Journal of Information Assurance and Security.
ISSN 1554-1010 Volume 9 (2014) pp. 062-071
© MIR Labs, www.mirlabs.net/jias/index.html

Enhanced P2P Botnets Detection Framework Architecture with Hybrid Analysis Approach

Raihana Syahirah Abdullah
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka.

E-mail: rasyahb@gmail.com

Faizal M.A., Zul Azri Muhamad Noh
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka.

E-mail: {faizalabdollah, zulazri}@utem.edu.my

Abstract—**Nowadays, botnets are the most advanced cybercrime as being powerful threaten to the internet infrastructure by risking the Internet stability and security. Millions of computers have been hijacking and infecting by botnets especially during peak activity. The P2P botnets exploit users and dominating the P2P technology which make botnets are harder to detect and terminated. As P2P botnets issues been highlighted as it's dramatically evolvement, this paper addresses on current problems relate to P2P botnets faced by users and recommending the improvement. Also, this paper concentrated on proposing P2P botnets detection framework. Also, an in-depth analysis of P2P botnets has been conducted to understand and cope with their behaviors and characteristics. The new improvement has been introduced at the propose botnets framework architecture to improve the effectiveness of P2P detection analysis. The framework architecture has been structuralized with hybrid analysis through the integration of static analysis and dynamic analysis. Prior to this matter, this research has proposed a new enhancement on framework architecture that has been reinforced by hybrid detection technique to improve the effectiveness and efficiency of P2P botnets detection.**

Keywords-P2P botnets, detection framework, hybrid analyzer, hybrid analysis, hybrid technique

I. INTRODUCTION

Security issues have been overshadowed by intensity of global Internet usage. Recently, the number of users rapidly evolving and continually growth due to the access of Internet today are easier than before. This situation portrays where users spend more time on Internet usage in most of their daily activities including those relate to their work. The fact is when the users start to turn on the computers, users will completely expose to the threats on the cyberspace. In different dimension, they actually exposed with variety of threats including botnets attack where technically influence by bots or zombies. Theoretically, there were many version of definition have been used by the previous researcher sto described the botnets. Botnets has been illustrate as network of compromised computer running malicious software and infected by all king of techniques: worms, Trojan horse, and viruses [1]. Then, these zombie computers are remotely controlled by an attacker called Botmaster. Before we jump into detailed discussion, the basic fundamental that need to be understand and highlight is botnets as one of the most emergent threats and also emerging phenomenon [2]. The architecture of early botnets which constructed and based on central control server is relatively simple especially on the IRC and HTTP. However, the presence of P2P botnets supported by the centralized and distributed communication makes it more invulnerability to detection. This is because P2P botnets have the flexibility, complexity and robustness that make it become benign application. Thus, P2P botnets were seen as an emerging phenomenon by able to adapt and thrive in its environment. This would allow a big impact on the internet structure especially in negative perspective.

Many research groups center and individual attempt to investigate and develop a new technique in P2P botnets detection. The in-depth analysis of P2P botnets is provided by Wang P. et. al. [3] through a detailed description of P2P botnets has been presented by systematically focused the study on P2P lifetime, types of P2P botnets and countermeasures for P2P botnets. Hossein R. Z. et. al. [4] have discussed on two approaches of botnets detection techniques based on setting up honeypots and Intrusion Detection Systems (IDS) guided by a complete survey that has been done. Meanwhile, Duan J. et. al. [5] has analyze different P2P botnets such as Storm and Nugache to examine the behavior of these botnets. Besides that, Sun D. et. al. [6] analyses different P2P botnets that only focuses on P2P functionality. Rather than that, many papers have been conducted to discuss how to distinguish flows between the normal Internet and botnets, some was paper by Liao W.H. and Chang C.C. [7], Han L. [8] and Chitrakar R. and Chuanhe H. [9] which used data mining to judge the P2P botnets detection.

Moreover, it is clear that many researchers try to find out and develop their way to detect and mitigate the botnets. In this case, Massi J. et. al. [10] have illustrated the network flow data equivalent with botnets behavior. Their flow is

based on the steps that will be use to identified characteristics of malicious network traffic from botnets that are common across some botnet attacks. To date, the research that has been conducted by Anil, S. and Remya, R. [11] purposely to analyze and detect P2P botnets in hybrid method using combination of genetic algorithm (GA), Self-Organised Map (SOM) and Support Vector Machine (SVM) as an anomaly detection. The equation of all research determine that differentiating and legitimate P2P communications is a difficult task and they have to admit that botnets detection is a challenging problem. Most of researchers have concentrated on general P2P detection technique make the current detection techniques are still not comprehensive enough.

Due to current issues, therefore, it is very important that security techniques of P2P botnets detection is developed that can prevent botnets attacks to system resources and data. This will definitely minimize losses face by the organization. However, completely preventing botnets in P2P network, at present, ineffective and less capable. Furthermore, the main impact faced by the organization due to the security incident especially botnets in P2P traffic will give the major losses to the organization. Hence, detecting the botnets especially in P2P network as possible may reduce the financial loss and data damage occurring inside the network of the organization. As reacting to the demand P2P botnets pattern, the hybrid technique with statistical approach offers the best solution for security manner which helped detection on P2P botnets activity with low false positive and high detection accuracy.

II. RELATED WORK

Currently, researchers more alerts and interested about the evolving P2P botnets as it become a universal issues. Many of researchers started to concern on tackling the prevention, detection and defenses on P2P botnets. There also encountered the ways to take down P2P botnets by modeling the models, structuring the architectures or developing the frameworks to make the detection systems. This situation has promises a good sign for computer security appropriately diminishing the botnets evolvement. Numerous frameworks detection systems have been proposed for detecting the botnets mostly in centralized botnets; IRC-based and HTTP-based.

