B-Corr Model for Bot Group Activity Detection Based on Network Flows Traffic Analysis

Dandy Pramana Hostiadi[1,2*], Waskitho Wibisono[1] and Tohari Ahmad[1]
[1] Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2] Department of Informatics, Institut Teknologi dan Bisnis STIKOM Bali Bali, Indonesia
[e-mail: dandy.17051@mhs.its.ac.id, waswib@if.its.ac.id, tohari@if.its.ac.id]
*Corresponding author: Dandy Pramana Hostiadi

Received March 19, 2020; revised August 1, 2020; accepted September 20, 2020; published October 31, 2020

Abstract

Botnet is a type of dangerous malware. Botnet attack with a collection of bots attacking a similar target and activity pattern is called bot group activities. The detection of bot group activities using intrusion detection models can only detect single bot activities but cannot detect bots' behavioral relation on bot group attack. Detection of bot group activities could help network administrators isolate an activity or access a bot group attacks and determine the relations between bots that can measure the correlation. This paper proposed a new model to measure the similarity between bot activities using the intersections-probability concept to define bot group activities called as B-Corr Model. The B-Corr model consisted of several stages, such as extraction feature from bot activity flows, measurement of intersections between bots, and similarity value production. B-Corr model categorizes similar bots with a similar target to specify bot group activities. To achieve a more comprehensive view, the B-Corr model visualizes the similarity values between bots in the form of a similar bot graph. Furthermore, extensive experiments have been conducted using real botnet datasets with high detection accuracy in various scenarios.

Keywords: Bot group activity, bot activity flows, similar intersection, network security, intrusion detection system

http://doi.org/10.3837/tiis.2020.10.014 **ISSN : 1976-7277**

1. Introduction

Botnet is one of the most dangerous threats, serious problems that need to be handled appropriately. The following are hazardous and disruptive activities of botnets: spamming activities, distributed denial-of-service (DDoS) attacks, identify theft, and other malicious activities [1]. Bot activities have a characteristic that differed from other malware characteristics. Botnet consists of a collection of infected computers called as bots and controlled by botmaster [2]. Bot would then attack a 'victim' computer based on the bot master's command transmitted to all bot clients.

General approaches of botnet activity detection could be performed in the following ways [3]: detection based on signatures, detection with honeypot/honeynet, detection by monitoring DNS traffic from each bot, and detection based on analysis of bot behavior. These approaches usually involve data processing techniques such as histograms [4], clustering approaches [2], [5, 6], and classification approaches [7 – 9]. In the bot activity detection model, the discussion regarding similarities to determine the relationship between the bot attack and the capability to detect the bot preparators were relatively few and incomprehensive. The similarities of bot activities were required to determine the group of attackers (bots) and the number of attacks in the attack scenario. The purposes were to figure out which possible hosts have a vulnerability to be attacked by the bot and to show a similar bots attack pattern.

The correlation alert approach was common to be adapted in the intrusion detection development model. The correlation measurements could show how strong the correlations are between different alerts and relationship between assailants. The correlation alert method was calculated based on Similarity-Based approaches, Sequential-Based approaches, and Case-based approaches [10]. The similarity-based approaches are commonly used due to the convenience of applying similarity analysis—based on the similarity attributes behavior.

In a bot attack scenario inside a computer network, a collection of bots received the command to attack many computers, and it is possible to have a similarity in the attack pattern and the target among those attacked computers. The similarity of bot group attack pattern and the target is called a bot group activity. Some intrusion detection models can detect bot activity, but none has detected bot group activities and explained the relationship between bot activities. The detection for this activity would then enable network administrators to isolate the activity or access from bot group activity. It can be developed in the future to determine the correlation between bots' activities.

Different from the approaches in models [1, 5 – 8, 11], this proposed model could detect suspected bot hosts and measure the similarities between bots by intersection-probability analysis. The process of discovering bot activities was done by the classification machine approach [7, 8, 11]. The process of measuring similarity between bot activities is analyzed using interception-similarity in the following: computers/targets attacked, port address, attack protocol, and total packets. The model proposed in this research is called the B-Corr model. □

This research aimed to determine bot group activity on the network, which resulted from measurements of bots' similarity using intersection-probabilities and made it applicable for network administrators to isolate actions from groups of bots that attack several computers on the network. The B-Corr model can be developed in the future to determine the relationship between bot activities known as correlation bot activities. The models proposed in this study have the following contributions:

- As a proposed model, the B-Corr model analyzes the similarity in each bot's activities based on network flows. B-Corr model analyzes network traffic, in which there were various activities between normal, background, and bot activities. The B-Corr model does not specialize in its analysis only on DNS traffic or HTTP, but also on IRC communication. Hence, it generated optimal measurement results between bot activities on the network.
- B-Corr model, as a proposed model, can find bot group activities by measuring the intersection-probability obtained from feature extraction derived from net flows activity. B-Corr model uses network flowed traffic header and converted it to generate new features to analyze similarities such as inbound, outbound, inbound degree, and outbound degree.
- B-Corr does not rely on detector application tools such as honeypot or malware detection tools to obtain similarity between bots based on the intersection behavior bots concept from network traffic flow analysis.

This paper's structure is presented into several sections; Section 2 unravels previous studies, which has become the foundation of this research. Section 3 elaborates B-Corr as a model proposed for bot group activity detection. Section 4 elaborates on the experiment and evaluation. Section 5 unravels the conclusion of the research.

2. Related Works

Ahmadian et al. [12] constructed an alert correlation model by conducting feature extraction on alert information such as IP address, port address, counter, and alert type. A group of alerts would then be broken down into smaller parts called as an episode. Each occurrence of different alerts type was grouped based on the features' similarity, then the occurrence frequency in each episode was calculated. Similarity measurement was done by calculating the probability of similarity between different alert features. The strength of correlation was calculated using the probability approach and CCM (Causal Correlation Matrix). The result was a series of visualization from attacks that occur in the network. In this research, attack data used were generated from the intrusion detection system using attack scenarios equipped with clear stages.

Ghasemimol et al. [13] constructed an alert correlation model by similarity approach and entropy-based measurement approach. The goal was to display a large number of raw alerts that contained information on a fewer display or general level without removing information (hyper-alerts). The alerts model as an entropy would then be grouped (clustered) based on the information similarity into a cluster alert. Cluster alerts were then optimized by hyper-alerts partial entropy (APE) to generate maximum entropy and build an HG (Hyper Alert Graph). In this research, an evaluation conducted on data attacks was generated from the intrusion detection system.

Both models in [12 - 13] detect malicious activity on a network with different behavior from botnet attacks. Bot attacks are dynamic and have different characteristics from malicious activities in the network. Nevertheless, measuring the similarity between malicious tasks is logically applied to find the similarity between bot activities. It is to detect a bot group activity, and it is possible to calculate the correlation between bot characteristics.

Alvarez et al. [1] conducted a detection of botnet, which has a single activity nature. The detection was done by extracting flow traffic, such as DNS queries, incoming and outgoing packets, data size (data bytes), and IP address. Traffic flow contraction of each feature was

termed as BI (Behavior Instance) and converted into a vector to be processed to detect Botnet activity with a classification approach using SVM (Support Vector Machines). The result then could detect half the spreading of the malicious host. In this research, the detected bot activity group did not have a similar relationship, even though the attack process was similar.

Kwon et al. [14] conducted a detection on group activities from the botnet. The activity process was performed by analyzing activities periodically, where communication activities while receiving and performing instructions from botmaster through C&C communication. In this process, periodic access was obtained frequently for DNS query access. Processing regular activities used a signal processing approach named DFT (Discrete Fourier Transform) as a DNS frequency counter query access. Post regular activities were generated to result in group activities, and measurements were made using a similarity approach. Similarities were applied to look for patterns similar to regular activities and grouped into one group activity. A similarity measurement used was pDist (power Distance) algorithm. It compares periodic structure two input signals of periodic activity. This research focuses on activities that have high frequency and intensity. The fact was that the bots do not always work intensely, but they affect the attack pattern, which records the computer network's activities.

Chowdhury et al. [5] detected botnet activity by representing bot activity in graph-based models. Each host was described as a vertex or node, and its relationship to other hosts was in activity directions. Features extracted from traffic flow include the following: in degree, out-degree, in degree weight, out-degree weight, clustering coefficient, node betweenness, and eigenvector centrality. These features were used as the input data in the clustering process. SOM (Self-organizing Map) method was used to classify each node's activities, which was a representation of host activities in the network. The result from SOM was a value in each cell, it represented a number of nodes, and the cell was a formed cluster. Cluster filters were formed based on the number of nodes in the cell and portrayed to actual implementation. This research enabled bot activity detection and found the number of bots in network flow data between bot activities, normal activities, and background activities. However, this research did not calculate similarities occurred in each interconnected bot.

