Multivariate Statistical Analysis on Anomaly P2P Botnets Detection

Raihana Syahirah Binti Abdullah, Faizal M. A., Zul Azri Muhamad Noh
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka Hang Tuah Jaya,
76100 Durian Tunggal, Melaka

***Abstract*—Botnets population is rapidly growing and they become a huge threat on the Internet. Botnets has been declared as Advanced Malware (AM) and Advanced Persistent Threat (APT) listed attacks which is able to manipulate advanced technology where the intricacy of threats need for continuous detection and protection. These attacks will be almost exclusive for financial gain. P2P botnets act as bots that use P2P technology to accomplish certain tasks. The evolution of P2P technology had generated P2P botnets to become more resilient and robust than centralized botnets. This poses a big challenge on detection and defences. In order to detect these botnets, a complete flow analysis is necessary. In this paper, we proposed anomaly detection through chi-square multivariate statistical analysis which currently focuses on time duration and time slot. This particular time is considered to identify the existence of botserver. We foiled both of host level and network level to make coordination within a P2P botnets and the malicious behaviour each bot exhibits for making detection decisions. The statistical approach result show a high detection accuracy and low false positive that make it as one of the promising approach to reveal botserver.**

Keywords—P2P botnets; anomaly-based; chi-square; multivariate; statistical-based

I. Introduction

The researches on botnets and P2P botnets evolution are vital to determine its evolvement in various perspectives. These finding related to its technology evolving and complexity from year to year. In [1] author has stated that as the time passing, botnets is also built with stronger techniques to perform attacks on a large scale. Significantly, this research bridged important relationship with botnets technology as depicted in Fig. 1 where early emergence of P2P botnets existence in year 2002 and rapidly growth until now with more robust, complicated and flexible P2P botnets.

Previous works show that several issues related on P2P botnets remained unexplored. As public know that the botnets is an emergent threat to computer network worldwide. In addition, P2P botnets posed with abnormal behaviors it affected to network operation and network security. Botnets employs fast-flux domain technology that widely adopted by bots servers to improve the productivity of botnets in real time [3], [4]. The fast-flux uses multiple IP address assigned to it that hidden behind a single server. These IP address are swapped in and out of flux with extreme frequency and very short time-to-live (TTL). These technique change the mapping of domain name to different bots with constant shifting that give the attackers additional strength to thwart down the bots servers and obscure their true origin [5]. Mean, it will allow bots to utilize a shifting number of bots servers and effectively hides the botnets attacks from being detected.

In the context of the botnets, fast-flux refers to the strategy of hiding their bots servers to protect botnets communication [6]. In fact, fast-flux is used to obfuscate the specific server involves in their cyber-attack criminal. Besides that, time of attack would be different and continuously changing that make recognition pattern of P2P botnets traffic become difficult. The inconsistency of time attack makes the detection of P2P botnets through time slot trickier and harder. That makes the detection through time slot is not an option and less preferred for previous frameworks. Unfortunately, the bots servers need to be traced and shut down to make sure the P2P botnets stop spreading their communication in a particular time. Taking down the bots servers technically disallowed the P2P botnets launch the attack. Hence, the revelation of bots server is important to deal with P2P botnets attack. Up to now, current P2P botnets detection framework unable to identify and reveal source of real P2P bots server [7]. To address this gap in understanding, the correlation on incoming packet through both of host traffic and network traffic is needed by tracking those P2P botnets in particular time slot. An exact detection framework that able to detect the P2P botnets in real time is extremely needed. So that, our paper enhances a comprehensive P2P detection framework that able to detect the P2P botnets in particular time slot effectively. The distinct P2P botnets across multiple hidden bots server will be identifying using multivariate statistical measurement in particular time slot.

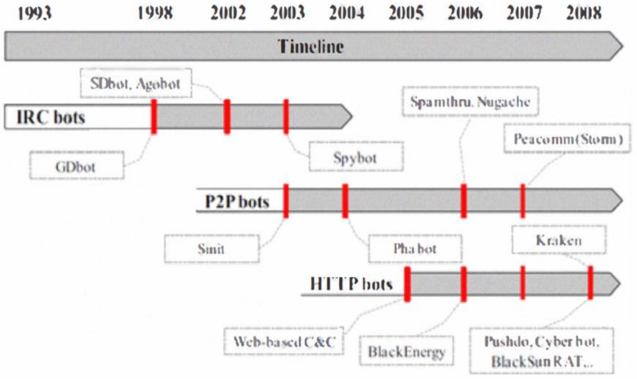

Fig. 1. Timeline of botnets evolution [2].