To date, a few frameworks detection systems for P2P botnet detection have been proposed [11][12][13][14][15][16][17] [18][19][20][21]. Through the significant reviews showed that many studies have been conducted relate to IRC and HTTP detection systems. However, the study on P2P botnets detection is still limited where offers room to be improved. Due to this point, this paper is proposing improvement framework architecture about P2P botnets detection. This significance research has been supported by Sundaram in his paper which highlighted the importance of building a security mechanism for preventing any intrusion from hacker so that we can take action and improve the system security [22].

A. The Revolution of P2P Botnets Detection System

Recently, the P2P botnets detection has attracted many researchers to study in this area after the emergence of P2P application issues extremely increased. Zeidanloo and Eternad [11][12] used a network-based anomaly detection to identify IRC, HTTP and P2P-based botnets. Despite it not require prior knowledge, [11] suffers from the effect of different flow interval durations was not presented and the accuracy approach is unknown. Chunyong and Ghorbani [13] and [11] are not mined the data due they depend wholly on classifier module to cluster the data. However, they are successful in detection on unknown botnets attack. Meanwhile, Yuanyuan et al. [14], Chunyong and Ghorbani [13] and Muthumanickam and Ilavarasan [17] performed the combined of host and network analyzer with directly trigger action to detection reports in real time detection. Nevertheless, Yuanyuan et al [14] and Muthumanickam and Ilavarasan [17] do not provide filtering module because the lack on payload inspection may occurhigh identification accuracy. The direct network flow and host log analysis provided by [17], make it do not feasible for offline detection. Then, Arshad et al [16] and Junjie et al [15] utilized the passive monitoring concern on network traffic. However, their detection has consistently shown that it is not required prior knowledge and draws our attention to the real time detection. Li et al [18] also conducted a single detection methods where are network-based analyzer. This detection system become ineffective against the new P2P-based botnets infected detection since the detection only done in Storm, Waledac and Skype application.

Hence, by reviewing the previous studies, most of them target to detect in network-based level [11][12][15][16][18] rather than host-based level [13][14][17]. Besides that, majority researchers used dynamic analysis [11][12][13][14][15][16][17][18] rather than static analysis. Means, the analysis only had been done during or after program execution. This situation lead researcher capable to detect malicious behavior and activity during or after program is running [23]. In addition, no research has been found using hybrid analysis that combined static and dynamic analysis. The research to date has tended to focus on anomaly-based detection rather than others technique to tackle the problem on unknown botnets detection with low level false positives alarm [11][14][16]. As the appropriate detection system, most of previous work had done with mined data in ensuring a good data preparation by remove outliers as a key to producing valid and reliable framework [14][15][16][17][18]. Statistically, a major problem with the current framework is it cannot reveal the bot servers and C&C migration due they not compromise with protocol and structure independent. So far, only [14] can trace and reveal the bot servers. Then, the research works has focus more on network layer and application layer [11][12][13][17][18]

due the layer will be fully visible the full picture ofP2P botnets infections.

The reviewed session has proved that there are limited P2P detection systems have been done in network security field which is very beneficial to both side of industries and users. It is a valuable works that every researcher struggled and useful knowledge to fight against security problem in real network environment. Technically, there are lots of botnets detection systems relates to IRC and HTTP-based. However, the studies for P2P-based detection and prevention are critically required more action of improvement.

In this paper, our work is not constrained by single detection technique and does not require the learning of data label to detection. Moreover, our approach offers more robustness because it does not rely on one single detection analyzer and one analysis approach. So far, all the previous detection systems have their own drawbacks and some room of improvement needs to be carried out. Toward enhancing the previous framework, this develop framework architecture provides better improvement since the detection are made by doing the hybrid analysis, hybrid technique and combination of host analyzer and network analyzer. Only several researches do the hybrid analyzer [13][14][17], compare with the rest do the single level detection [11][12][15][16][18]. Otherwise, the difference between this improvement frameworks with others are the detection involve on the majority detection on OSI layer rely on Data Link Layer, Network Layer, Transport Layer and Application Layer. The rest of previous studies only focus on Application Layer [11][13][17][18]. The details of improvement P2P botnets detection framework architecture are defined on Section V.

B. P2P Botnets Behaviour

Most of studies have reveals the behaviours concentrated on survey and literature of the P2P botnets. They are collected the common trends and characteristics of P2P botnets from the traditional P2P variants. Zang et al [23] and Leder et al. [24] conduct a review on several P2P variants to fine the flow of classification and encounter with offensive approach. Works done by Junfeng et al [25] have come out with a complete comparison between P2P variants on several features. The study makes the comparison on Storm and Nugache as P2P variants. Only two studies focused in some technical study approach on a P2P variant. They were studies by Donghong et al [1] and Chao et al. [26] that designed the mechanism to differentiate the P2P variants. The mechanisms are involved the command, control, infections, propagation, exploits, attack and survivability mechanism. With the selected P2P variants, they have illustrates the characterization and summarization on the particular P2P variants. These studies continually contributed much in considering the evolution of P2P botnets behaviours in general conceptual.

Furthermore, the general concept will help the researcher to have better understanding about P2P variants on how the P2P botnets will infect the host and network. This can be done by analyzing the P2P botnets itself by doing a reverse engineering. The reverse engineering analysis consists of two main approach which are static analysis and dynamic analysis. The integrated of analysis recognized as hybrid analysis will discovers the P2P behaviours and characteristics that give beneficial on understanding of P2P botnets in order to develop an effective P2P botnets detection techniques.

C. Hybrid Analysis Approach

In hybrid analysis approach, two of analysis approaches were combined. It is the combination of static analysis and dynamic analysis approach. Due to static analysis, it has capability to detect malicious activity before the programs been executed. This allow the raw codes of infected files can be revealed and give initial perceptions to administrator before the entire analysis is performed. Instead, dynamic analysis by chance has the ability to detect the malicious activity during and after the programs executed. The dynamic analysis essentially completes the whole analysis on fully diverse logs and network packet in a good manner. Based on analysis by [27], the combination of hybrid analysis approach has given an implication that there are complement each other weaknesses.

