Journal of Theoretical and Applied Information Technology
31st July 2018. Vol.96. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THRESHOLD IDENTIFICATION FOR HTTP BOTNET DETECTION

[1]NUR HIDAYAH M. S, [1]FAIZAL M. A, [2]WAN AHMAD RAMZI W. Y, [1]RUDY FADHLEE M. D

[1]Department of System and Computer Communication, Faculty of Information and Communications Technology, Universiti Teknikal Malaysia Melaka (UTeM).

[2]Department of Computer System and Electrical Technology, Masjid Tanah Community College, Melaka.

E-mail: [1]nurhidayahmohdsaudi@gmail.com, [1]faizalabdollah@utem.edu.my, [2]ramzi016@yahoo.com, [1]rudyfadhlee@gmail.com

ABSTRACT

Over the past years, botnets have gained the attention of researchers worldwide. A lot of effort has been given to detect the presence of a botnet. Many researchers focus on developing the systems and compare the detection method to detect the botnet activity. Identifying an appropriate threshold value is essential in order to differentiate between normal and abnormal network traffic. The suitable value of the threshold can minimize false positive rate botnet activity. Therefore, in this paper, we will identify the appropriate static value of the threshold for detecting HTTP botnet. The likelihood ratio tests and classification table were two test that will be used in order to access the fit of the model. The comparative analysis with another researcher also has been conducted. The result found showed about 95% of the data are declared as an attack when the sample of data has been compared with the value of the threshold. Thus, the value of the threshold is acceptable discrimination to use in detecting HTTP botnet activity.

Keywords: *Threshold, Malware, Botnet, HTTP Botnet, Logistic Regression*

1. INTRODUCTION

These recent years have witnessed an annual increase in the incidents of cyber-attacks on the Internet. Most of the attacks include emails spamming, distribution denial-of-service attacks and theft credential from the victim's computer. All these attacks usually might contribute to serious disasters and breach the computer security policies such as Confidentiality, Integrity, and Availability [1]. Besides, with various computer-processed device platforms, cybercriminals will have various choices in strategizing the attack and resulting in complexity to oppose the crime. HTTP botnet is considered by security organizations as the biggest threat since the attack is based on HTTP protocol which is widely used to open a website. Usually, HTTP botnets use a centralized C&C where a single C&C gives an order to the network of bots [2] as shown in Figure 1. HTTP botnets conceal their C&C connection in the HTTP traffic and are transmitted over the Internet by emulating the behaviors of authorized Web connection [3]. In addition, [4] have described that an HTTP bot is grouped to communicate with a certain web server using an HTTP post, which contains exclusive identifiers for the botnet, and in response, the web server will conduct the HTTP commands that it has been set up by. Thus, due to the complexity of the attack, the botnets were thoroughly examined and ways to detect them in network traffic, especially when using the HTTP protocol were studied.

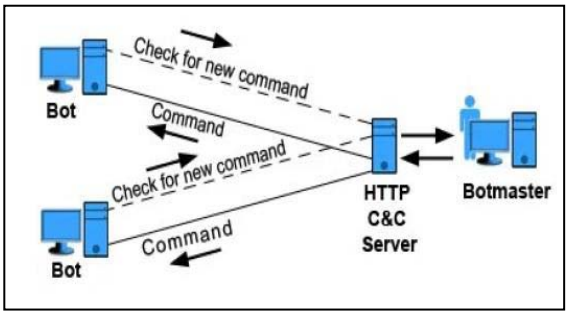

Figure 1. HTTP Botnet [5]

Moreover, the effective analysis of botnet detection system is a key element to the life cycle of botnets [6]. According to [7], a botnet lifecycle is comprised of six phases as shown in Figure 2. The primary phase is initial infection where attackers attempt to infect the target PCs in various misuse strategies to distribute new bot clients. For example,

Journal of Theoretical and Applied Information Technology
31st July 2018. Vol.96. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

sending an email with malware connections are prompt to a misuse program. Then in a second injection, when email attachments are open by the unsuspicious user, the contaminated PC will download bot copies through the hypertext transfer protocol (HTTP), peer-to-peer (P2P), or file transfer protocol (FTP) from remote servers and spontaneously install to an exploited mechanism. This mechanism then changes into a "zombie" and runs the malicious code.

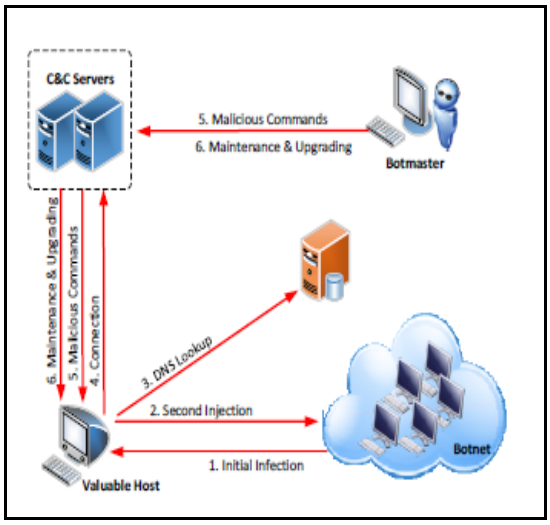

Figure 2. Botnet Life Cycle [4]