Choi et al. [15] conducted detection on botnet activity. Detection activities from botnet were done by looking at activities access queries to a specific DNS address, analyzed from the source of IP address information. Query activity to the domain was analyzed periodically on time partitions ran during traffic time. Each array was traffic information containing IP source address, and DNS query address measured for similarity in each other time partition. Any information from the same IP source and DNS query at different partition times were indicated as duplicated access. Activities of each bot in a network were indicated as the access DNS queries to the same destination. Each similarity was calculated as similar activities and categorized as group botnet activities. This research was focused on similarity analysis based on DNS traffic. Still, a real botnet activity was based not only on DNS access analysis and could perform various activities according to the attack's purpose.

Choi et al. [16] developed a weighted measurement for bots traffic record similarity. The purpose of the weighted score was to detect botnet group activities. An analysis was done from the traffic information distribution by taking the source’s IP address information to the IP address destination at each time partition. Each bot activity was distributed by the source of IP address to the IP destination address on the one-time partition. Each distribution activity was compared to other time partitions and given a value of 1 if there were similarities and 0 if there was no similarity. A list of each IP address source to IP destination address on each time partitions was sorted into one vector column order. After each vector was produced, it will be compared to the next vector and calculated by similarity measurements, called Kulczynski,

Cosine, and extended Jaccard similarity (Tanimoto). This study described the similarity between bot activities. However, the similarity was not a measurement to compare the two bots. A similarity between two bots could generate two similar values based on the activity direction behavior.

Eslahi et al. [4] analyzed detection group activities from botnets based on distribution data packets running in time series. Each partitioned time changed to a histogram graph that runs in partition time on hour based, and time sequence partitions measured each histogram graph. Measurement of similarity in each histogram point used the Euclidean Distance for calculation. Assumption group activity from botnet was to have the same size of information periodically and automatically be labeled to indicate group activity detection from botnet. New test data will be tested on this model and use classification approaches such as C4.5, Random Forest, Naive Bayes, Support Vector Machine, and Feedforward Neural Network (FNN). The result was a detection model with a more optimal FNN classification than four other classification methods. This research conducted an activity detection, and the result did not show a similarity of bot behavior in the network. However, it showed a high level of detection in detecting bot activity.

Other previous studies of bot detection models were conducted to detect bot activity and obtain unexpected computer information as a bot in the network. In [4, 15, 16], the activity of a group of bots was detected by looking for similarities based on the concept of time segmentation with an assumption that the bot was active in each segment. The detection model with a concept of correlation [10, 12, 17] was to measure the level of proximity between attack alerts generated from the intrusion detection model and not specified to a bot attack scenario. The bot detection model in [1, 7 – 9, 18, 19] could detect bot activities, but could not measure the relation between bot activities in a group activity attack scenario. In related research of botnet activity detection, it did not measure how strong the relationship of bot activities between bot attacks in a bot group or how similar the bot's behavior in bot group activities. Moreover, other related studies also did not elaborate on an attack pattern by bot group activities to its target.

Aside from previous studies, this proposed model was a measurement model of similarity based on the intersection activities concept to defined bot group activities from network flows traffic. The proposed model, B-Corr, adopts the intersection-probabilities concept to measure the similar bot’s behavior by looking at attack similarity of bots group on a target computer. B-Corr also analyzed the use of paths (port addresses) using protocols and the same value of total packets and grouped them into bot attack groups called bot group activity. The B-Corr model was focused on the similarities measurement of two bots based on bot activity flows detected by the bot activity detection system, and it can be developed to determine the correlation between bots activities.

3. B-Corr Model for Bot Group Activity Detection

In general, bot activity is known as a group of bots that have their behavior inside a network, and this is challenged for a model activity [20], especially in finding relationships between activities. The proposed B-Corr model uses intersection-probabilities to measure the similarity between bot flows activities. B-Corr model detects every host activity suspected of having bot activity using a machine learning classifier. The results from the generated classification become the input data for B-Corr model. Several stages in B-Corr model include a list of suspected IP bots, then extracting flows into features that described activities in each

suspected bot by analyzing header traffic net flows. After generating the features, B-Corr model then measures the intersections-probability of each feature. The results of B-Corr similarity measurement are stored in the Similarity Matrix table. To achieve more comprehensive view, the results of intersection-probability measurements are visualized in a similarity graph. Each of the suspected bots was represented as a vertex, and each vertex was connected through an edge, which has two similarity values based on the intersection relationship point of view. The B-Corr model is shown in **Fig. 1**.

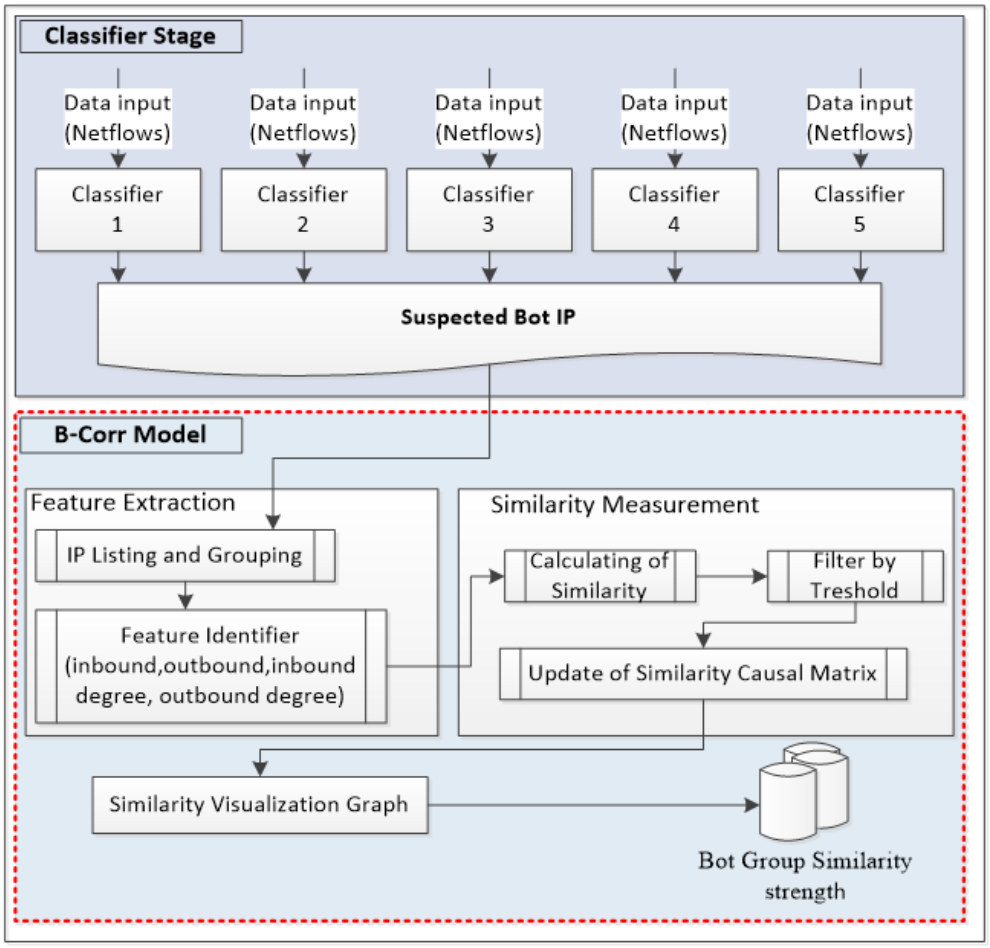

Fig. 1. The Proposed B-Corr Model

3.1 CTU 13 Dataset Overview

CTU is a public dataset owned by Czech Technical University through a Stratosphere IPS Laboratory lab project. The dataset contains network traffic with malware activities at Czech Technical University and recorded in 2011. CTU dataset and captured malware are a dataset consist of botnet flows, normal traffic from normal users, and background network [20 – 22]. The purpose of this dataset is to obtain an actual record of botnet malware activity. The CTU-13 dataset consists of thirteen records with activities (called scenarios) that originated from different botnet examples. Each scenario runs a different particular botnet malware, in which several protocols use and perform different actions. Some studies use the CTU-13 dataset to model botnet activity [5, 19 – 23].

CTU datasets in each scenario have a different number of bots. In CTU-9, CTU-10, CTU-11, and CTU-12 datasets, there are more than one active bots performing attacks with the same type of bot. This research states that each dataset consists of a similar group of bot attacks from some computers. In contrast, CTU-1, CTU-2, CTU-3, CTU-4, CTU-5, CTU-6, CTU-7, CTU-8, and CTU-13 datasets are attacked by bot on several computers. The description of the CTU dataset is shown in **Table 1**.