III. METHODOLOGY

The methodology of overall analysis process illustrated in Figure 1 is started by performing analysis in each of P2P botnets infected file. The hybrid analysis approach discover the two level of analysis: static analysis and dynamic analysis which the analysis is done on every single host log and network packet captured in order to distinguish whether its payload is malicious or spam, either it corresponds to a remote check for vulnerabilities, or whether it follows unusual conventions with respect to normal P2P options. For final hybrid analysis result, these two types of analysis approach are correlated.

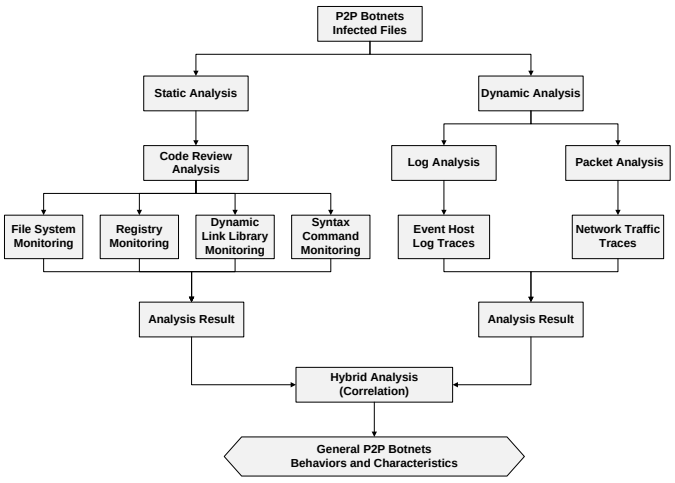

Figure 1: P2P Botnets Hybrid Analysis Approach [28]

Moreover, the hybrid analysis as depicted in Figure 1 is performed by considered the two perspectives level: host-level and network-level environment are inherent from our previous study in [29]. In host-level, the host logs were analysed by the characteristics on file system monitoring, registry and log monitoring. Instead, in the network-level, the characteristics on the full-payload from network packet were examined.

At static analysis, analysis is done on file system, registry, dynamic link library and syntax command monitoring. This static analysis had been done in reviewing the real codes of infected files to reveal and study their true characteristics. The P2P botnets generally will create and load their code file and make the dramatic changes in registry in order to exploits the host. Then, it attempt to inject the code by execute the dynamic link library (*.dll)* and generate it to stack overflow status. Then, every single of P2P infected files was captured by their own syntax command where it will correspondent to botnets server. The important of code review analysis is it can extract the information on system and application log on each of infected host includes the event status, process status, action status and more.

Subsequently, the dynamic analysis had been done on the event host log and network traffic dataset. The dataset has been collected by implementing the P2P testbed setup in a controlled environment. The event host log has captured by process monitor and process explorer in order to gather the information on local host. Meanwhile, the overall network traffic is captured by tcpdump service. Through this dataset, the P2P botnets behaviour and characteristic fully observed to ensure the interaction on the botnet server and the effect on each of infected files to real environment. After that, the combination of analysis result on static and dynamic analysis will be correlated together to construct the general behaviour of P2P botnets.

A. P2P Botnets Testbed Scenario

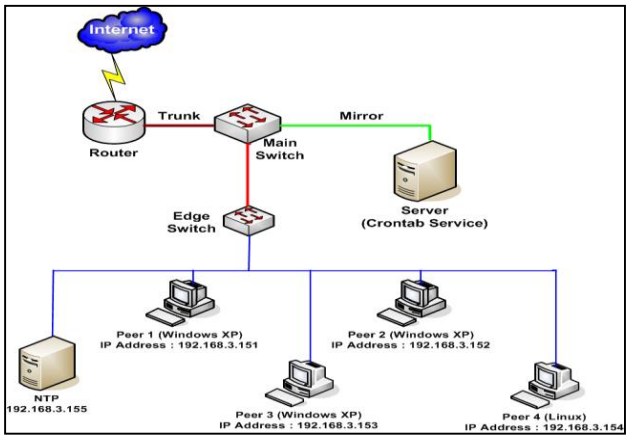

Figure 2: P2P Testbed Setup

At first stage, the controlled environment has been implemented known as P2P testbed environment. The experimental testbed lab is conducted to capture the P2P Botnet activities with similar configuration have been used by Faizal [30]**.** In this paper, the two P2P botnets attack scenarios: host-level scenario and network-level scenario are designed using a complete P2P botnets life cycle consists of attack steps: Scan, Exploit, C&C Connection and Impact/Effect. These attack steps is mapped to basic botnets life cycle as motivated by [23][31][32]. Each level of scenario is attained through hybrid analysis. The hybrid analysis has been done by diverse the analyzer into two categories which are host analyzer and network analyzer.

B. Data Collection

We have captured the P2P botnets traffic in 1 week running continuous period from P2P scenario as described above in a controlled environment. The testbed used in this research consist of one router, two switches, four peer that placed with a fresh installation of Windows XP 32-bit and Linux, one NTP server and one server to perform the capturing packet process. Besides that, each peer had installed by process explorer, process monitor and the activation of P2P botnets infected files is released for each of peers. The P2P botnets infected files are provided by the MYCERT of Cyber Security Malaysia. The activation has been done to launch the P2P botnets attack in the second stage. Afterwards, the event logs are collected by process explorer (procex) and process monitor (procmon) application provided by sys internal. Meanwhile, the network traffic was collected via tcpdump service using the crontab server. This event logs and network packet data will be analyzed through the next stage.

