008). http://www.niser.gov.my

[3] Microsoft, Ruth, A. & Hudson, K. (2003). *Security + Certification: CompTIA Exam SYO-101*. USA. Microsoft Press.

[4] Module for CEH (2009)

[5] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur and J. Srivastava, , " A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection", *SIAM International Conference on Data Mining*, 2003

[6]Faizal MA., Asrul HY., Shahrin S. 2007. "An Earlier Detection Framework for Network Intrusion Detection System". *Proceeding of the Second International Conference on Advances in Information Technology, Bangkok, 1 – 2 November 2007*.

[7] Mohd Faizal Abdollah, Asrul Hadi Yaacob & Shahrin Sahib. 2007. Improved Fast Attack Detection Model for Network Intrusion Detection. *Proceeding of International Conference on Engineering and ICT, UTeM.*

[8] Wenke Lee. 1999. A Data Mining Framework for Constructing Feature and Model for Intrusion Detection System. PhD thesis University of Columbia.

[9] Zhang, D & Leckie, C. (2006). An Evaluation Technique for Network Intrusion Detection Systems. *In Proceeding of the First International Conference on Scalable* Information Systems, Hong-Kong, June 2006.

[10] Bro. (2009). http://www.bro-ids.org.

[11] Snort. (2009). http://www.snort.org.

[12] Allen, J., Christie, A., Fithen, W., Mc Hugh, J., Pickel, J. & Stoner, E. (2000). State of the Practice on Intrusion Detection Technologies. *Technical Report on Networked Systems Survivability Program.* University of Carnegie Mellon, Pittsburgh, USA.

[13] Karl Levitt, "Intrusion Detection: Current Capabilities and Future Directions", *In Proceeding of the 18th Annual Computer Security Applications Conference, IEEE*, 2002

[14] Wang Y., Huang GX. & Peng DG. "Model of Network Intrusion Detection System Based on BP Algorithm". *Proceeding of IEEE Conference on Industrial Electronics and Applications*, IEEE, 2006

[15] Gaurav Tandon & Philip K. Chan. (2007). Weighting versus Pruning in Rule Validation for Detecting Network and Host Anomalies. *In Proceeding of KDD 07 Conference, ACM USA.*

[16] Idika, N. & Mathur P. A. (2007). A Survey of Malware Detection Technique. *In Proceeding of Software Engineering Research Center Conference, SERC-TR286*.

[17] Hussain A., Heidermann, J. and Papadopoulos, C. (2003). A Framework for Classifying Denial of Service Attacks. *In Proceeding of 2003 ACM SIGCOMM, Germany, 2003*.

[18] Kanlayasiri, U., Sanguanpong, S. & Jaratmanachot, W. (2000). A Rule Based Approach for Port Scanning. *In Proceeding of Electrical Engineering Conference*. Thailand.

[19] KDDCUP99 dataset. (2009). http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[20] Gates, C & Damon, B, Cpt. (2005). Host Anomalies from Network Data. In *Proceeding from the Sixth Annual IEEE SMC, 2005.*

[21] Darpa99. (2009). http://www.ll.mit.edu/

[22] Andy Field. (2005). Discovering Statistic Using SPSS, *2nd edn, Sage Publication London Schuyler W.Huck*

[23] Hosmer D.W and Stanley, L. (2000). Applied Logistic Regression Second Edition. *USA. John Wiley and Son Inc*