Table 1. CTU-13 dataset description

ID	Duration (hours)	Bots Count	Bots Name	Total NetFlows	Botnet Flows	Normal Flows	Background Flows
1	6.15	1	Neris	2.824.637	39.933 (1.41%)	30.387(1.07%)	2.753.290(97.47%)
2	4.21	1	Neris	1.808.123	18.839(1.04%)	9.120(0.5%)	1.778.061(98.33%)
3	66.85	1	Rbot	4.710.639	26.759(0.56%)	116.887(2.48%)	4.566.929(96.94%)
4	4.21	1	Rbot	1.121.077	1.719(0.15%)	25.268(2.25%)	1.094.040(97.58%)
5	11.63	1	Virut	129.833	695(0.53%)	4.679(3.6%)	124.252(95.7%)
6	2.18	1	Menti	558.920	4.431(0.79%)	7.494(1.34%)	546.795(97.83%)
7	0.38	1	Sogou	114.078	37(0.03%)	1.677(1.47%)	112.337(98.47%)
8	19.5	1	Murlo	2.954.231	5.052(0.17%)	72.822(2.46%)	2.875.282(97.32%)
9	5.18	10	Neris	2.753.885	179.880(6.5%)	43.340(1.57%)	2.525.565(91.7%)
10	4.75	10	Rbot	1.309.792	106.315(8.11%)	15.847(1.2%)	1.187.592(90.67%)
11	0.26	3	Rbot	107.252	8.161(7.6%)	2.718(2.53%)	96.369(89.85%)
12	1.21	3	NSIS.ay	325.472	2.143(0.65%)	7.628(2.34%)	315.675(96.99%)
13	16.36	1	Virut	1.925.150	38.791(2.01%)	31.939(1.65%)	1.853.217(96.26%)

3.2 Classifier Stage

Bot's activity detection process is required to obtain botnet group activity. CTU-13 dataset is a dataset containing bot activity. This research detects bot activity in CTU-1 scenario dataset through CTU-13 scenario using machine learning classification. The classification approach is adopted from previous studies [7, 8, 11], which introduced several classification methods such as K-Nearest Neighbor, Gaussian Naïve Bayes, Logistic Regression, and Decision Tree and Support Vector Machine. This method was chosen since classification results have a high accuracy degree in botnet activity detection. The process of classification is shown in **Fig. 2**.

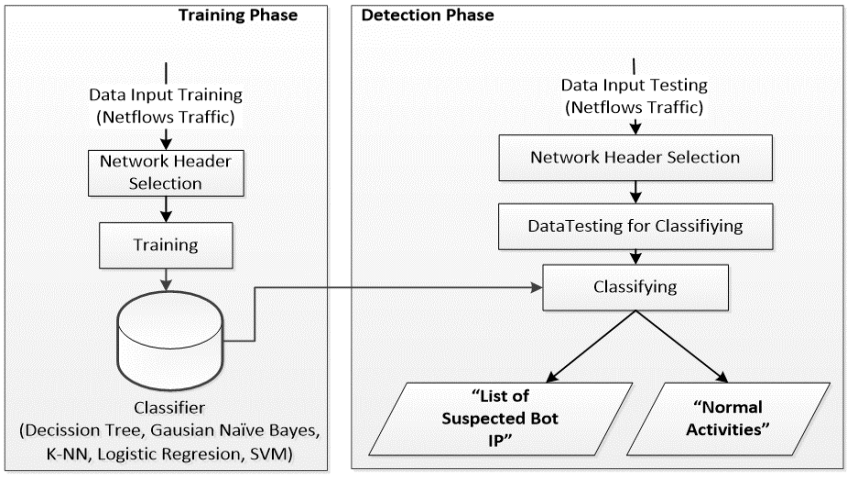

Fig. 2. Classifier processing to produce suspected Bots list

To classifying, 8 of 14 network headers were chosen as flow headers. They are duration, protocol, source port, destination port, source IP, destination IP, total packets, and total bytes. The choice of header network flows based on a study [24], which resulted in high degree accuracy in bot activity detection. At the classification stage, it took Network Flows Traffic (*FT*) consisting of a set of headers. Network headers used were Duration (*Dur*), Protocol (*P*), Source port (S_{port}), Destination Port (D_{port}), Source IP (S_{IP}), Destination IP (D_{IP}), Total Packets (T_{pkts}) and Total Bytes (T_{byts}) and they were defined as:

$$FT = \{D_{ur}, P, S_{port}, D_{port}, S_{IP}, D_{IP}, T_{pkts}, T_{byts}\} \quad (1)$$

The result of the detection phase has generated a list of suspected bots contained in each of the CTU-13 dataset attack scenarios. From the list that was detected as a bot, it is used as an input data on B-Corr model. B-Corr model calculates the similarity between bot flows activities by introducing the concept of an intersection-probability to obtain activity from a group of similar bots called bot group activity. B-Corr model produces similarity values and its strength between bots in group activities that attack the target computer.

3.3 Definition and Notation of B-Corr Model

B-Corr model applies an analysis of network traffic flows. Network traffic has a network header. Let Netflow traffic be set as group records, and Netflow traffic be denoted as *FT* in the model as :

$$FT = r_K; K \in R\,; K = \{1, 2, 3, \dots, j\} \quad (2)$$

where r_k was a collection of suspected bot activities records, *K* was an index of records activity with natural numbers. Each record (r_k) has a network header (∂) and was defined as

$$\partial \in r_k\,;\ \partial = \{S_{IP}, D_{IP}, S_{port}, D_{port}, P, T_{pkts}\} \quad (3)$$

where $S_{IP} = \{s_{IPm}\}_{m=1,\dots,n}$ was a set collection of source IP addresses, $D_{IP} = \{d_{IPm}\}_{m=1,\dots,n}$ was a set collection of destination IP address, $S_{port} = \{s_{portm}\}_{m=1,\dots,n}$ was a set collection of source port address, $D_{port} = \{d_{portm}\}_{m=1,\dots,n}$ was a set collection of destination port address, $P = \{p_m\}_{m=1,\dots,n}$ was a set collection of protocol and $T_{pkts} = \{t_{pktsm}\}_{m=1,\dots,n}$ was a set collection of total packets. If it were declared, each network header in the index (*i*), network header (∂), was defined :

$$\partial \in r;\ \partial_{(i=6)} \begin{cases} i_1 \rightarrow \partial_1 = S_{IP}; \\ i_2 \rightarrow \partial_2 = D_{IP}; \\ i_3 \rightarrow \partial_3 = S_{port}; \\ i_4 \rightarrow \partial_4 = D_{port}; \\ i_5 \rightarrow \partial_5 = P; \\ i_6 \rightarrow \partial_6 = T_{pkts}; \end{cases} \quad (4)$$

3.4 B-Corr Feature Extraction

From the classification process at the bot detection phase, it extracts new features in the measurement of similarity by selecting the following network headers: protocol, source port, destination port, source IP, destination IP, total packets. Each element of network header ($\partial_{(i=6)}$) was extracted to gain feature, denoted as value $Feature\partial$. In addition, for each element of $\partial_{1(m)}$ with the following names; inbound ($In_{(bound)}$), outbound ($Out_{(bound)}$), outbound degree ($Out_{(bound\ degree)}$), and inbound degree($In_{(bound\ degree)}$), it is defined as :

$$Feature\partial_{1(m)} = \{In_{bound_{\partial 1(m)}}, Out_{bound_{\partial 1(m)}}, In_{bound\ degree_{\partial 1(m)}}, Out_{bound\ degree_{\partial 1(m)}}\} \quad (5)$$

This research showcases each suspected bot from classifier results as communication between nodes. Communication of each bot could be illustrated in a graph, where bots interact as vertex and communication flow as edges [5]. It assumes each bot as a vertex connected through the edge. Each vertex, represented as a suspected bot, has two types of flows. They are inbound flows and outbound flows. If the suspected bot is a vertex, then inbound flows are flows that enter a vertex. Where outbound flows are flows that come out of a vertex. Inbound and outbound flows can be developed into inbound and outbound degrees.