IV. ANALYSIS AND FINDINGS

To evaluate the performance of proposed improvement detection framework architecture, it is tested on hybrid

analysis that was concentrated on Layer 4 components. Layer 4 known as Analysis Layer will analyze the whole of detection process in recognizing the attack pattern and behavior of P2P botnets. From the hybrid analysis, it is found that there are significant different between a normal P2P traffic and abnormal P2P traffic in the static analysis and dynamic analysis. This analysis also bringing together a considerable contribution especially on early warning of P2P botnets activities in the network as P2P botnets detection. The details of the analysis are described in the next section.

A. Code Analysis

Code analysis act as the static analysis process where has enlighten the capabilities to detect the malicious activity before the programs files is running. The operation process involved in every P2P botnets infected files has thread create as its main process start. The information of P2P botnets infected files has discovered in four main components: file monitoring, registry monitoring, dynamic link library monitoring and syntax command monitoring as illustrated in Figure 3.

Table 1 shows the list of operation process that occurred in the P2P botnets infected files. Operation process like *Create File, Load File, Write File* and *Delete File* are considered as dangerous event process that happen to allow the attacker replace the original files in operating systems. Here, we can see the attempting of malicious activity in trying creates an *.exe* file and delete essential files in system directory to exploits the victim host. Otherwise, the frequent changes in registry will be noticed as high possibility of malicious activity has arisen. The operation processes in registry that need to be addressed are *RegOpenKey, RegSetValue and RegCloseKey*. The *TCP/UDP Reconnect, TCP/UDP Disconnect, TCP/UDP Send* and *TCP/UDP Send* are indicates the TCP and UDP communication is luring in the network.

Table 1: Operation Process by P2P Botnets

	Invalid Hash	Allaple.L	RBot	Palevo	srvcp	tmnbtib
Thread Create	√	√	√	√	√	√
Query Name Information File	√	√	√	√	√	√
Load Image	√	√	√	√	√	√
Create File	√	√	√	√	√	√
Read File	√	√	√	√		
Load File	√	√	√			
Write File			√	√		
Delete File			√	√		
Close File	√	√	√	√	√	√
RegOpenKey	√	√	√	√	√	
RegCreateKey		√		√	√	
RegQueryValue	√	√	√	√		
RegSetValue		√	√	√	√	
RegCloseKey	√	√	√	√	√	
File System Control	√	√	√	√		
Query Open					√	
Create File Mapping	√		√	√	√	
Set End Of Information File	√	√	√	√	√	
TCP/UDP Reconnect		√			√	√
TCP/UDP Disconnect		√			√	√
TCP/UDP Receive		√				
TCP/UDP Send		√				
Dynamic Link Library	√	√	√	√	√	√
Syntax Command	√	√	√	√	√	√

The most important part that torching an intention are stack overflow occurrence that generates by *dynamic link library* (*.dll*). In fact, a stack overflow is an undesirable condition which a particular P2P botnets tries to use more memory space than the call stack has available to inject their codes. As a result, the P2P botnets excessive demand for memory space to launch such attack and the host may crash immediately. Thus, the risk of P2P botnets exploits in a maximized situation. Besides that, each of P2P botnets has their own *syntax command* from bots server to instruct the recruit bots to do the malicious task such as steal any information in host and network.

Actually, the in-depth reviewing of P2P botnets codes were revealed the details of behavior and characteristic. It can be demonstrated by an example on Palevo analysis as shown in figure 6. All of dangerous event process involved *Create File, Write File* and *Delete File* are occurred in this infected files to exploits the original directory files. Then, the significant registry which is *RegOpenKey, RegSetValue* and *RegCloseKey* also encompass make the host is expose in malicious activity. The dynamic link library become as stack overflow on *user32.dll, kernel32.dll, gdi32.dll, ole32.dll, comdlg.dll* and *ntfll.dll.* Other monitoring is syntax command as define in Figure are used by bot server to update, make connection and permit any command to launch various attack.

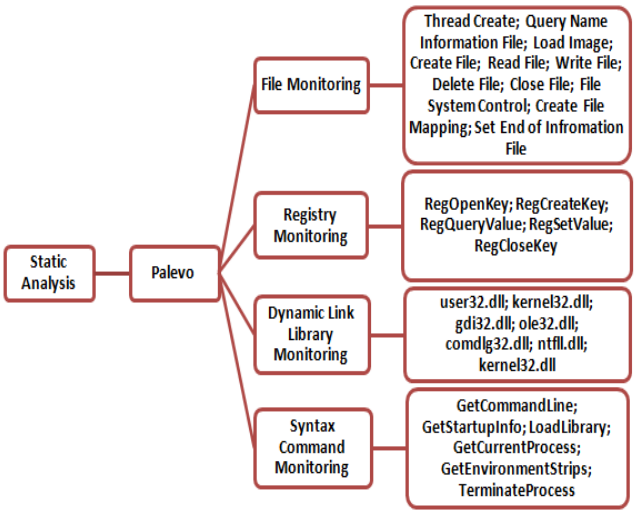

Figure 3: Palevo Operation Process in Static Analysis

B. *Packet and Log Analysis*

While the packet and log analysis is act as more on the dynamic analysis process. In dynamic approach, the capabilities are concern on the detection of malicious activity during or after program files execution. The P2P botnets activities are captured through the event host logs and whole network traffic. The Palevo dynamic analysis as depicted in Figure 4 shows the P2P botnets attempting to contact the bots server by using vulnerable port and certain IP address as remote address. The Palevo make the connection via the bots server by receiving the command to launch a series of attack on MITM, poisoning, ARP, TCP flag and ICMP flooding attack. Yet, a similar analysis activities are also been done in Invalid Hash, Allaple.L, RBot, srvcp and tnnbtib.