Following from that, in DNS lookup, new bot clients need to interface with C&C servers in the wake of turning out to be genuine bots. Generally, the name of C&C and DNS server needs to register by Botmaster in order to prevent from being known. Meanwhile in the association stage otherwise called "Mobilizing", if the bot clients need to deliver information about zombie mechanism and acquire to gain updates, they need to associate with C&C servers. The zombie mechanism then turns into a part of the botnet army. After the connection phase, the actual botnet command will be started. Bot clients wait for the instructions which will be sent by the botmaster and the harmful program will be implemented then execute to attack the victim's machine when bot accepts the commands. The last phases are maintenance and update which it is needed to keep the bots lively and maintained. Botmaster could improve malware codes to repair any bugs in order to enhance the performance of their bot programs to be more intelligent. Other than that, by changing the pattern of the harmful program from time period to random or changing the C&C server's addresses, the current detection techniques can also be avoided.

Additionally, botnets detections network utilises behavior-based detection and it can be divided into two categories using anomaly-based and signature-based approaches. In this paper, anomaly-based detection will be employed in the experiment. It is because of the difficulty of discovering the unique communication patterns in the network traffic which do not imitate the signature-based approaches [9]. Furthermore, this technique also detects the botnets using network abnormalities such as high network latency, traffic on unfamiliar ports, high volumes of traffic, and abnormal system behavior that might indicate the existence of botnets activity in the network. Besides, anomaly-based analysis possessed the capability to detect botnets and even novel attacks [10]. According to [11], an anomaly-based approach has difficulties in determining the value of threshold due to incomplete profile from behavior which can lead to the false alarm.

However, the existing problem of botnet detection is the difficulty in identifying an appropriate value of threshold to distinguish between normal and abnormal network traffic. As a result, a new technique to identify the value of the threshold is necessary, especially for the detection of a botnet attack. This statement motivated by [12] which stated that the appropriate value of threshold to minimize the false positive still becomes an issue which needs to be solved. Setting an inappropriate value of the threshold will generate the false alarm of the botnet activity. The author [8], claimed that identifying a good threshold can minimize the false positive rate. Hence, a new method to find the threshold value is required in order to reduce incorrect alarm produced by the anomaly-based detection for botnet recognition.

The remainder of this paper is presented as follows: Section 2 discusses related studies and Section 3 explain the methodology used for this paper. Section 4 presents some analyses of the results. Section 5 concludes the paper and suggests future work directions.

2. RELATED WORK

The main objective of discovering the value of threshold in this project is to differentiate between normal and abnormal activity present in the network. There are several studies that have been previously conducted that are related to the threshold. In this research, [13] utilized static threshold mechanism as a part of distinguishing the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

port scan movement to recognize the attacker. He recommended that the selected threshold can be manually adjusted. Additionally, [14] also applied static threshold mechanism to recognize the attacker. Standard deviation and mean from typical records of the host have been used by this researcher to differentiate between the common and uncommon data. However, [15] suggested the threshold alert notice malicious activity in the network traffic. The threshold alert is used to distinguish between benign traffic and malware. Moreover, [16] deliberate that if the sum of payloads is below the value threshold of 2 KB, then suspicion should be raised. The author [17] also discovered the universal entropy of packet sizes by defining the smaller result in packet size have a higher tendency to be attacked. He stated that the attack is detected when the distance between the probability distribution of packet sizes is greater than the value of the threshold. In contrast with the author, [18] studies the detection of intrusion at the host or network by using log analysis. He clusters the log events and uses a filtering threshold to decrease the size of events for examination. The experiment outcome of this author shows filtering threshold significantly impacts the result of identifying the anomalies at the network or host with the rate of detection is about 87.26% and 85.24% of anomalous events.

In addition, the researcher [19] examines network attacks by using rank distribution data. The determination of threshold values for major network variables based on the collected data of rank distribution under normal network condition. When the threshold increase, it shows that the identification of attacking IP addresses and subsequent blocking of their access. Other than that, author [20] defines a significant value of the threshold for botnet identification. Determination of the threshold value can discover the unknown properties of the normal traffic patterns. The result from his research show that the value of the threshold is set at 0.2 with an average percentage of correctly recognized bots is moderately large (> 80%). Nevertheless, author [21] suggested the method of structural analysis-based learning to categorize between the botnet and benign application. The research used machine learning method in order to achieve high detection of accuracy. The result shows the value of the threshold is set to 0.05 as acceptance value to detect botnet application. Conversely, author [22] reported that 0.9 is the optimum value for the threshold to distinguish between benign and botnet. His research has proven that with the value of the threshold, the detection of zero-day fast-flux botnets can be recognized.

Briefly, from the aforementioned related works were significant in using the threshold as a method to find anomalies activity. In recent times, various alternative techniques have been proposed by the researcher in distinguishing botnet detection. Nonetheless, the method still lacks in distinguish the behaviors of malware and affect the rate of false negatives. Therefore, this research is focused on identifying the value of static threshold in detecting HTTP botnet. Then, the value of static threshold will be tested and validated with several samples of data in order to establish its reliability. This research is supported by the author [38] which pointed that a proper identification threshold is required for botnet detection. The value of the threshold also may assist in detecting an intruder and recognize malicious activity in the network system.