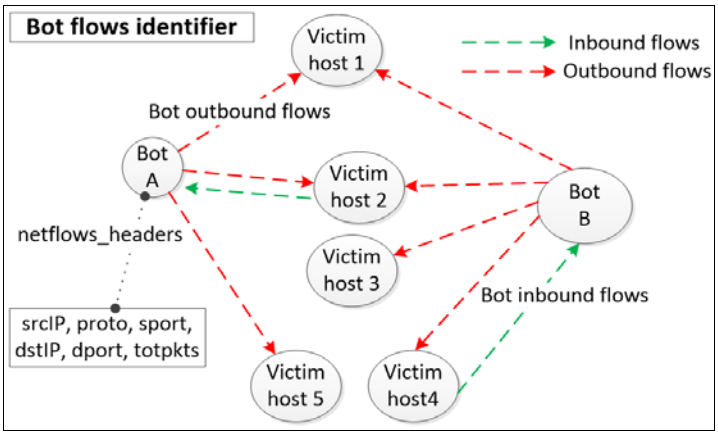

Fig. 3. Bot flows identifier

In **Fig. 3**, bot A has outbound value 3 and inbound value 1. For example, bot A has 20 repeat flows to victim host 1. To victim two are 33, and to victim five are 10, and recurring flows from victim host 2 to bot A were 13. Then, bot A's outbound degree value is 63, and bot A's inbound degree value is 13.

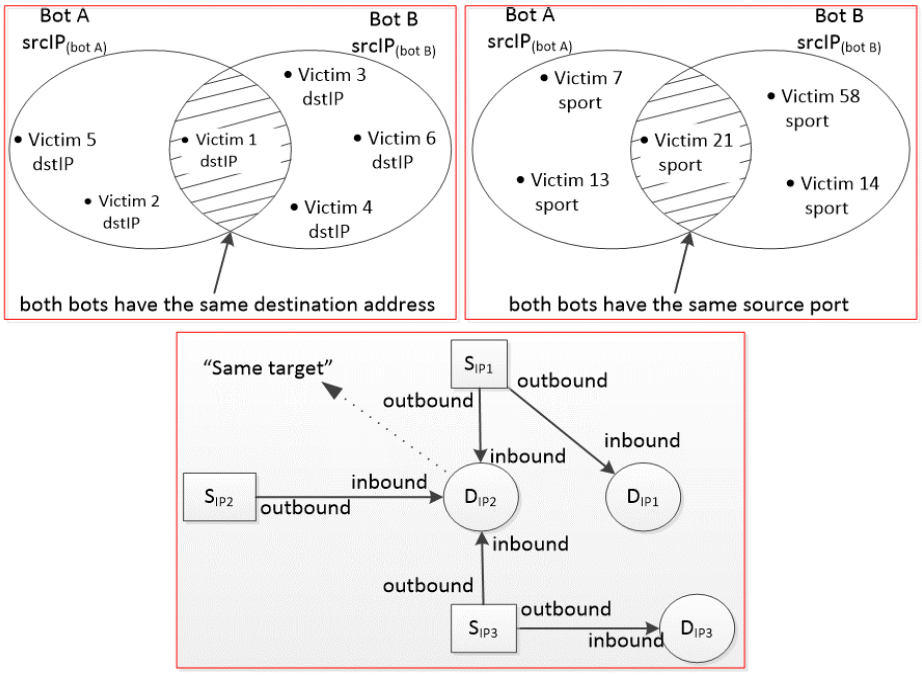

Fig. 4. The identified feature between Bots behavior

Inbound, outbound, inbound degree and outbound degree values are basic features used to calculate the relationship between bots based on similar flows. There was a strong suspicion that every suspected host would have slice flows against a victim host's attack. **Fig. 4** shows that sliced flows can be in the form of net flows header slices called destination IP, source port, destination port, protocol, and total packets.

3.5 B-Corr Similarity Measurement

Similarity measurement was a phase to measure activity among bots obtained from the bot behavior analysis process in the extraction process feature. It calculates each similarity of behavior, such as calculation from the similarity of destination IP addresses, source port addresses, destination port addresses, protocols, and total packets between two different source IP addresses. Comparisons between the two bots adopted a matrix correlation approach [17].

Similarity measurement is completed by two elements value comparison of ∂_i , which is (∂_{if} , ∂_{it}). Then, the intersecting value of feature element $\partial_{1(m)}$ was identified. Similarity measurements defined as intersection-probabilities feature $\partial_{1(m)}$ for each network header,

$$\boldsymbol{sim}\ \partial_{f,t|_i} = \frac{n(\partial_f \cap \partial_t)}{n(\partial_t)}\ , \boldsymbol{f \neq t} \tag{6}$$

where *f* and *t* are compared element values of *i*. *i* is a value of network header that has inbound ($In_{(bound)}$), outbound ($Out_{(bound)}$), outbound degree ($Out_{(bound\ degree)}$), and inbound degree($In_{(bound\ degree)}$) and element values of $\partial_{if} \neq \partial_{it}$. Equation result (equation 6) produced one-way measurements, where if first $S_{IP(f)}$, which is a suspected bot A, and $S_{IP(t)}$, which is a suspected bot B, produce similarity of suspected bots A→B. This intersection value $\partial_f \cap \partial_t$ was proportional to the value of ∂_t. Besides, for the measurements of bot B→A similarity, the intersection comparison value $\partial_t \cap \partial_f$ was ∂_f.

Referring to equation (6) measurement similarity of destination IP address (D_{IP}) header, in which the amount of data intersection of *Feature* $\partial_{2(f,\ inbound)} \cap$ *Feature* $\partial_{2(t,\ inbound)}$ was calculated as follows:

$$\boldsymbol{A \rightarrow B;}\ sim\ \partial_{f_{D_{IP}}, t_{D_{IP}}|S_{IP}} = \frac{n\left(\partial_{f_{D_{IP(inbound)}}} \cap \partial_{t_{D_{IP(inbound)}}}\right)}{n\left(\partial_{t_{D_{IP(in_{bound\ degree})}}}\right)}\ , \boldsymbol{f \neq t} \tag{7}$$

$$\boldsymbol{B \rightarrow A;}\ sim\ \partial_{t_{D_{IP}}, f_{D_{IP}}|S_{IP}} = \frac{n\left(\partial_{t_{D_{IP(inbound)}}} \cap \partial_{f_{D_{IP(inbound)}}}\right)}{n\left(\partial_{f_{D_{IP(in_{bound\ degree})}}}\right)}\ , \boldsymbol{f \neq t} \tag{8}$$

Source port address measurement was performed to highlight the similarity from the number of suspected bots when attacking by using the same port on bot activity type using the same port to attack the target host. Similarity measurement over source port header was done by calculating the amount of data intersection from the following *Feature* $\partial_{3(f,\ outbound)} \cap$ *Feature* $\partial_{3(t,\ outbound)}$:

$$\boldsymbol{A \rightarrow B;}\ sim\ \partial_{f_{S_{port}}, t_{S_{port}}|S_{IP}} = \frac{n\left(\partial_{f_{S_{port(outbound)}}} \cap \partial_{t_{S_{port(outbound)}}}\right)}{n\left(\partial_{t_{S_{port(out_{bound\ degree})}}}\right)}\ , \boldsymbol{f \neq t} \tag{9}$$

$$\boldsymbol{B \rightarrow A};\ sim\ \partial_{t_{S_{port}}, f_{S_{port}} | S_{IP}} = \frac{n\left(\partial_{t_{S_{port(outbound)}}} \cap \partial_{f_{S_{port(outbound)}}}\right)}{n\left(\partial_{f_{S_{port(out_{bound\ degree})}}}\right)}, \boldsymbol{f \neq t} \quad (10)$$

Bot behavior activity analysis during target host attack, in general, utilized weakness of active port on the target side. This allowed the activity from several bots to find the target active port's weakness. Hence, the rationale used to measure similar bot's activity was based on the attack on the destination port. The amount of data intersection used from $Feature\partial_{4(f,\ inbound)} \cap Feature\partial_{4(t,\ inbound)}$ is as follows:

$$\boldsymbol{A \rightarrow B};\ sim\ \partial_{f_{D_{port}}, t_{D_{port}} | S_{IP}} = \frac{n\left(\partial_{f_{D_{port(inbound)}}} \cap \partial_{t_{D_{port(inbound)}}}\right)}{n\left(\partial_{t_{D_{port(in_{bound\ degree})}}}\right)}, \boldsymbol{f \neq t} \quad (11)$$

$$\boldsymbol{B \rightarrow A};\ sim\ \partial_{t_{D_{port}}, f_{D_{port}} | S_{IP}} = \frac{n\left(\partial_{t_{D_{port(inbound)}}} \cap \partial_{f_{D_{port(inbound)}}}\right)}{n\left(\partial_{f_{D_{port(in_{bound\ degree})}}}\right)}, \boldsymbol{f \neq t} \quad (12)$$