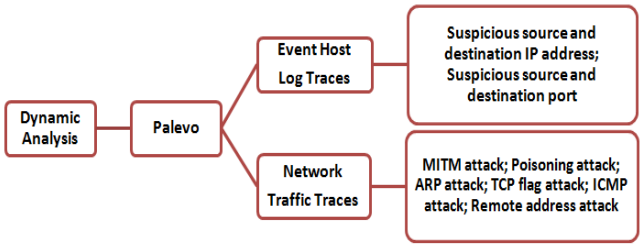

Figure 4: Palevo Dynamic Analysis

C. *Behavioural Analysis*

The hybrid analysis result as described in previous section is utilized to construct the P2P botnets behaviour model. The new edition of P2P botnets has been proposed in this research called as a general P2P botnets behaviour model. The proposed model consists of scan and exploits, C&C communication and impact/effect as shown in Figure 5. The details are:

1) Scan and Exploit - P2P botnets scan the vulnerable port and suspicious IP address. Then, it had exploited the victim host and network by capturing the private and confidential information.

2) C&C Communication - P2P botnets make the connection with C&C communication that known as bots server via P2P network to establish communication among bots

3) Impact/Effect - P2P botnets launch various attack to accomplish certain malicious task

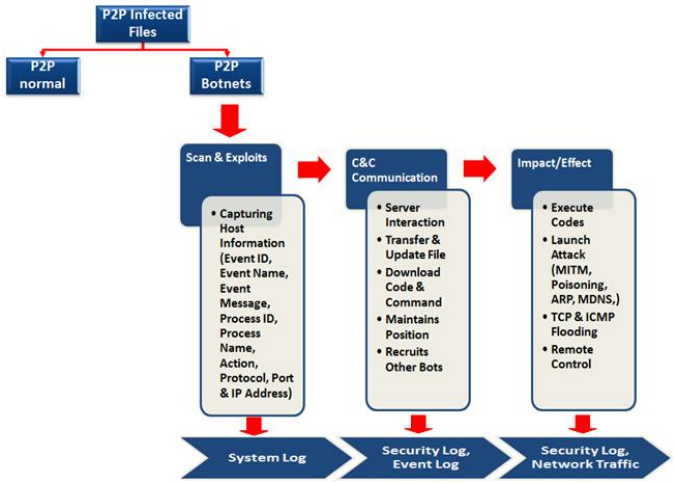

Figure 5: General P2P Botnets Behavior Model

The comprehensive P2P botnets behavior model with suitable log to be monitored essentially illustrated in figure 5. The analysis had been done through a basic of P2P botnets life cycle involving Scan & Exploit, C&C Communication and Impact/Effect phase. At first, the Scan & Exploit phase will capturing the whole details information in system log including on event ID, event name, event message, process ID, process name, action, protocol, port and IP address. Subsequently, the server interaction with C&C Communication is establish in second phase to accomplish the task on transfer and update files, download codes and command, maintains the positions and recruits other bots as a member. The logs to be monitored in second phase are security log and event log. In a final phase which is Impact/Effect, the P2P botnets will executes code to launch malicious attack, flooding attack and remote control access. This malicious event is monitored on security log and network traffic log.

D. *Attack Anatomy*

After all, the recognition of P2P botnets attack anatomy has been constructed by accomplish on both of hybrid analysis which are static analysis and dynamic analysis. As discussed in previous section, this phase mapped the overall operation process with event log and network traffic traces. The correlation process contributed in creating the sequence of P2P botnets activity and event. Hence, the findings are beneficial for further purpose on build the general P2P botnets behaviour. Consequently, an overall correlation

phase represent as a complete sequence of RBot activities and events are demonstrated in Figure 6.

E. Evaluation Results

In order to validate the analysis efficiency and detection accuracy of our detection technique, we respectively labeled the dataset as P2P botnets files in both logs; host log or network traffic. On this basis, the dataset and detail information was shown in Table 2 below:

Trace	Duration	No. of Peers	Size	# of packets flows	Detec-ted as P2P Bot
Normal	7 days	5	7.4 GB	12,495,782	X
Invalid Hash	7 days	5	21 MB	258,645	√
Allaple. L	7 days	5	23.6 MB	290,668	√
RBot	7 days	5	17.8 MB	219,233	√
Palevo	7 days	5	23.1 MB	284,510	√
Srvcp	7 days	5	19.6 MB	241,402	√
Tnnbtib	7 days	5	25.4 MB	312,838	√