3. RESEARCH METHOD

Figure 3 illustrates the process of threshold selection. This process was used to detect the botnet. A collection of normal and botnet dataset has been collected and going through data preprocessing. About 57 feature of data was analyzed by using feature selection to select the influence features (7 of the feature). Then the influence feature will testing by using a Likelihood Ratio test and classification table. The result of the logistic regression equation was analyzed to identify the fit of the models. Then, the probability graph is generated based on the selected model which can identify the appropriate value of threshold in the detection of the botnet. When the event goes beyond the number of thresholds that have been declared in the probability graph, then the system will generate an alert that there was botnet attack in the network. Thus, selecting a suitable value of the threshold is important in order to detect a botnet attack.

Journal of Theoretical and Applied Information Technology
31st July 2018. Vol.96. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

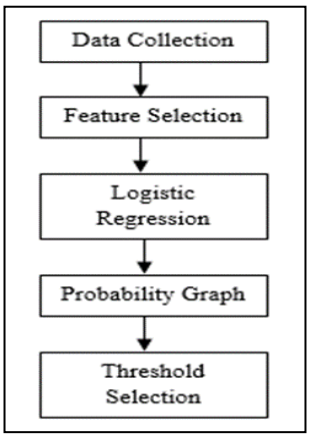

Figure 3. Process of Threshold Selection

3.1 Feature Selection

Feature selection is the technique of choosing the relevant features from dataset to provide the good prediction results, reduce the cost of computational, and improved the interpretability of the model. Feature selection has been applying in many practitioners for reducing dimensionality by aiming at a subset of relevant features from the original based on specific criteria [23]. According to [24], feature selection is a method that affects the most in detection among numerous data by eliminate the redundant, reduce effects from noise and irrelevant features. Furthermore, author [25] stated that the objective of feature selection is to increase the accuracy of the model in term of reducing the complexity data for learning and setup.

Besides, filter model, wrapper model, and embedded model are three categories of features selection methods. The filter method based on reliable features of training dataset with an independence of any predictor to select the best features [23]. The filter model depends on specific measures such as consistency, dependency, and correlation. Contrast with the wrapper model where the process of feature selection involves optimized classifiers to obtain the set of features for improving the classification model performance [26]. This model will repeat the process until the high accuracy or performance is achieved. Moreover, the embedded model is the combination of filter model and wrapper model where the model learn and identify significant features that contribute to the accuracy of the model [27]. Thus, among three methods of feature selection, the wrapper model is chosen as it achieves better recognition rate and avoids overfitting since the model used the cross-validation measured of predictive accuracy [28]. This statement is supported by [29] which claim that the wrapper model as it can handle large dimensional data and it uses independent subset evaluation. Figure 4 shows the process of the wrapper feature selection model.

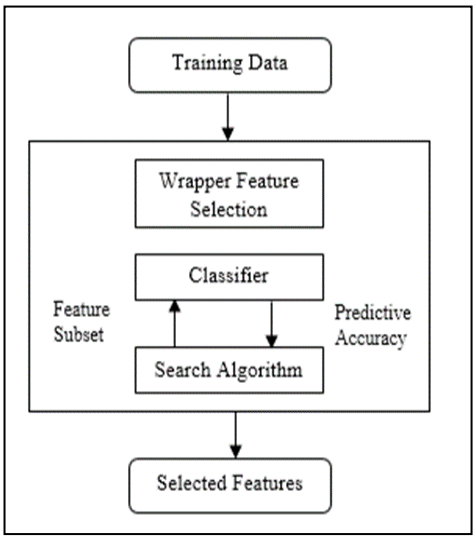

Figure 4. Process of Wrapper Feature Selection Model [26]

Additionally, feature selection also can be done by the heuristic method such as forward selection, backward elimination, and optimized selection. Table 1 shows the heuristic methods for feature selection. From Table 1, it concludes that forward selection is the best option as the selected attribute provide the maximum accuracy to the model compared with backward elimination, the selected attribute giving the minimum accuracy to the model. For that reason, forward selection will be used in this research in order to obtain high accuracy with the significant feature. The forward selection algorithm is shown in Figure 5.

Table 1. The Heuristic Method for Feature Selection

Method	Content
Forward Selection	• Start with empty selection attribute. • The performance estimate using cross-validation for each added attribute.
Backward Elimination	• Start with full selection attribute. • The performance estimate using cross-validation for each removal attribute.
Optimized Selection	• Select significant attributes. • Optimize and search problem.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Initialize feature set $F^k = \emptyset$ at $k = 0$
Iterate
- Find the best feature j to add to F^k with most significant cost reduction
- k++ and $F^k = F^{k-1} \cup \{j\}$

Figure 5. Forward Selection Algorithm [28]

3.2 Logistic Regression

A kind of development which is used when the reliant variable is a division and the independents are any sort is called binary logistic regression. Logistic regression is utilized to foresee the likelihood of the dichotomous event. According to Discovering Statistics Using SPSS, logistic regression refers to various regressions with an outcome variable that is a categorical variable while the predictor variables are continuous or categorical [30]. Besides, [31] stated that logistic regression is convenient as any value from negative infinity to positive infinity, it can take as an input, whereas the values between 0 and 1 as an output.