A set of active bots that attacked the target host collection using the same protocol. Similar to the attack on DNS, bots tend to utilize UDP protocol. Some of the bots which intensely interacted using protocols are vulnerable and difficult to trace by security systems. To measure the similarity of bot activities, an analysis of similarity protocol is used from the suspected bot. The amount of data intersection from $Feature\partial_{5(f,\ inbound)} \cap Feature\partial_{5(t,\ inbound)}$ was calculated as :

$$\boldsymbol{A \rightarrow B};\ sim\ \partial_{f_P, t_P | S_{IP}} = \frac{n\left(\partial_{f_{P_{outbound}}} \cap \partial_{t_{P_{outbound}}}\right)}{n\left(\partial_{t_{P_{out_{bound\ degree}}}}\right)}, \boldsymbol{f \neq t} \quad (13)$$

$$\boldsymbol{B \rightarrow A};\ sim\ \partial_{t_P, f_P | S_{IP}} = \frac{n\left(\partial_{t_{P_{outbound}}} \cap \partial_{f_{P_{outbound}}}\right)}{n\left(\partial_{f_{P_{out_{bound\ degree}}}}\right)}, \boldsymbol{f \neq t} \quad (14)$$

In addition to the IP address, port address, and protocol information, total packets are the determinants of similarity between bots' activity. In general, bots with similar types have the same number of packets or have an immediate difference. For example, in a DoS attack, a certain number of packets would flood in according to the bot master's command. Then, each connects bot client would attack according to the bot master's instructions. It was why the research was based on an analysis of suspected bot's behavior similarity. Hence, the amount of data intersection was counted from $Feature\partial_{6(f,\ inbound)} \cap Feature\partial_{6(t,\ inbound)}$ as :

$$\boldsymbol{A \rightarrow B};\ sim\ \partial_{f_{T_{pkts}}, t_{T_{pkts}} | S_{IP}} = \frac{n\left(\partial_{f_{T_{pkts(outbound)}}} \cap \partial_{t_{T_{pkts(outbound)}}}\right)}{n\left(\partial_{t_{T_{Pkts(out_{bound\ degree})}}}\right)}, \boldsymbol{f \neq t} \quad (15)$$

$$\boldsymbol{B \rightarrow A};\ sim\ \partial_{t_{T_{pkts}}, f_{T_{pkts}} | S_{IP}} = \frac{n\left(\partial_{t_{T_{pkts(outbound)}}} \cap \partial_{f_{T_{pkts(outbound)}}}\right)}{n\left(\partial_{f_{T_{Pkts(out_{bound\ degree})}}}\right)}, \boldsymbol{f \neq t} \quad (16)$$

The total similarity between two bots was stored in the Similarity Matrix table. The results of each A→B and B→A resemblance were listed, total similarities were calculated as (ω):

$$\boldsymbol{\omega} = \frac{1}{c} \sum_{c=1}^{c} \boldsymbol{sim}(\partial_{f,t|_i})_c \tag{17}$$

For each sim value $\partial_{f,t}$ from each i, it was determined by threshold value which states that the two elements ∂_{if} and ∂_{it} are similar with the following rules:

$$(\partial_{if}, \partial_{it}) \begin{cases} similar\ if\ sim\ \partial_{f,t|_i} > threshold \\ similar\ if\ sim\ \partial_{f,t|_i} < threshold \end{cases}, where\ \partial_{if} \neq \partial_{it} \tag{18}$$

Each obtained similarity was accumulated and compared to total flows contained in each compared bot. The result of the comparison obtained value of two-way similarity between two bots. The first bot was called bot a1, and the second bot was bot a2. Then, the produced similarity measurement was the similarity between bot a1 to bot a2, and bot a2 to bot a1. The value threshold was used to determine the similarity strength between two bots, where bots that have strength similarity values were considered similar bots.

3.6 B-Corr Similarity Visualization graph

In this approach, the similarity between bots activities is visualized as a similarity visualization graph and shows the bot relationship based on similarity. Each bot became vertex and was connected through the edge. Hence, the value of graph (G) was defined as $G = \{V,E\}$, where V was the collection of bots (S_{IP}), and E was the collection of similarity values ($corr\ \partial_{f,t|i}$) between bots.

Values between two bots with high similarity values in a range of 0-1 could have one or two different values when illustrated in graph node. For example, there are three bots connected, as shown in **Fig. 5**, called bot A, bot B, and bot C. From bot A node, bot A has a similarity value of 0.92 to bot B and a similarity value of 0.91 to bot C. Meanwhile, bot B has a similarity value of 0.87 to bot A and bot C has a similarity value of 0.84 to bot A. This indicated that bot A was a bot that interacts strongly with both bots; bot C and bot B. However, the similarity value between bot B and C are not as strong as the similarity with bot A.

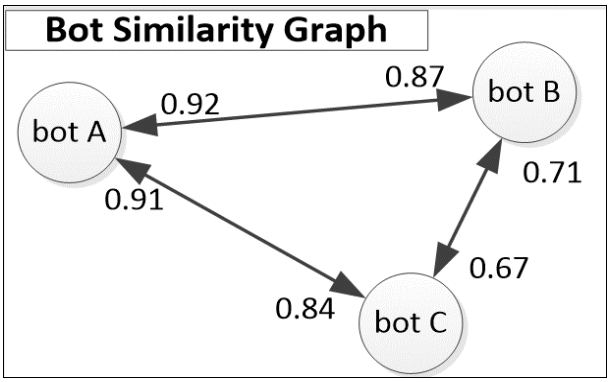

Fig. 5. Similarity measurement between two Bot

4. Experiment and Evaluation

Botnet activity detection model and the process of measuring the similarity in activity between bots on large data sets have challenges. Besides, the process of measuring and

detecting activities requires sufficient computing resources to support data processing. This research employs the CTU-13 dataset, which consists of 13 bot scenarios with a large amount of total network flow. Therefore, this study uses a computer with an Intel Core i-7-8700 CPU 3.20GHz, 16GB RAM, 250SSD storage, and the Python programming language version 3.7 to detect bot group activity.

4.1 Bot Detection

As explained in Section 3, this research adopts a bot detection model in [7, 8, 11] to detect bot activities that used several methods Decision Tree (DT), Gaussian Naïve Bayes (GNB), K-NN, Logistics Regression (LR), and SVM methods. The classification results of these five methods are shown in **Table 2**.

Table 2. Classification result of several methods

Methods	Accuracy (%) Detection in CTU Dataset Scenario												
	1	2	3	4	5	6	7	8	9	10	11	12	13
DT	89.43	88.95	90.64	91.54	93.00	88.16	88.57	90.63	90.65	92.16	92.19	89.02	93.15
GNB	90.27	91.61	87.33	92.17	90.04	86.55	88.79	91.98	91.39	88.76	90.63	93.13	89.88
K-NN	91.20	90.32	88.01	94.44	92.18	89.64	88.87	91.85	89.94	87.89	91.71	91.59	91.75
LR	89.61	90.19	90.51	89.94	86.29	90.73	90.03	93.70	89.85	91.30	90.03	92.82	90.68
SVM	93.69	90.15	92.73	91.64	88.28	88.62	92.73	89.89	91.83	89.78	86.68	89.79	88.93

In this research, the accuracy of each classification method from bot activity is illustrated in the graphic shown in **Fig. 6**, and it is shown that the best classification model was the K-NN (K-Nearest Neighbor) method with the highest accuracy of 94.44% in bot attack detection of the CTU-4 scenario. The lowest accuracy was 86.29% in bot attack detection of the CTU-5 scenario using the Support Vector Machine (SVM) classifier method. The average detection accuracy for all 13 scenarios was 90.72% in the CTU-13 dataset. This showed that the classification process for bot attack activities detection was quite effective, with the lowest accuracy of 86.29% and average classification result of 90.72%.