TABLE I. EVOLUTION OF BOTNETS GENERATION LANDSCAPE

Botnets Generation	Application-based	Types of botnets
1st Generation (1988 - 1998)	IRC-based Centralized C&C Server	GM, GT-Bot, GD-Bot, SD-Bot, Ago-Bot, Spy-Bot, Eggdrop
2nd Generation (1999 - 2001)	IRC-based P2P-based	Pretty Park, Slapper, Gao-Bot
3rd Generation (2002 - 2011)	HTTP-based P2P-based	Sinit, Pha-Bot, Spamthru, Nugache, Peacomm (Storm), Web-based C&C, Black Energy, Bobax, Torping, MyTob, Spyeye, Kraken, Srizbil, Pushdo, Cyber-Bot, BlackSunRAT, Rustock, Coreflood, Zeus, Waledac, Spamit, Bredolab, McColo, Mariposa, Conficker
4th Generation (2012 - now)	P2P-based Hybrid-based Encrypted Communication	Koobface, Kelihos, Grum

The botnets evolvement as clarified by [2], [8] on application based can be simplified as Table I. So in near future, the combination of HTTP and P2P protocols used botnets known as hybrid P2P botnets may be arise with stronger asymmetric cryptography, stronger encryption and private key usage for communication between bots. This integrated P2P architecture provides robust network connectivity, individualized encryption and control traffic dispersion, limited botnets exposure by each captured bot and easy monitoring and recovery by its Botmaster which are hardly traceable compared to other existing botnets in current and in future.

Thus, P2P is not a panacea yet. While this area of tackling on P2P botnets offers great potential and promise, there are still many challenges need to be addressed before the full potential can be realized. P2P botnets dominating most of security problems where the exactly solution and effective detection remain mystery. This security problems become endless challenged for researchers to explore and investigate these issues. Standing on open concept, the P2P ideology of openness and sharing makes these security issues more acute to be handled. By allowing other nodes to access a node's content/service, the node becomes more vulnerable to be attacked in that situation where it acts only as a client. Similarly, many nodes that used to transfer messages had causing the network being more vulnerable to denial-of-service (DoS) attacks. So, it is relatively easy for any malicious node to flood the network with queries. The attacks are harder to detect especially at the application level.

The rest of paper is organized as follows. In Section II, we provide details background on the anomaly-based detection using chi-square multivariate statistical concept. Section III will describe the methodology of overall signature detection process. Next, Section IV will provide details detection module of our proposed P2P botnets chi-square multivariate statistical detection analysis with details results and discussion. Finally, our paper is concluded in Section V.

II. BACKGROUND

Anomaly intrusion detection is able to detect intrusive behaviours according to deviant behaviours and use situation of computer resources. It makes an attempt to describe acceptable behavioural characteristics with quantitative method, to differentiate abnormal and potential intrusive behaviours [9]. However, intrusive activities are not always in agreement with abnormal activities. What anomaly intrusion detection needs to do is to construct abnormal activity set and find out intrusive activity subset therein. Bearing no relation to the system, anomaly intrusion detection has comparatively high universality, and may be able to detect new and unknown attack methods never occurring before as mentioned by [9], [10]. Based on these strengths, this research has considered applying the statistical test in multivariate model with processing the chi-square as an anomaly-based detection. The next sub-section will describe the statistical test used in this research.

The statistical test is included in anomaly-based of Intrusion Detection System (IDS). According to the [11] stated that anomaly detection using statistics will observes the activity of subjects and generate profiles to represent their behaviour. The anomaly-based detection modelled by comparing the data to normal patterns using statistical method that deviates from normal activity [12]. As network events are processed, the system updates the current profiles and periodically calculates an anomaly score by comparing the current profile with stored profile using a function of abnormality. If anomaly score is higher than a targeted threshold, a system is generates the alert as detected. Besides that, the statistical tests are not require labelled data and allowed for zero-day attack detection [13].