Table 2: Datasets Used in Our Evaluation

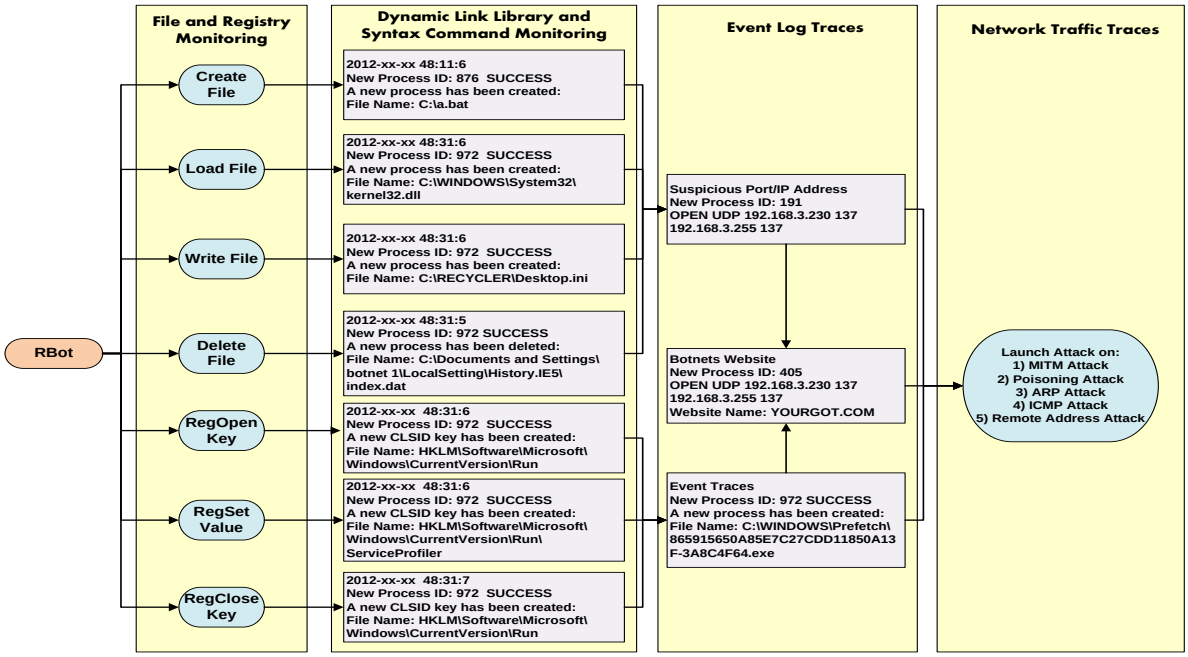

Figure 6: Attack Anatomy in RBot Summary

V. Proposed The Improvement of P2P Botnets Detection Framework Architecture and Component

In this research, the marriage of several P2P botnets detection system has been done in order to improve the functionality and capability of detection system. The architecture is designed based on the finding from the revolution of P2P botnets detection systems. Technically, this detection framework architecture has structured using the multi-layer detection systems which contain input and output layers and four hidden layers that represent the mined data, analyzer, analysis modules and detection tasks. Figure 7 graphically visualizes the architecture of framework. The proposed architecture consists of six main layers and structured with their own components as our previous paper in [33]. Then, each component involves a set of modules and steps. Entire layers describe as followed:

A. Layer 1: Input Layer

First layer describes as the input layer that entail by *Monitoring Module*. In this layer, the dataset has gathered from host level and network level. The data has been labeling through capturing packet in a controlled environment called as P2P testbed network. The host logs categories have involved the system log, application log and security log. Meanwhile, the network log consists of full payload packet for P2P network traffic.

B. Layer 2: Mining Data Layer

Layer 2 describes as the mined data layer involve with *Filtering Module*, *Pre-processing Module* and *Mined-Data Module*. In this layer, filtering is a process to take selectively control the flow of data from network packet where it allows only useful data. Then, the unnecessary data will be taken from the raw data are applied to make the next process easy and smooth. It is set as a data mining that

exclusively denote by pre-processing stage. The selection has been done in selecting the useful attributes and may consume most of the time process. The approach that had be applied in pre-processing stage is k-means clustering and classifying concept which are capable to detect novel attack

without any prior notice and capable to find natural grouping of data based on similarities among the patterns [8]. Overall, the major task implicate in mining data layer are data reduction and data discretization that obtains reduced representation in analytical and numerical results.

C. *Layer 3: Analyzer Layer*

The third layer represents the analyzer layer. In this layer, the hybrid analyzer had been used as a combination of host-level and network-level denote as *Host Analyzer Module* and *Network Analyzer Module*.P2P botnets have some unique characteristics and attacking behaviours that entirely different either in host or network level. The detection of P2P botnets using host behaviours and networks behaviours has their own benefits and limitations. In host-based, the monitoring behaviours have done in a single host and the events occurring within that host for suspicious activity. While in network-based, it will monitors network traffic for particular network segments or devices that analyses the network and application of protocol activity to identify suspicious activity [34]. If the detection is made at one level only probably it hardly provide reliable detection results. Thus, in order to detect bots more effectively, a combined host-based and network-based analysis is needed. These combination levels of analysis are complement each other in finding malicious activity occur in the P2P network. This paper utilizes the combination of both approach simultaneously in differentiating a normal P2P and abnormal P2P behaviours.

D. *Layer 4: Analysis Layer*

Layer 4 enlightens the *Hybrid Analysis Module* and *Attack Pattern Identification Module*. The in-depth analysis provides two levels of analysis approach which are static approach and dynamic approach. The use of both static and dynamic approaches remains as hybrid analysis approach which complements each other disadvantages. Hence, static approach has highlight the capabilities to detect the malicious activity before the program is running or executed while the dynamic approach has the capabilities to detect the malicious activity during or after program execution.

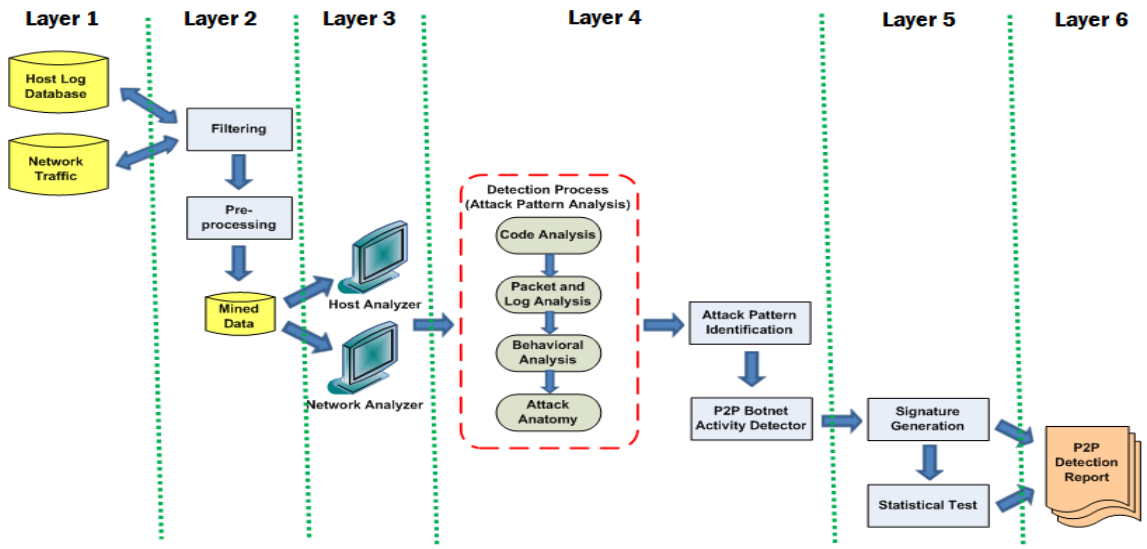

Figure 7: Proposed Improvement of P2P Botnets Detection Framework Architecture [33]