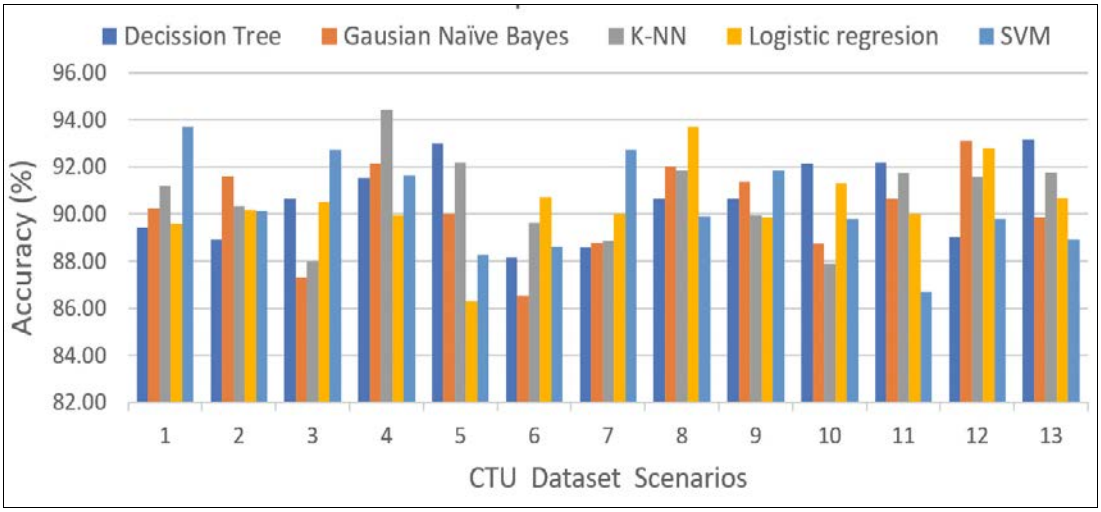

Fig. 6. Classification result

4.2 B-Corr Feature Extraction

After the classification process was completed, list of host classified as bots were generated. Network header which were $S_{IP}=\{s_{IPm}\}_{m=1,\ldots,n}$, $D_{IP}=\{d_{IPm}\}_{m=1,\ldots,n}$, $S_{port}=\{s_{portm}\}_{m=1,\ldots,n}$, $D_{port}=\{d_{portm}\}_{m=1,\ldots,n}$, $P=\{p_m\}_{m=1,\ldots,n}$ and $T_{pkts}=\{t_{pktsm}\}_{m=1,\ldots,n}$, obtained and calculated for each element value of $\partial_{1(m)}$ to become new feature set which were $In_{(bound)}$, $Out_{(bound)}$, $Out_{(bound\ degree)}$, and $In_{(bound\ degree)}$. Extraction result of new feature set shown in **Table 3**.

Table 3. Feature extraction for similarity measurement (Ex. CTU-9_IP_list)

IP	Bot inbound Flows	Bot outbound Flows	Bot inbound_degree Flows	Bot outbound_degree Flows
147.32.84.100	8	3	8	2
147.32.84.165	40	2553	63	22792
147.32.84.191	39	3056	54	18775
147.32.84.192	39	2804	129	20305
147.32.84.193	41	2691	65	17961
147.32.84.20	12	5	12	19
147.32.84.204	41	2512	56	18784
147.32.84.205	46	2711	62	17536
147.32.84.206	37	2206	50	18553
147.32.84.207	38	2535	45	16002
147.32.84.208	38	2444	134	17909
147.32.84.209	39	2991	62	16379
147.32.84.100	8	4	8	3
147.32.84.165	40	2553	63	22792

4.3 B-Corr Similarity Measurement

The threshold value is used to determine the similarity between bots. Measurement of the threshold value is done by measuring the similarity in a real CTU dataset containing bot traffic activity and comparing it with the results of similarity measurements on the B-Corr model. The determining of the threshold value in the CTU dataset scenario is shown in **Fig. 7**.

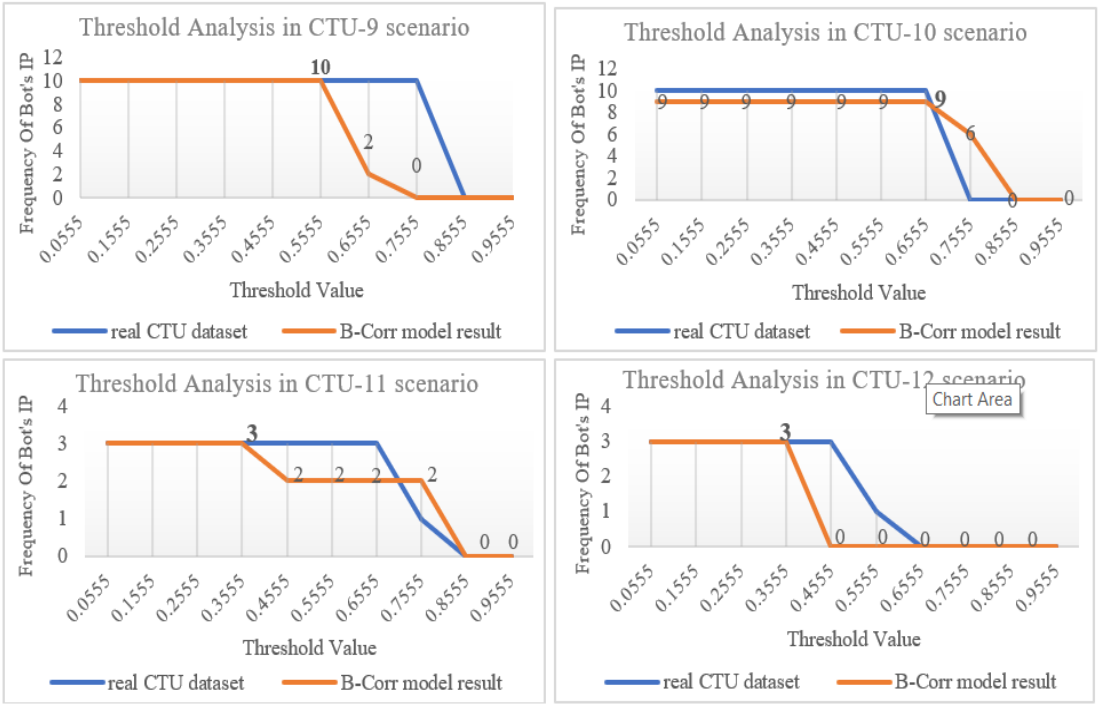

Fig. 7. Threshold analysis for CTU dataset scenario

The threshold values tracking is only obtained in scenario datasets of CTU-9, CTU-10, CTU-11, and CTU-12. This is due to the scenario dataset consisted of more than one bot. In the real CTU dataset, each amount of bot's IP is calculated in the range of threshold values with a width of the threshold value of 0.1, and the range of the threshold value is 0.0555 to 0.9555. The threshold value in the real CTU dataset with the highest number of bot's IP used as the range of threshold values in the B-Corr model. The largest threshold value from the B-Corr model results, which has the highest number of bot's IP is a valid threshold value and used as a similarity threshold value for bot activities. From the tracing of the threshold value that has been carried out, each CTU-9, CTU-10, CTU11, and CTU-12 dataset scenario has different threshold values. The threshold values are shown in **Table 4**.

Table 4. Threshold values for each CTU dataset scenario

Dataset Scenario	Threshold Value	Number Bots IP (B-Corr Model)
CTU-9	0.5555	10
CTU-10	0.6555	9
CTU-11	0.35555	3
CTU12	0.3555	3

The similarity measurements of bot flow activities using threshold values in **Table 4** produce a similarity matrix table. An example of measuring the B-Corr model's similarity in the CTU-9 scenario is shown in **Fig. 8**.