In addition, the advantages of logistic regression include flexibility and the ability to apply logistic regression to many subject areas. This statement is supported by [32] which studies the applicability of logistic regression to calculate the probability that a packet contains malware. Logistic regression can replace all the signatures which are related to a single malware family with the same accuracy as signature detection. Thus, it is an appropriate alternate to discriminant analysis as it does not require strict expectations, whether normality and equality.

Moreover, recognizing the fit of the model is fundamental before selecting an appropriate threshold from the graph of probability which is made from the logistic regression model. The reason is to assess the efficiency of the model in defining the outcome variable. When the model fits, the model will then give a decent effect to the model in expecting the result. The precision of the detection will also become higher. The likelihood ratio test (1) also known as chi-square test model and the rate of correct prediction was the two tests used to measure the fit of the model [30].

$$x^2 = 2 [\text{Log Likelihood(New(with predictor)} - \text{Log Likelihood(Baseline(without predictor))} \quad (1)$$

However, the rate of the correct prediction can obtained by using classification table. The rate of the attack detection and normal detection is based on the classification table which is comprised of false positive and false negative. False positive is represents as non-malicious which is it misclassified as attack. Meanwhile, false negative is represents as attack but it is misclassified as normal. Table 2 depicts an example of the classification table.

Table 2. Classification Table [33]

Classified		Predicted	
		Normal	Attack
Observed	Normal	A	B
	Attack	C	D

From the table above, it can be concluded that:

i) Detection Attack Rate = d / (c + d)
ii) False Positive (FP) = b / (b + d)
iii) Detection Normal Rate = a / (a + b)
iv) False Negative (FN) = c / (a + c)
v) Overall Detection Rate = (a + d) / (a + b + c + d)

3.3 Threshold Identification

The estimated probability of the logistic model is the basis of the threshold identification. By using equation (2), the regression equation of the model can be computed.

$$P(Y) = e^{a+bx} / (1 + e^{a+bx}) \quad (2)$$

The regression model used the cut-off value from the Receiver Operating Characteristic curve (ROC). The simplification of the set of possible combinations of sensitivity and specificity possible for predictors known as ROC curve [34]. In this paper, the cut-off value for the regression model in detecting the attack was 0.8 or a probability of 80%. This cut-off value is based on the assumption that in order to eliminate any bias caused by the attack or normal network traffic which may reflect the accuracy of the result. Furthermore, according to [35], the selected cut-off value probability of 80% is considered as an acceptable discrimination.

4. RESULTS AND DISCUSSION

Based on previous study [36], there are seven features that are involved in botnet detection which are avg_segm_size_b2a, initial_window_bytes_a2b, unique_bytes_sent_b2a,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

max_win_adv_a2b, max_win_adv_b2a, min_segm_size_a2b and max_segm_size_a2b. All these features are analyzed by using Log likelihood test and Wald test. Then, the results were compared and discussed by using the statistical values of Log likelihood test and Wald test.

From Table 3, it can be concluded that only three features gave a decent effect on the model for expecting the result. The features are avg_segm_size_b2a, initial_window_bytes_a2b, and min_segm_size_a2b as the value of Wald test is different from zero. Meanwhile, the other four features which are unique_bytes_sent_b2a, max_win_adv_a2b, max_win_adv_b2a, and max_segm_size_a2b are not selected. This is because the result of Wald test is significantly from zero, which means that the features selected did not give a decent effect to the model in expecting the result. Therefore, only avg_segm_size_b2a, initial_window_bytes_a2b, and min_segm_size_a2b feature with the value 2173.349, 7445.696, and 13961.988 can be used in identifying the threshold selection. From that, the probability graph will be produced in order to determine the appropriate value of threshold in botnet detection.

Table 3. Result of Features Influence

Features	Wald test	-2 Log Likelihood
Avg_segm_size_b2a	2173.349	402052.506
Initial_window_bytes_a2b	7445.696	400666.766
Unique_bytes_sent_b2a	0.322	400651.702
Max_win_adv_a2b	0.854	398849.06
Min_segm_size_a2b	13961.988	378687.144
Max_segm_size_a2b	0.124	378445.002
Max_win_adv_b2a	0.354	377280.474

4.1 Classification Table

The fit of the model can be assessed by using the classification table. Table 4 and Table 5 show the results of the classification table of the null model and the full model.