E. *Layer 5: Detection Layer*

The layer fifth expresses the hybrid detection techniques layer indicate the *Signature Generation Module* and *Statistical Test Module*. The hybrid techniques born as the weaknesses and limitation of previous theory have been overcome including the researches towards evaluating the analyzer and detection level. In the other word, hybrid has been choosing to advance their ability. The implementation of hybrid technique used to have maximum accuracy, effectiveness and efficiency in detection rate evaluation. The hybrid technique has using the combination on data mining-based, signature-based and anomaly-based. It has encounter the attributes relate on behavior features either in host-level or network-level. Then, the hybrid technique introduces as improvement in P2P botnets detection system.

Mined data by using algorithm k-Means Clustering to classify a certain number of clusters (assume k clusters) fixed a priori.
The algorithm composed with steps:

[1] Initialize k centers location
[2] Assign each X_i to its nearest Cluster Center C_i
[3] Updates each Cluster Center C_i as the mean of all X_i that have been assigned as closest to it
[4] Recalculate the positions of the k centers

[5] Repeat Steps ii, iii and iv until the centers no longer move
[6] Exit

Algorithm 1: K-Means P2P Clustering

The rule-based method role as a signature-based technique to detect known P2P botnets
[1] Load captured packet and host log (Input= S1, S2)
[2] Initialize S with S_i=normal/abnormal, S_{ii}=classification attack types, S_{iii}=classification sub attack types
[3] If S_i=abnormal then S_i=abnormal;
[4] Else S_i=normal
[5] End
[6] If S_{ii}=classification attack types then S_{ii}=classification attack types;
[7] Else S_{ii}=none;
[8] End
[9] If S_{iii}=classification sub attack types then S_{iii}=classification sub attack types
[10] Else S_{iii}=none;
[11] End
[12] If S_i=abnormal or If S_{ii}=classification attack types or S_{iii}=classification sub attack types then Detected, d=1
[13] Else if S_i=normal then Detected, d=0 (proceed with statistical test on anomaly-based section)
[14] End
[15] End

Algorithm 2: Rule-based detection

This hybrid techniques used as the backbone in proposing the implementation of P2P botnets detection system to modeling the intrusion report. The detection layer has been started by mined the data through pre-processing technique indicates by k-means clustering. Then, the host log and network traffic will be analyzed through the signature-based technique by rule-based module, thus generating an attack alarm if a known pattern is matched. Further statistical tests as anomaly-based revealed on anomalous volume that allow as second detection for the unknown intrusion events. The detection will combine the steps on algorithm 1 followed by algorithm 2 and at last do the algorithm 3 that mainly indicate this procedure as hybrid techniques detection.

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
[3] Calculate statistics of packets distribution, let it with $X^2=\sum (O-E)^2/E$
[4] Exit

Algorithm 3: Statistical Test detection

F. Layer 6: Output Layer

The final layer is describes the *P2P Detection Report Module* represents an individual output of P2P botnets detection system. The P2P detection system can be applied with standard detection method.

Therefore, this research has proposed a new detection in improving the detection in P2P botnets by revealing their behaviors and characteristics within several improvements has been made. Finally, the real study on P2P botnets detection will be compared to the previous systems. Both of the result will be compared. After that, this P2P detection will be developed whereas it is useful for security uses in future.

VI. CONCLUSION AND FURTHER WORK

Considering the evolvement of P2P botnets at the real world, we proposed an improvement of detection framework architecture in order to improve the functionality and capability of current detection system that combines on hybrid analysis; static analysis and dynamic analysis. Our evaluation based on reverse engineering and passive monitoring data has been shown in the following results. With combination of hybrid techniques, hybrid analyzer and hybrid analysis, our framework is able to detect all P2P botnets dataset as early detection before program under inspection is executed.

In further work, we try to make comparison the validation results with one of commercial product and other framework results. The comparison will be made on detection rates and metrics influence the accuracy of detection. In addition, we plan for further research on a real-time detection and mobile environment platforms.

ACKNOWLEDGMENT

The authors would like to express the appreciation to Inforslab Group 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] Donghong, S., et al. The New Architecture of P2P-Botnet. in Cybercrime and Trustworthy Computing Workshop (CTC), 2010 Second. 2010.

[2] Duan J. et al. "Descriptive Model of Peer-to-Peer Botnet Structures." *2010 International Conference on Educational and Information Technology, ICEIT 0201.* (pp. 153-157). Beihang University, China: IEEE, 2010.

[3] Wang P. et. al. "A Systematic Study on Peer-to-Peer Botnets." Central Florida University, USA: IEEE, 2009.

[4] Hossein R. Z. et. al. "A Proposed Framework for P2P Botnet Detection." *IACSIT International Journal of Engineering and Technology, Vol.2, No. 2* (pp. 161-168). IEEE, 2010.

[5] Sun D. et al. "The New Architecture of P2P Botnet." *2010 Second Cybercrime and Thrustworthy Computing Workshop.* (pp. 34-40). Tsinghua University, China: IEEE, 2010.