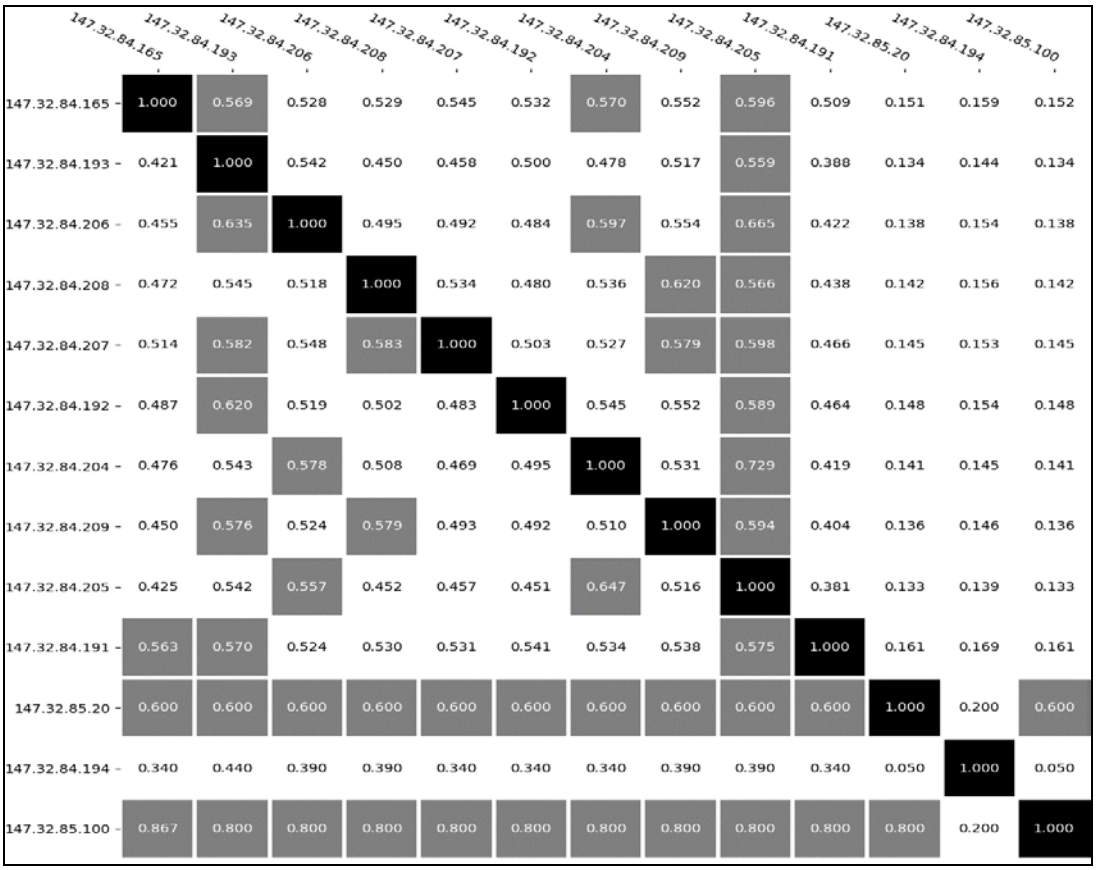

Fig. 8. B-Corr similarity matrix table (CTU -9 Dataset)

The similarity between bots tracking is obtained only in scenario datasets of CTU-9, CTU-10, CTU-11, and CTU-12. Tracking results in the activity of a group of bots that has a similarity are shown in **Table 5**.

Table 5. B-Corr model result

CTU Dataset	Number of Bot	IP Address of Bot	B-Corr Model	
			Similar activities	Non-similar activities
9	12	147.32.84.165, 147.32.84.191, 147.32.84.192, 147.32.84.193, 147.32.84.204, 147.32.84.205, 147.32.84.206, 147.32.84.207, 147.32.84.208, 147.32.84.209, 147.32.84.100, 147.32.84.20	47	109
10	10	147.32.84.165, 147.32.84.191, 147.32.84.192, 147.32.84.193, 147.32.84.205, 147.32.84.206, 147.32.84.207, 147.32.84.208, 147.32.84.209, 147.32.85.76	81	29
11	3	147.32.84.165,147.32.84.191, 147.32.84.192	46	129
12	3	147.32.84.165,147.32.84.191, 147.32.84.192	32	40

In those CTUs' scenario datasets, an analysis is done to obtain the behavior from bot group activity. The analysis was completed by observing an intersection between bot groups on the same computer target. The analysis is shown in **Fig. 9**.

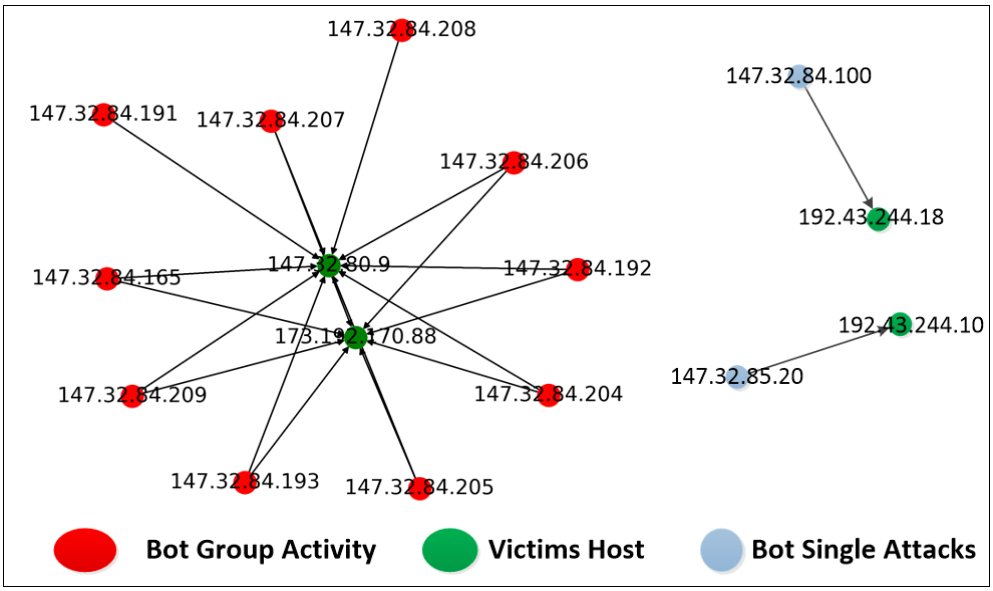

Fig. 9. Bot group activity

Fig. 9 shows an example of a bot activity group search in the CTU-9 scenario dataset. It could be noticed that there were two attack targets by a set of bots on the computer target with an IP address of 147.32.80.9, an IP address of 173.192.70.88. Two of these targets were the targets of a set of bots with the following IP addresses: 147.32.84.165, 147.32.84.191, 147.32.84.192, 147.32.84.193, 147.32.84.204, 147.32.84.205, 147.32.84.206, 147.32.84.207, 147.32.84.208, 147.32.84.209. Two target computers with IP addresses 147.32.80.9 and 173.192.70.88 have the same vulnerability to be attacked by a group of bots called bot group activity. Besides, the bots with IP address 147.32.85.20 and 147.32.85.100 are not constituted by bot group activity due to the attack on the target computer that did not have any similarity in the targeted attack but was a bot attacking activity.

4.4 B-Corr Similarity Visualization graph

A similarity visualization graph is a graph used to describe the similarities in some bots that have similarities. The value of this similarity is illustrated in the form of a similarity node. The example of similarity graph visualization between bots in scenario dataset CTU-9 is shown in **Fig. 10**.

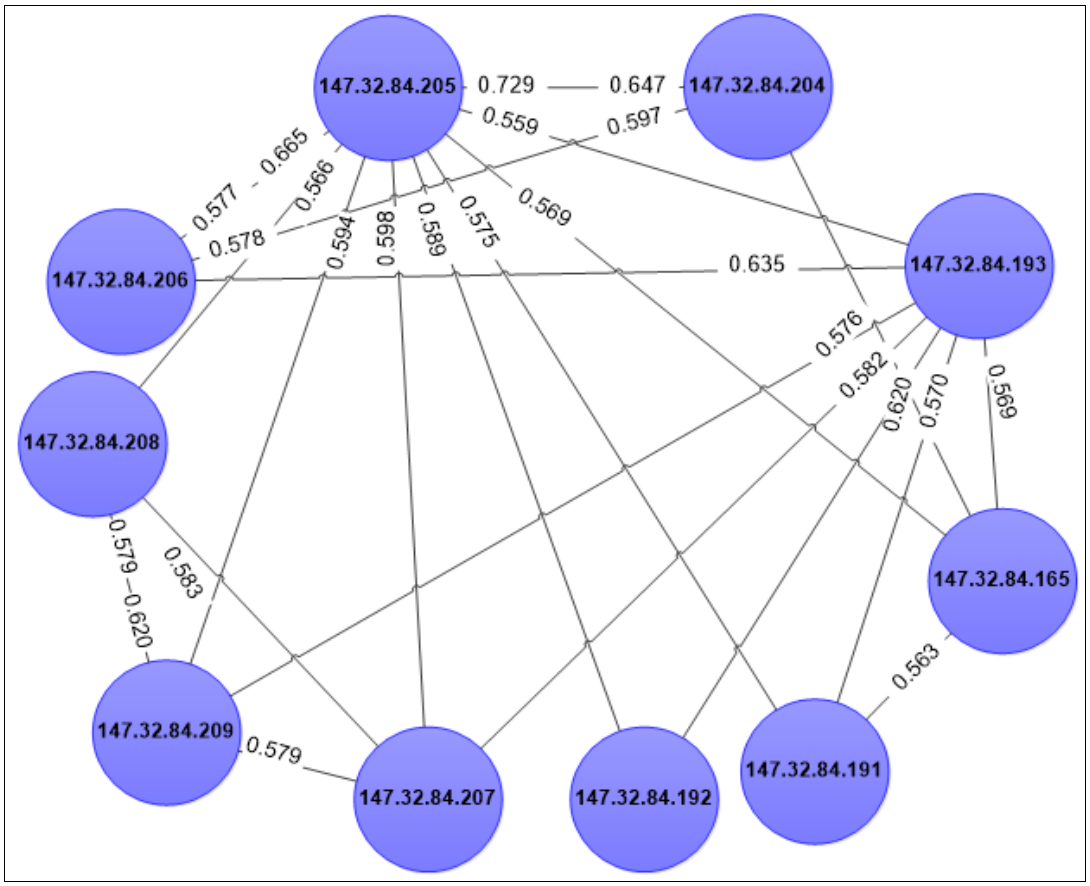

Fig. 10. Bot similarity visualization graph CTU_9

4.5 Analysis with CTU-13 Dataset

Bot group activities attack was performed by a group of bots with similarity values with a similarity approach. The similarity means that there was a resemblance of similar attacks and methods to the target. Bot group activities in CTU-9, CTU-10, CTU 11, and CTU-12 scenario datasets were obtained in this research. In another scenario dataset, no bot group activity was found due to other datasets that did not consist of more than one bot attack. Hence, the calculation of similarity between bots could not be measured in scenario datasets CTU-1, CTU-2, CTU-3, CTU-4, CTU-5, CTU-6, CTU-7, CTU-8, and CTU-13.