Table 4. Classification of Null Model

Observed		Predicted Class	
		Normal	Botnet
Class	Normal	60069	0
	Botnet	613398	0

Table 5. Classification of Full Model

Observed		Predicted Class	
		Normal	Botnet
Class	Normal	3881	56188
	Botnet	472	612926

Table 4 demonstrates that the finding normal rate of the model is 100% accurate in categorizing the normal while the false negative was also very high which is 91.08%. This showed that the organization was very dangerous because many attacks were not discovered. Inappropriately, the model assumed most of the data were normal when using constant, which shows the model also did not have the abilities to identify the attack. Then, after the predictor was incorporated into the model, the accuracy of the detection attack became high and the false positive reduced as represented in Table 5. The detection attack rate of the model is 99.92% accurate in categorizing the attack and only 8.39% is a false positive. The false negative was reduced to 80.82% from the full model. Although the attack recognition rate was only 99.92%, it was still satisfying as the current botnet detection system has 80% of the abilities to distinguish the botnet [37]. The model is capable of differentiating the classification of normal and attack since it has the better expectation. Besides, for null model, the total percentage of the organization table was 8.92%. The result of the overall percentage increased to 91.59% after the full logistic regression model was applied to the data. Thus, the model was appropriate, fits, and it is suitable for expecting the outcome variable since it indicates an increase in the correct percentage for the classification between the attack and standard.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.2 Threshold Identification

4.2.1 Avg_segm_size_b2a Feature

Figure 6 shows the graph of the threshold for avg_segm_size_b2a feature that was created the logistic regression model. The cut-off value of the logistic probability model is 0.8, thus, the threshold was 1.6. Therefore, the value of 2 avg_segm_size_b2a per bytes can be set as an attack inside the real-time system.

The fitted logistic regression equation was computed as follows:

$$P(Y) = e^{-5.415+3.391x} / (1 + e^{-5.415+3.391x})$$

4.2.2 Initial_window_bytes_a2b Feature

The graph shown in Figure 7 was produced by a fitted logistic regression equation for the initial_window_bytes_a2b feature. The logistic regression equation that computed the threshold is:

$$P(Y) = e^{-3.407+3.222x} / (1 + e^{-3.407+3.222x})$$

4.2.3 Min_segm_size_a2b Feature

Figure 8 shows the graph generated from fitted logistic regression for the min_segm_size_a2b feature. The cut-off value of 0.8 of threshold was 1.6. Thus, the value of 2 min_segm_size_a2b per bytes can be set as an attack traffic in the real time system. The logistic equation which generates the graph is illustrated as follows:

$$P(Y) = e^{-3.454+3.407x} / (1 + e^{-3.454+3.407x})$$

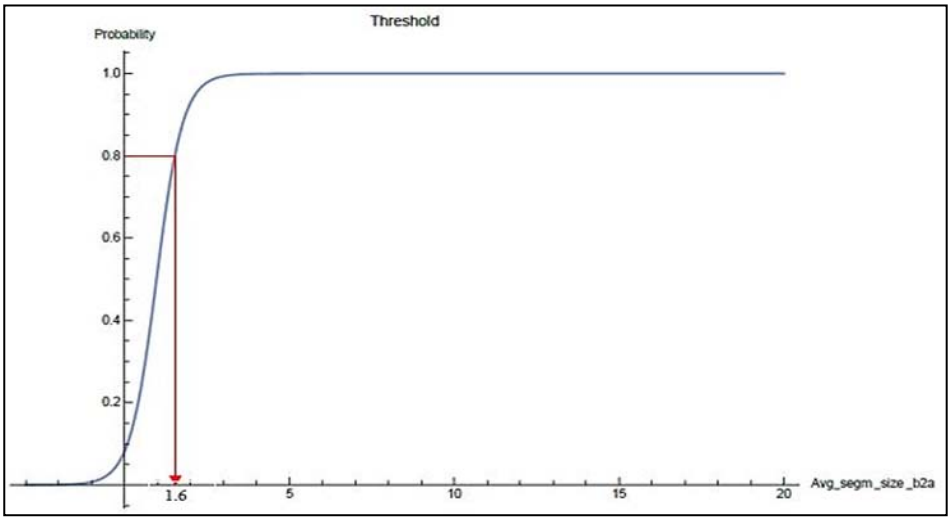

Figure 6. Threshold of the Avg_segm_size_b2a Feature

Journal of Theoretical and Applied Information Technology
31st July 2018. Vol.96. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