Moreover, the statistical anomaly modelling is regularly performed with one of following models which are Operational model or Threshold Metric, Markov Process Model or Marker model, Statistical Moments or Mean and Standard Deviation Model, Multivariate Model and Time Series Model [14]. According to the various researches that have been carried out, each of the technique has performed dissimilarly in different environments and scenarios. Furthermore, [14] had recommended choosing the multivariate models because it can deal with huge amount of network data that possibly changed its behaviours over time, enough resources for computations and the higher security level. Multivariate models are the appropriate choice since they produce better results with less false alarm rate as compared to mean and standard deviation model. Hence, this multivariate model is recommended for host based data and network traffic data, since the bulk of data to be tested is huge. Indeed, in case of distributed attacks, this model can prove to be a very promising technique.

Theoretically, the Intrusion Detection System (IDS) deals with a huge amount of high dimensional data and have a large numbers of behaviour and a high frequency of events occurrence. Multivariate models can be applied for multiple behaviours to measure the suitability towards many intrusions contained multiple subjects and events. Subsequently, multivariate models also considered the correlations between two or more metrics [15]. Thus, a multivariate anomaly detection technique is required for intrusion detection. Otherwise, the IDS also demands on a minimum delay of processing for every event as an early detection for intrusions. Chi-square statistic is a good candidate for intrusion detection in multivariate statistical models with low computation cost. Chi square worked as multivariate but it owned property of robustness that can overcome the IDS problems. Practically, chi square is used to examine the differences between the observed and expected pattern data. It is a goodness-of-fit test that applied to bin data where the data placed into classes. The testing result from [16] has demonstrated the reliable and robust intrusion detection performance of the chi-square technique. They are also highly recommended the deployment of the chi-square technique in IDSs. Due to its good feedback and potential, thus, this research will applied the multivariate in chi-square technique as statistical test to detect the unknown attack in P2P botnets.

The chi-square (x^2) test is used to verify the difference between the measurement and the expected distribution [17]. Furthermore, chi-square test is detected the significant association between two categories of variables [18]. The strength of association between two variables can be tested with developing the hypotheses. In line with that, the chi-square test is defined by the hypothesis that contained with null hypothesis and alternatives hypothesis. There are two rival hypotheses that related with each other. The null hypothesis is tested for possible rejections under an assumption this is going true while, the alternative hypothesis is tested to be accepted and declared as false. The simple hypothesis can be as:

H_0 = This data follow a specific distribution.

H_1 = This data do not follow a specific distribution.

Beside using the classification table in data mining test and signature identification, the chi-square statistical test has been used to indicate whether the parameters can detect the unknown attack or not in the final result. Therefore, the chi-square has been chosen as one of the test that used as the P2P botnets detection in this research. According to [18] and [19], the chi-square formulation of calculation is defined as (1).

$$x^2 = \sum (O_t - E_i)^2 / E_t \quad (1)$$

Where, O_t is the observed frequency for bin *i* and

E_i is the expected frequency for bin *i* and

E_t is expected frequency

III. Methodology

Even though the signature-based detection has been completely done but, several of undetectable P2P botnets are noticeably existed. This situation happens due to the capability of signature-based where it can only detecting the known attack instead of the unknown attack. Standing on the fact, the anomaly-based detection is alternately necessary to conquer this problem. The integration of signature-based and anomaly-based are technically complement of each other weaknesses. As a result, this research presents an anomaly detection technique based on the chi-square statistic. In this case, if the pattern is not recognized in the signature-based, then it will be processed through raise an "anomaly" alarm that allowed as second detection for the unknown intrusion events. This technique is tested for defining its performance in distinguishing normal events from intrusive events in each variant. The study also reveals that the multivariate statistical technique based on the chi-square test statistic as illustrates in Fig. 2 indicates the intrusive events are detected as unknown attack.

```
Statistical tests as an approach in anomaly-based indicate the
evaluation on anomalous traffic volume that act as second
detection for the unknown intrusion events
    [1] Determine categories of packets
    [2] Let time slot = T13 and attributes = TCP Flag
    [3] Calculate statistics of packets distribution, let it with
        $X^2 = \sum (O-E)^2/E$
    [4] Exit
```

Fig. 2. Statistical test detection.