[6] Liao W. H. & Chang C. C. "Peer to Peer Botnet Detection Using Data Mining Scheme." Tatung University, Taiwan: IEEE, 2010.

[7] Han L., "Research of K-MEANS Algorithm based on Information Entropy in Anomaly Detection", Fourth International Conference on Multimedia Information Networking and Security, 2012.

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

[9] Massi J. et al. "Botnet Detection and Mitigation." *Proceedings of Student-Faculty Research Day, CSIS.* Pace University: IEEE, 2010.

[10] Anil, S. and Remya, R., "A hybrid method based on Genetic Algorithm, Self-Organised Feature Map and Support Vector Machine for better Network Anomaly Detection", ICCCNT, IEEE, 2013.

[11] Zeidanloo, H. R., Hosseinpour, F. and Eternad, F.F.: New Approach for Detection of IRC and P2P Botnet. *International Journal of Computer and Electrical Engineering Vol. 2(No. 6): 1793-8163*, 2010

[12] Zeidanloo, H. R. a. A., A.B.: Botnet Detection by Monitoring Similar Communication Patterns.*(IJCSIS) International Journal of Computer Science and Information Security Vol. 7(No. 3): 36-45*, 2010

[13] Yin, C. and Ghorbani, A.: P2P Botnet Detection Based on Association between Common Network Behaviors and Host Behaviors: IEEE, 2011

[14] Yuanyuan, Z., H. Xin, et al.: Detection of Botnet using Combined Host-and Network-Level Information. *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2010.

[15] Junjie, Z., R. Perdisci, et al.: Detecting Stealthy P2PBotnet Using Statistical Traffic Fingerprints. *IEEE/IFIP 41st International Conference on Dependable Systems & Networks (DSN)*, 2011

[16] Arshad, S., M. Abbaspour, et al.: An anomaly-based Botnet detection approach for identifying stealthy Botnet. *IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE)*, 2011

[17] Muthumanickam, K. and Ilavarasan, E. : P2P Botnet Detection: Combined Host and Network-Level Analysis: IEEE, 2012

[18] Li, H. et al.: P2P Botnet Detection based on Irregular Phased Similarity: IEEE, 2012

[19] Zang, X., et al.: Botnet Detection through Fine Flow Classification. *CSE Department Technical Report CSE11-001*, 2011

[20] Guofei, G., P. Roberto, et al.: BotMiner: Clustering analysis of network traffic for protocol-and-structure-independent Botnet detection. *Proceedings of the 17th conference on Security symposium. San Jose, CA, USENIX Association*,. 2008

[21] Su, C. and E. D. Thomas: P2P Botnet detection using behavior clustering and statistical tests. *Proceedings of the 2nd ACM workshop on Security and artificial intelligence. Chicago, USA, ACM.*, 2009

[22] Sundaram, A. An Introduction to Intrusion Detection. ACM , 2 (4), 3-7, 1996.

[23] Zang, X., et al.: *Botnet Detection through Fine Flow Classification*. CSE Department Technical Report CSE11-001, 2011

[24] Leder et al.: *Proactive Botnet Countermeasures Approach*, 2009

[25] Junfeng et al. : *Descriptive Model of Peer-to-Peer Botnet Structure*, International Conference on Educational and Information Technology (ICEIT), 2010

[26] Chao et al.: *Botnet: Survey and Case Study*, Fourth International Conference on Innovative Computing, Information and Control (ICICIC), 2009

[27] Robiah Y. et al.: *A New Generic Taxonomy on Hybrid Malware Detection Technique*, International Journal of Computer Science and Information Security Vol. 5, no. 1 56-61, 2009

[28] Raihana Syahirah Abdullah, Faizal M.A., Zul Azri Muhamad Noh: Tracing P2P Botnets Behavior via Hybrid Analysis Approach. European Journal Scientific Research, January 2014

[29] Raihana Syahirah Abdullah et al., "Preliminary study of host and network-based analysis on P2P Botnet detection"; *TIME-E Confernece IEEE Bandung, Indonesia*: 2013

[30] Faizal M. A., Mohd Zaki M., Shahrin S., Robiah Y, Siti Rahayu S., Nazrulazhar B.: Threshold Verification Technique for Network Intrusion Detection System. *(IJCSIS) International Journal of Computer Science and Information Security*, Vol. 2, No. 1, 2009

[31] Chandrashekar, J. et al: The Dark Cloud: Understanding and Defending against Botnets and Stealthy Malware. *Intel Technology Journal Vol.13 Issues 2*, 2009

[32] Feily, M., A. Shahrestani, et al.: A Survey of Botnet and Botnet Detection. *Third International Conference on Emerging Security Information, Systems and Technologies (SECURWARE)*, 2009.

[33] Raihana Syahirah Abdullah, Faizal M.A., Zul Azri Muhamad Noh, Mohd Zaki Mas'ud, Siti Rahayu Selamat, Shahrin Sahib: Enhanced P2P Botnets Detection Framework Architecture with Hybrid Analyzer: Host-based and Network-based. Information and Assurance Conference (AIS), Tunisia, IEEE, 2013

[34] Sabahi, F. and Movaghar, A.: Intrusion Detection: A Survey. *The Third International Conference on System and Networks Communication*, 2008

AUTHORS' INFORMATION

Raihana Syahirah Abdullah She is currently a PhD student at Universiti Teknikal Malaysia Melaka. Her research area include computer and network security.

PM Dr Mohd Faizal Abdollah is currently a senior lecturer in Universiti Teknikal Malaysia Melaka. The research area are system communication computer cluster in IDS, malware, forensic and network security

Dr Zul Azri Muhamad Noh is currently a senior lecturer in Universiti Teknikal Malaysia Melaka. The current research interests include advanced networking and distributed system research cluster in quality of service (QoS), wireless LAN, packet scheduling algorithm, and multimedia communication.