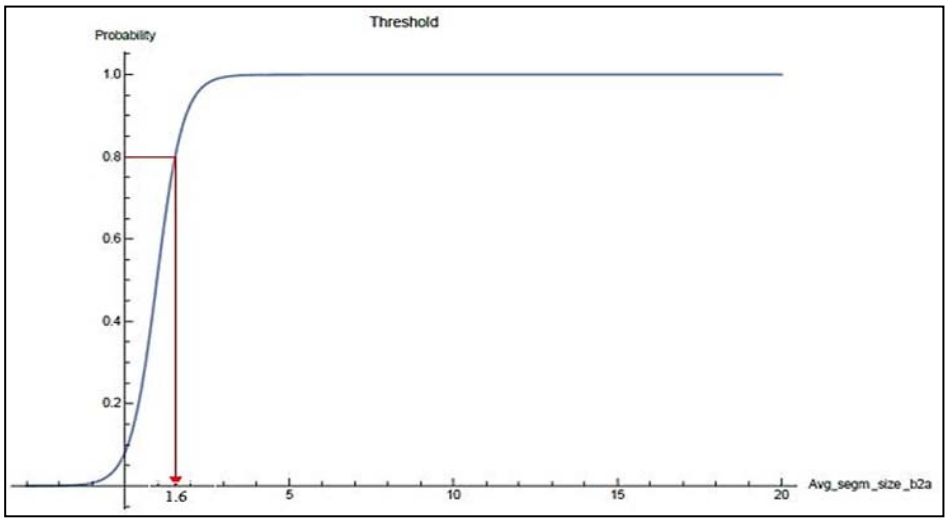

Figure 7. Threshold of the Initial_window_bytes_a2b Feature

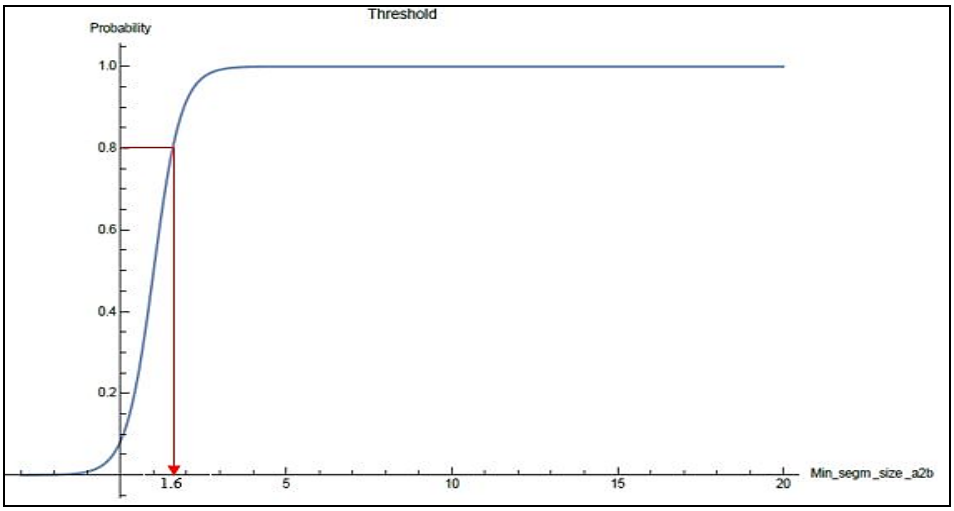

Figure 8. Threshold of the Min_segm_size_a2b Feature

Referring to the result above, only three significant feature was selected from the seven feature in order to identify the value of threshold in detecting botnet activity. This three feature gave the best result in distinguishing botnet detection as the value of Wald test is greater than zero. The value Wald test give the significant commitment to the feature in anticipating the better result. The selected feature was used to generate the threshold graph by using the probability of logistic model. When the event goes beyond the number of thresholds that have been declared in the probability graph, then the system will generate an alert that there was botnet attack in the network. Thus, the obtained value of threshold will be used to distinguish botnet attack and more importantly, it can help to reduce false alarm.

In addition, the testing is conducted in order to show whether the selected threshold is able to detect a botnet attack. The result was evaluated based on the accuracy detection rate criteria. Figure 9 shows the testing and result of validation procedure. Then, the results were compared to the value of threshold in the probability graph that was generated based on the selected feature influence in botnet detection. For this project, the value of overall threshold was 2 per bytes and the result for testing achieved approximately 95% accuracy of detection rate. So, if the value of equal or greater than 2 per bytes, it shows the presence of botnet attack in the network otherwise there is normal activity. Therefore, it is concluded that the value of threshold which is 2 per bytes in this project can be used to detect a botnet attack as the testing result

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

provide the 95% of the capabilities to distinguish the botnet.

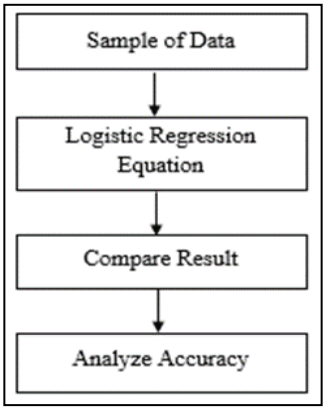

Figure 9. Result Validation Procedure

5. OPEN RESEARCH ISSUE

Selecting a significant feature and an appropriate value of the threshold is not a simple task to be accomplished. The inadequate data and irrelevant feature will affect the result of threshold value in distinguish botnet activity. For instance, the value of threshold for previous work cannot be used in this research as the study not focus on HTTP botnet detection. The value of threshold should minimize the rate of false alarm and consequently the rate of accuracy will be increase with the improved false alarm rate. Besides, the existing works proposed the value of threshold to differentiate between benign and malware activity not on HTTP botnet detection. Nowadays, the type of botnet attack has undergone significant changes and difficult to be identified as the HTTP botnets hide their communication through HTTP traffic. This remains as an open challenge in the research community. Moreover, constraints on botnet detection (i.e. cannot differentiate and recognize the new botnet activity precisely) need to be improved.