Primarily, for this research, the chi-square statistical test is based on the parameters in the P2P botnets behaviours which detected in integrated analyser module and integrated analysis module. The statistical test is conducted with the discrete of mathematical calculation in the chi-square value. In term of parameters, the TCP Flags and ICMP Flood have been added to use in this research. Both of these parameters are still relevant to be used in this research because the analysis result shown that the behaviours are suitable for this research. The other parameters also has been added in this test including Remote Address Attack, MITM Attack and Poisoning Attack. In order to capitalize on verifying the presence of attacks, the chi-square test calculation has been used which previously applied by [19]. Positively, this research is exploring the power of chi-square statistic to detect the anomalous network activities that appropriately related to the dataset of this research. In overall, [19] was more concentrated to anomaly-based in their detection system compares to this research where the anomaly-based concerned as one of the element that complementary to signature-based. For that, this research increases the power of detection whereby the anomalous volumes of P2P traffic are competently detected. The power of detection technically contributes by the integrated approach that gives extra advantage to this research. Moreover, the detection rate also has been increase through this integrated approach. Thus, this research absolutely defeat previous research [19] where both known and unknown attack have been successfully detected compare to their detection system that only detects unknown attack.

Strengthen to the advantages and unique, chi-square statistic being a good test for detecting the intrusions. In [19], author has successfully proved in their research where chi-square is qualified for the statistical testing to detect various attacks in network activities. In depth, they confirmed that chi square test is suitable with denial-of-service attack and flooding attack. This situation proves that their detection system was more concentrated on the TCP flag and flooding attack. Oppositely on this research, where the priority also stresses on the Remote address attack, MITM and Poisoning attack that happened on P2P botnets. The different of detections give an extra promotion for this research to cover up the limitation on previous research. Furthermore, across to the difference of familiarity, this research differently concentrates on mathematical calculation of chi square formula instead of [19] focused on developing java script coding. By conducting the similar pattern attack as [19], this research had successful get the detection result of attack through the mathematical calculation. The mathematical calculation s exactly acceptable to be used for this research because it can achieve a positive result on detecting the attack. In fact, the testing had shown 100% of intrusion capable to be identified whereby completely defeats the previous research.

Otherwise, the main feature in chi-square statistical test calculation is using time distribution known as time slot. The used of time slot is also referring to previous research by [19] which proved this feature is significant to detect unknown attack. Time slot is used an interval time together with an event counter or resource measure, and take into account the order and the inter-arrival times of the observations as well as their values. Thus, the observed traffic for instance will be labelled as abnormal if its probability of occurrence is parallel with hypothesis at a given time [15]. The process flow for the statistical test is illustrated in Fig. 3. The statistical test is starts with the data inputs from captured P2P botnets dataset. The data preparation has been done with converting the dataset from PCAP file format to CSV data format with selected attributes. All of unnecessary data is eradicated and only useful attributes or behaviours are extracted from PCAP format to CSV format. The selected attributes or behaviours are concerning on TCP Flags, ICMP Flood, Remote Address Attack, MITM Attack and Poisoning Attack. Only the significant information will be processed in the packet and other information from the packets are removed. For this stage, five probabilistic of attributes or behaviours have their own significant information have been categorized as:

1) *TCP Flags*: TCP packets are categorized in three categories which are SYN, RST/ACK, FIN/ACK.

2) *Remote Address Attack*: A list of IP addresses that detected as C&C server.

3) *MITM Attack and Poisoning Attack*: Frequently occur and trusted as an attack.

4) *ICMP Flood*: The flooding will be calculated together with every attack as listed above.

Following by that, these attributes or behaviours are distributed with three main columns where the first column contained the time slot during connection occurs, the second column contained the categories of attributes and behaviours, and the last column will total out the average per second. The distribution is saved in CSV data store as an observed data entries in chi-square statistics. Then, the CSV format is manipulated through mathematical calculation in chi-square value that manually done. In this phase, the selected attributes or behaviours are analysed. After the chi-square calculation is performed, this chi-square value is passed to the decision and conclusion of intrusions. In decision phase the chi-square calculated value is compared with chi-square tabulated value, which is also called critical value [20]. The intrusions will officially declare as occurred when the chi-square calculated value is greater than critical value and vice versa. As a matter of fact, [21] stated that the large difference between the observed and expected frequencies is an intrusion. Thus, the principal step in anomaly-based technique faced with the problem of detecting unknown botnets through show existence of bots in the network. Anomaly-based technique also has the extra capabilities in terms of reducing false negative alert and detecting multistep attack [22]. Nevertheless, it cannot reduce the false positive alert which can only be reduced by using signature-based technique. Hence, this has given an implication that there are complement each other weaknesses. The fully results are briefly discusses in the next section.