6. CONCLUSION

This paper emphasis on the using of the threshold to detect anomalies activity. The difficulty in identifying an appropriate the value of threshold to distinguish between normal and abnormal network traffic is the current problem in the detection of HTTP botnet. Identifying a good threshold can minimize the false positive rate. A low threshold may generate many false alarms while a higher threshold may miss botnet attack detection. Thus, selecting a suitable threshold is an important focus of this paper. The main contribution of this work is the three feature such as Avg_segm_size_b2a, Initial_window_bytes_a2b, and Min_segm_size_a2b with the value of threshold 2 per bytes can be used to detect botnet. The limitation of this study is the feature extracted from TCP headers features and only use six variant of botnets. For future works, it is recommended for future studies implement a dynamic technique in order to identify the value of the threshold for detecting botnet activity.

ACKNOWLEDGEMENTS

This work has been supported under Universiti Teknikal Malysia Melaka research grant Gluar/CSM/2016/FTMK-CACT/100013 and KPT MyBrain15. The authors would like to thank to Universiti Teknikal Malaysia Melaka and all members of INSFORNET research group for their incredible supports in this project.

REFERENCES

[1] Liao HJ, Lin CH, Lin YC, Tung KY. "Intrusion detection system: A comprehensive review", *Journal of Network and Computer Applications*, Vol. 36, Issue 1, pp. 16-24, 2013.

[2] Tyagi, R., Paul, T., Manoj, B.S. and Thanudas, B. "A novel HTTP botnet traffic detection method", *In: India Conference (INDICON), 2015 Annual IEEE*, 17-20 Dec. 2015, New Delhi, India: IEEE. pp. 1-6, 2015.

[3] Sakib, M.N. and Huang, C.T. "Using anomaly detection based techniques to detect HTTP-based botnet C&C traffic", *In: Communications (ICC), 2016 IEEE International Conferenc*e, 22-27 May 2016, Kuala Lumpur, Malaysia: IEEE. pp. 1-6.

[4] Hsu, F.H., Ou, C.W., Hwang, Y.L., Chang, Y.C. and Lin, P.C. "Detecting Web-Based Botnets Using Bot Communication Traffic Features". *Security and Communication Networks*, *2017*.

[5] P. Pierluigi, 2013 "HTTP-Botnets: The Dark Side of a Standard Protocol!" Retrieved from https://securityaffairs.co/wordpress/13747/cyber-crime/http-botnets-the-dark-side-of-an-standard-protocol.html [Accessed on March 8, 2018].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[6] Silva SS, Silva RM, Pinto RC, Salles RM. "Botnets: A survey", *Computer Networks*. Vol. 57, Issue 2, pp. 378-403, 2013.

[7] Limarunothai R, Munlin M. "Trends and Challenges of Botnet Architectures and Detection Techniques", *Journal of Information Science And Technology*. Vol. 5, No. 1, pp. 51-57, 2015.

[8] Fredrikson, M., Jha, S., Christodorescu, M., Sailer, R. and Yan, X. "Synthesizing near-optimal malware specifications from suspicious behaviors", *In: Malicious and Unwanted Software:" The Americas"(MALWARE), 2013 8th International Conferenc*e, 22-24 Oct. 2013 Fajardo, PR, USA: IEEE. pp. 45-60.

[9] Karim A, Salleh RB, Shiraz M, Shah SA, Awan I, Anuar NB. "Botnet detection techniques: review, future trends, and issues", *Journal of Zhejiang University Science C*, Vol. 15, Issue 11, pp. 943-983, 2014.

[10] Abdullah, R.S., Abdollah, M.F., Noh, Z.A.M., Mas'ud, M.Z., Selamat, S.R. and Yusof, R. "Revealing the criterion on botnet detection technique", *IJCSI International Journal of Computer Science Issues*, Vol. 10, Issue 2, pp. 208-215, 2013.

[11] Jing-xin W, Zhi-ying W, Kui D. "A network intrusion detection system based on the artificial neural networks", *In Proceedings of the 3rd International Conference* on *Information Security*, 14-16 November, Shanghai, China: ACM, pp. 166-170, 2004.

[12] Derrick EJ, Tibbs RW, Reynolds LL. "Investigating new approaches to data collection, management and analysis for network intrusion detection", *In Proceedings of the 45th Annual Southeast Regional Conference*, 23-24 March, Winston-Salem, North Carolina: ACM, pp. 283-287, 2007.

[13] Kanlayasiri U, Sanguanpong S, Jaratmanachot W. "A rule-based approach for port scanning detection", *In Proceedings of the 23rd Electrical Engineering Conference*, Chiang Mai, Thailand, pp. 485-488, 2000.

[14] Gates C, Becknel D. June. "Host anomalies from network data", *In Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop, IAW'05*, 15-17 June, West Point, New York, USA: IEEE, pp. 325-332, 2005.

[15] Canali, D., Lanzi, A., Balzarotti, D., Kruegel, C., Christodorescu, M. and Kirda, E. "A quantitative study of accuracy in system call-based malware detection", *In Proceedings of the 2012 International Symposium on Software Testing and Analysis*, 15-20 July, Minneapolis, USA: ACM, pp. 122-132, 2012.

[16] Cai, T. and Zou, F. "Detecting HTTP botnet with clustering network traffic", *In 2012 8th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM),* 21-23 September Shanghai, China: IEEE, pp. 1-7, 2012.