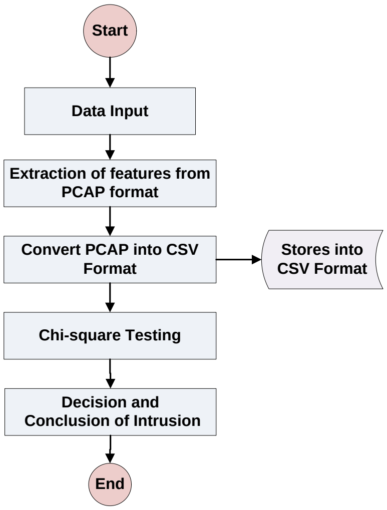
Start
Data Input
Extraction of features from PCAP format
Convert PCAP into CSV Format
Stores into CSV Format
Chi-square Testing
Decision and Conclusion of Intrusion
End

Fig. 3. Process flow of statistical test.

IV. Result and Discussion

The calculation of the statistical approach has been performed by using the time slot as a parameter. Technically, the calculation has been done through five probabilistic attributes or behaviours: TCP Flags, Remote Address Attack, MITM and Poisoning Attack and ICMP Flood that obtained from the analysis part in Chapter Four. The classification of attack and the selection of attributes for each variant determined through the tremendous anomalous volume on dataset. In this research, the calculation will only be done at the first and final time slot whereby the rest of other time slot between the first and final are ignored.

The calculation for first and final time slot has been done to determine the started and ended of location of the attacks happen. In this case, if the final time slot contained the attack, it meant that the attack happened from the beginning until the end of time slot. However, if the final time slot is clean from the attacks, the calculation will moving to the time slot before the last one. This calculation is repeated orderly. Table II shows the numbers of packet average per seconds for Palevo variant based on categories TCP flag set and ICMP packet. This dataset are detected in seven time slots by the existence of TCP flags activities at that time. The TCP flags sets are involved the TCP SYN/ACK, TCP RST/ACK and TCP FIN/ACK.

TABLE II. TIME DISTRIBUTION IN PALEVO DATASET

Time Slot	**Categories and No. of Packets Average Per Seconds**				**Total**
	TCP SYN/ACK	**TCP RST/ACK**	**TCP FIN/ACK**	**ICMP**	
31243-31251	24197.28	24197.97	24197.97	24199.22	96792.44
31254-31257	24198.00	24198.27	24198.29	24450.97	97045.53
31414-31416	24202.42	24202.42	24202.72	24750.96	97358.52
55285-55287	41018.23	41018.23	41018.23	41018.18	164072.86
55288-55292	41018.23	41018.75	41018.75	41019.30	164075.03
55293-55298	41018.75	41019.30	41019.30	41020.80	164078.15
55300-70810	41019.30	47078.56	58660.36	58950.20	205708.42

The first calculation in chi-square statistical test is entailed by the average of packets per second and relative frequencies as tabulated in Table III. The average of packets per second are derived from Table II with calculating the total of each category dividing with number of categories. Then, the relative frequencies can be easily calculated as dividing the total number of average packets per second by the total average packets per second for each category.

TABLE III. AVERAGE PACKET DISTRIBUTION AND RELATIVE FREQUENCIES IN PALEVO DATASET

Categories	**No. of Average packets per second**	**Categories**	**Relative Frequencies**
SYN/ACK	236672.21	SYN/ACK	0.24
RST/ACK	242733.49	RST/ACK	0.25
FIN/ACK	254315.61	FIN/ACK	0.25
ICMP	255409.62	ICMP	0.26
Total	989130.95	Total	1

Before continuing to the next chi-square statistical calculation, the main hypothesis that need to derive in this test are:

H_0= The first and last Time Slot has the specified distribution or there is no intrusion, and

H_1= The first and last Time Slot does not has the specified distribution or there is an intrusion

The observed values and x^2 test calculation for the first and last time slot is defined respectively are shown in Tables IV and V. From these two tables, the x^2 goodness-of-test statistic calculation to be:

$$x^2 = \sum (O - E)^2 / E = 4986377.62 \text{ and } 2865.95$$

Let assume that the hypothesis test is performed at 5% significance level so ($\alpha = 0.05$). There are four types of categories within in the test, so k = 4 and the degree of freedom becomes as df = 4 - 1 = 3. Then, this research diligently check the chi-square table in Appendix E, with α = 0.05 and df = 3, the chi-square tabulated value are 7.82. As a result, let do the significant comparison here where the chi-square tabulated value = $x^2_{0.05}$ = 7.82 and chi-square calculated value = x^2 = 4986377.62 and 2865.95. Hence, the chi-square calculated value is greater than chi-square tabulated value, so the null hypothesis H_0 is rejected and the H_1 is accepted. It means that there is an intrusion or anomaly in the Palevo dataset at the first and last time slot. This research has defined the differences between observed and expected frequencies. So, it can be concluded that there is TCP flags set attack in the Palevo dataset. The statistical approach that applies in anomaly-based detection has proved that the undetectable P2P botnets in signature-based module can be detected through this approach. The result in Table VI shows the P2P botnets can be detected effectively in the anomaly-based rather than signature-based result. The false negative concerns as the undetectable numbers of attack that fail to be detected in signature-based module. This situation happened when the unknown attack has been detected normal. Significantly, the unknown attack is tackles by conducting the anomaly-based detection whereby the chi-square statistical test with multivariate process has been performed. The classification of attack that has been selected through the tremendous anomalous volume on the dataset which alerting of P2P botnets symptom.

TABLE IV. COMPUTATION OF CHI-SQUARE TEST STATISTIC FOR THE FIRST TIME SLOT IN PALEVO DATASET

Time Slots	**Categories**	**Relative Frequencies (f)**	**Observed Frequencies (O)**	**Expected Frequencies (E=n*f)**	**(O - E)**	**(O-E)2**	**(O-E)2/E**
31243-31251	SYN/ ACK	0.24	24197.28	23159.80	1037.47	1076362.89	46.48
	RST/ACK	0.25	24197.96	23752.93	445.029	198051.07	8.33
	FIN/ ACK	0.25	24197.97	24886.31	-688.34	1.24091E+1	4986297.58
	ICMP	0.26	24199.21	24993.37	-794.15	630688.11	25.23
	Total	1	96792.43				**4986377.62**

TABLE V. COMPUTATION OF CHI-SQUARE TEST STATISTIC FOR THE LAST TIME SLOT IN PALEVO DATASET

Time Slots	**Categories**	**Relative Frequencies (f)**	**Observed Frequencies (O)**	**Expected Frequencies (E=n*f)**	**(O - E)**	**(O-E)2**	**(O-E)2/E**
55300-70810	SYN/ ACK	0.24	41019.29	49220.44	- 8201.14	67258833.47	1366.48
	RST/ ACK	0.25	47078.55	50481.00	- 3402.44	11576634.01	229.32
	FIN/ ACK	0.25	58660.35	52889. 72	5770.63	33300225.67	629.62
	ICMP	0.26	58950.20	53117.24	5832.95	34023406.29	640.53
	Total	1	205708.42				**2865.95**

TABLE VI. THE INTEGRATION OF SIGNATURE-BASED AND ANOMALY-BASED DETECTION RESULT

Variant	**False Negative (Undetectable) in Signature-based**	**Successful detected in Anomaly-based**
Allaple.L	2	6
RBot	6	8
Palevo	3	7
Srvcp	2	6

Additionally, the result shows that detection not only capable to identify the undetectable value in signature-based but also the statistical test able to detect more than predictable value in anomaly-based. This outcome demonstrates that the others unknown attack also has been successful detected. The incremental of the effectiveness towards the integration of detection techniques known as integrated technique with the integrated approach help on boosting the detection values. At the same time, the integrated technique with the integrated approach will complementary the weakness and integrate the best result. Other than that, result proves that the correlation between anomaly-based and signature-based are essentially needed and relevantly to be used in detecting the P2P botnets.

V. CONCLUSION

Currently, the technique or approach that has been chosen by most of researchers are not comprehensive enough because they cannot reveal the botserver in specific time. But, this study presents a statistical approach in order to detect existence botserver in specific time manner. The proposed anomaly detection module is based on chi-square multivariate analysis. The result show that the proposed detection module have high detection accuracy with ability to detect some unknown P2P botnets and produce a high detection rate with low false alarm rate. Hence, the developing detection module based on anomaly-based has been the most promising approach to fight against botnets threat by take down the real botserver.

The further research can be covered on different parameter and technique by increasing the accurate of detection.

ACKNOWLEDGMENT

The authors would like to express the appreciation to Inforsnet Group Research of Universiti Teknikal Malaysia Melaka (UTeM) and MyBrain15 Programme by Ministry of Higher Education Malaysia (MoHE) in encouraging the authors to publish this paper. This work was supported by the MoE of Malaysia under Grants FRGS/1/2014/ICT04/FTMK/02/ F00212.