[17] Xiang, Y., Li, K., Zhou, W. "Low-rate DDoS attacks detection and trace back by using new information metrics", *IEEE Transactions on Information Forensics and Security*, Vol. 6, Issue 2, pp. 426-37, 2011.

[18] Hajamydeen, AI., Udzir, NI., Mahmod, R. and GHANI, AAA. "An unsupervised heterogeneous log-based framework for anomaly detection", *Turkish Journal of Electrical Engineering & Computer Sciences*, Vol. 24, No. 3, pp. 1117-1134, 2016.

[19] Janratchakool, W., Boonkrong, S. and Smanchat, S. "Finding the Optimal Value for Threshold Cryptography on Cloud Computing", *International Journal of Electrical and Computer Engineering*, Vol. 6, No.6, pp.2979-2988, 2016.

[20] Mulay, P. "Threshold computation to discover cluster structure: a new approach", *International Journal of Electrical and Computer Engineering*, Vol. 6, No.1, pp.275-282, 2016.

[21] Kirubavathi, G. and Anitha, R. "Structural analysis and detection of android botnets using machine learning techniques", *International Journal of Information Security*, Vol. 17, No. 2, pp.153-167, 2018.

[22] Almomani, A. "Fast-flux hunter: a system for filtering online fast-flux botnet", *Neural Computing and Applications*, Vol. 29, No. 7, pp.483-493, 2018.

[23] Alkasassbeh, M. "An Empirical Evaluation for the Intrusion Detection Features Based on Machine Learning and Feature Selection Methods", *Journal of Theoretical & Applied Information Technology*, Vol. 95, No.22, pp. 5962-5976, 2017.

[24] Lee, H., Choi, D., Yim, H., Choi, E., Go, W., Lee, T., Kim, I. and Lee, K. "Feature

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Selection Practice for Unsupervised Learning of Credit Card Fraud Detection", *Journal of Theoretical & Applied Information Technology*, Vol. 96, No.2, pp. 408-417, 2018.

[25] J. Tang, S. Alelyani, and H. Liu, "Feature Selection for Classification: A Review", *Data Classification Algorithms Application*, pp. 37–64, 2014.

[26] Lee, S.J., Xu, Z., Li, T. and Yang, Y. "A novel bagging C4. 5 algorithm based on wrapper feature selection for supporting wise clinical decision making", *Journal of biomedical informatics,* 78, pp. 144-155, 2018.

[27] Arunadevi, J. and Nithya, M.J. "Comparison of Feature Selection Strategies for Classification using Rapid Miner", *International Journal of Innovative Research in Computer and Communication Engineering,* Vo.4, Issue 7, pp. 13556-13563 2016.

[28] Panthong, R. and Srivihok, A. "Wrapper feature subset selection for dimension reduction based on ensemble learning algorithm", *Procedia Computer Science*, 72, pp.162-169, 2015.

[29] Kumar, V. and Minz, S. "Feature Selection: A literature Review", *Smart Computer Review*, Vol. 4, No.3, 2014.

[30] Andy Field. "Discovering statistics using SPSS", 3rd ed. London, UK: SAGE Publications Ltd, 2009.

[31] Sanabila, HR., Fanany, M.I., Jatmiko, W. and Arimurthy, AM. "Bootstrapped multinomial logistic regression on apnea detection using ECG data", *In Conference International of Advanced Computer Science and Information Systems (ICACSIS 2010)*, pp.181-186, 2010.

[32] Hughes, K. and Qu, Y. "A theoretical model: Using logistic regression for malware signature based detection", *In The 10th International Conference on Dependable, Autonomic, and Secure Computing (DASC-2012)*, 2012.

[33] Abdollah, M.F. "Fast attack detection technique for network intrusion detection system", Ph. D, Universiti Teknikal Malaysia Melaka, Malaysia, 2009.

[34] Pepe, M., Janes, H., Longton, G., Leisenring, W. & Newcomb, P. "Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker", *American Journal of Epidemiology*, Vol. 159, Issue 9, pp. 882-90, 2004.

[35] Hosmer Jr, D.W. and Lemeshow, S. "Applied logistic regression". 2nd ed. New Jersey, USA: John Wiley & Sons, 2004.

[36] Hidayah, N.M., Faizal, M.A., Selamat, S.R., Fadhlee, R.M. and Ramzi, W.A.W. "Revealing the Feature Influence in HTTP Botnet Detection", *International Journal of Communication Networks and Information Security,* Vol. 9, No. 2, pp. 274-281, 2017.

[37] Eslahi, M., Rohmad, M.S., Nilsaz, H., Naseri, M.V., Tahir, N.M. and Hashim, H. "Periodicity classification of HTTP traffic to detect HTTP Botnets", *In 2015 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 12-14 April 2015, Langkawi, Malaysia: IEEE, pp. 119-123, 2015.

[38] V. Matta, M. Di Mauro and M. Longo, "DDoS Attacks With Randomized Traffic Innovation: Botnet Identification Challenges and Strategies," in *IEEE Transactions on Information Forensics and Security*, Vol. 12, No. 8, pp. 1844-1859, Aug. 2017.