REFERENCES

[1] Tyagi, A. K. and Aghila, G. (2011), " A Wide Scale Survey on Botnet", International Journal of Computer Applications 34(9):10-23, November 2011.

[2] Tung-Ming, K., Hung-Chang, C., and Guo-Quan, W., 2011. Construction P2P firewall HTTP-Botnet Defense Mechanism. *International Conference on Computer Science and Automation Engineering CSAE, IEEE.*

[3] Ching Hsiang Hsu, chu Ying Huang and Kuan Ta Cheu (2010). Fast-flux Bot Detection in Real Time. *Proceeding of RAID.*

[4] Chia Mei Chen, Sheng Tzong Cheng and Ju Hsien Chou. (2013). Detection of Fast-flux domains. *Journal of Advances in Computer Networks, Vol. 1, No. 2.*

[5] Nazario, J., and Holtz, T., 2008. As the net churns: Fast-flux Botnets Observations. *3rd Proceeding International Conference on Malicious and Unwanted Software.*

[6] Martinex-Bea Sergi, Sergio Castillo-Perez, Joaquin Garcia-Alfaro (2013). Real Time Detection Malicious Fast-flux Detection using DNS. *11th Annual Conference on Privacy, Security and Trust (PST).*

[7] Muthumanickam, K. and Ilavarasan, E. (2012), "P2P Botnet Detection: Combined Host-and Network-Level Analysis", *ICCCNT, India.*

[8] Charles, L. 2013. *Malware Threats in our Cyber Infrastructure*, Yogyakarta: Swiss German University.

[9] Chitrakar, R. and Chuanhe, H.: Anomaly Detection using Support Vector Machine Classification with k-Medoids Clustering. *IEEE,* 2012

[10] Bao, X., Xu, T. and Hou, H. (2009), "Network Intrusion Detection Based on Support Vector Machine," International Conference on Management and Service Science (MASS).

[11] Zhang, W., Yang, Q. and Geng, Y. 2009. A Survey of Anomaly Detection Methods in Networks. *International Symposium on Computer Network and Multimedia Technology CNMT.*

[12] Ngadi, M., A., Yazid, M. I., and Hanan, A., 2005. A Study on Advanced Statistical Analysis for Network Anomaly Detection. *Project Report, Faculty of Computer Science and Information System, Skudai, Johor.*

[13] Ferragut, E.,M., Laska, J. and Bridges, R.A. 2012. A New, Principled Approach to Anomaly Detection, *11th International Conference on Machine Learning and Applications, IEEE.*

[14] Qayyum, A., Islam, M.H., and Jamil, M. 2005. Taxonomy of Statistical Based Anomaly Detection Techniques for Intrusion Detection. *International Conference on Emerging Technologies, IEEE.*

[15] Teodoro, G. P., Diaz-Verdejo, J., Macia-Fernandez, G., and Vazquez, E., 2009. Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges. *Computers and Security 28*, pp 18-28.

[16] Ye, N., Emran, S.M., Li, X. and Chen, Q. 2001. Statistical Process Control for Computer Intrusion Detection. *Proceedings of DARPA Information Survivability Conference and Amp; Exposition II DISCEX, IEEE* .

[17] Oshima, S., Nakashima, T. and Nishikido, Y., 2009. Extraction of Characteristics of Anomaly Accessed IP Packets using Chi-Square Method. *International Conference on Complex, Intelligent and Software Intensive Systems, IEEE.*

[18] Andy Field, 2005. *Discovering Statistic using SPSS 2nd edition*, London: Sage Publication.

[19] Rahul, R., Zubair, K., and Khan, M.H., 2012. Network Anomalies Detection using Statistical Technique: A Chi-Square Approach. Vol. 9, Issues 2, pp. 3.

[20] Cristianini, N. and Shawe-Taylor, J., 2006. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Beijing: Publishing House of Electronics Industry.

[21] Richard, J. 2001. A Rough Set Aided System for Sorting WWW Bookmarks", Springer.

[22] Robiah Y, Siti Rahayu S., Mohd Zaki M., Shahrin S., Faizal M. A., Marliza R: A New Generic Taxonomy on Hybrid Malware Detection Technique. *(IJCSIS) International Journal of Computer Science and Information Security*, Vol. 5, No. 1, 2009