To determine the B-Corr model's performance, the IP address from the B-Corr model results was compared with the IP bot list contained in CTU datasets (CTU-9, CTU-10, CTU-11, and CTU-12). The result of searching the IP address list for bot group activity is shown in **Table 6**.

Table 6. Experiment between CTU description and B-Corr model result

Dataset CTU Scenario Numbers		**9**	**10**	**11**	**12**
Real CTU-13 Activities	Number of Bot	10	10	3	3
	Bot IP Address	147.32.84.165, 147.32.84.191, 147.32.84.192, 147.32.84.193, 147.32.84.204, 147.32.84.205, 147.32.84.206, 147.32.84.207, 147.32.84.208, 147.32.84.209	147.32.84.165, 147.32.84.191, 147.32.84.192, 147.32.84.193, 147.32.84.204, 147.32.84.205, 147.32.84.206, 147.32.84.207, 147.32.84.208, 147.32.84.209	147.32.84.165, 147.32.84.191, 147.32.84.192	147.32.84.165, 147.32.84.191, 147.32.84.192
B-Corr Model Results Activities Detections	Number of Bot	12	10	3	3
	IP Address of Bot Group Activities	147.32.84.165, 147.32.84.191, 147.32.84.192, 147.32.84.193, 147.32.84.204, 147.32.84.205, 147.32.84.206, 147.32.84.207, 147.32.84.208, 147.32.84.209	147.32.84.165, 147.32.84.191, 147.32.84.192, 147.32.84.193, 147.32.84.205, 147.32.84.206, 147.32.84.207, 147.32.84.208, 147.32.84.209	147.32.84.165, 147.32.84.191	147.32.84.165, 147.32.84.191, 147.32.84.192
	IP Address of non-Bot Group Activities	147.32.84.100, 147.32.84.20	147.32.85.76	147.32.87.220	-

Table 6 above shows that the B-Corr model in the CTU-9 scenario dataset generated twelve similar bot IP addresses. Then, the twelve similar bot IP addresses were compared with the CTU-9 dataset description. The comparison search result found that in the CTU-9 scenario, the B-Corr model could detect ten lists of IP addresses as a bot group activity. In the CTU-10 B-Corr scenario dataset, the model acquired nine similar Bot IP addresses. Comparing the nine similar IP address bots, CTU-10 dataset description, and comparison results, the B-Corr model could detect nine IP address lists as a bot group activity. In the CTU-11 scenario dataset, the B-Corr model obtained three similar bot IP addresses. From the three similar IP addresses, bots compared to the CTU-11 dataset description, the comparison search results found that the B-Corr model was able to detect two IP address lists as a bot group activity in the CTU-11 scenario. In the CTU-12 scenario dataset, the B-Corr model obtained three similar bot IP addresses. From the three similar IP address bots, the CTU-12 dataset description compared, and the results show that the B-Corr model was able to detect three IP address lists as a bot group activity. In the CTU-9 scenario dataset, two bot IP addresses were not included in a bot group activity. In the CTU-10 scenario dataset, one bot IP address was not included in a bot group activity. In the CTU-11 scenario dataset, one bot IP address was not included in a bot group activity. The IP address bots were not included in a bot group activity because the IPs did not have any similarity to the target computer. The detection accuracy of the bot group activities with the B-Corr model is shown in **Table 7**.

Table 7. Accuracy of Bot group activities detection using B-Corr model

Skenario Numbers	9	10	11	12	Total Accuracy
Real CTU-13 Activties	10	10	3	3	**89.16 %**
B-Corr Model Detections	10	9	2	3	
Accuracy (%)	100 %	90 %	66.7 %	100 %	

However, they have similarities to attacks based on similarities in network header information, such as protocols, port addresses, and total packets. Overall, the tracing results from the proposed B-Corr model were able to detect similarity between bot activities. It is then compared with CTU dataset descriptions on CTU-9, CTU-10, CTU-11, and CTU-12. Comparison results show that the IP address in bot group activity generated by the B-Corr model has an accuracy of 89.16% against the description of the bot IP address list in the CTU dataset. The accuracy of detection results states that the model can detect the number of IP address bots or the number of bot actors with similarities and are referred to as bot groups.

5. Conclusion

This paper introduces a new model of similarity measurement with an intersection-probabilities approach called a B-Corr model. The main idea of this proposed model is to detect the bot group activities by measuring the degree of intersection-probabilities between bot flows in attacking targets on a computer network. B-Corr model analyzes the network header information, such as IP address, port address, protocol, and total packets. The concept of intersection-probabilities is obtained from inbound, inbound, outbound, and outbound degree feature extraction, then the similarity is calculated. For each similar bot that has similar value, it is grouped based on the same target to obtain a bot attack group called bot group activity. The detection for this activity would then enable us to isolate the activity or access from bot group activity, and it can be developed to determine the correlation between bots activities. The bot group activity results were compared with the IP address list in CTU scenario descriptions of CTU-9, CTU-10, CTU-11, and CTU-12. Comparison of CTU-9, CTU-10, CTU-11, and CTU-12 scenario datasets due to these scenarios have more than one bot. Accuracy of detecting bot group activity from specific IP addresses generated by the B-Corr model reached 89.16 % of the bot's IP address list in the CTU description.

In further research, a study on group bot activity is planned to be conducted based on attack steps, correlation measurement between bots and its visualization. The goal is to provide an overview of bot attack stages on the target computer and complete the B-Corr model's reliability.

References

[1] J. Álvarez Cid-Fuentes, C. Szabo, and K. Falkner, “An adaptive framework for the detection of novel botnets,” *Comput. Secur.*, vol. 79, pp. 148–161, 2018. Article (CrossRef Link)

[2] C. Y. Wang et al., “BotCluster: A session-based P2P botnet clustering system on NetFlow,” *Comput. Networks*, vol. 145, pp. 175–189, 2018. Article (CrossRef Link)

[3] K. Alieyan, A. Almomani, A. Manasrah, and M. M. Kadhum, “A survey of botnet detection based on DNS,” *Neural Comput. Appl.*, vol. 28, no. 7, pp. 1541–1558, 2017. Article (CrossRef Link)

[4] M. Eslahi, W. Z. Abidin, and M. V. Naseri, "Correlation-based HTTP Botnet detection using network communication histogram analysis," in *Proc. of 2017 IEEE Conf. Appl. Inf. Netw. Secur. AINS 2017*, pp. 7–12, 2017. Article (CrossRef Link)

[5] S. Chowdhury et al., "Botnet detection using graph-based feature clustering," *J. Big Data*, vol. 4, no. 1, 2017. Article (CrossRef Link)

[6] T. S. Wang, H. T. Lin, W. T. Cheng, and C. Y. Chen, "DBod: Clustering and detecting DGA-based botnets using DNS traffic analysis," *Comput. Secur.*, vol. 64, pp. 1–15, 2017. Article (CrossRef Link)

[7] C. Hung and H. Sun, "A Botnet Detection System Based on Machine-Learning using Flow-Based Features," *Securware*, vol. The Twelft, no. c, pp. 122–127, 2018.

[8] X. Hoang and Q. Nguyen, "Botnet Detection Based On Machine Learning Techniques Using DNS Query Data," *Futur. Internet*, vol. 10, no. 5, p. 43, 2018. Article (CrossRef Link)

[9] L. Mathur, M. Raheja, and P. Ahlawat, "Botnet Detection via mining of network traffic flow," *Procedia Comput. Sci.*, vol. 132, pp. 1668–1677, 2018. Article (CrossRef Link)

[10] S. Salah, G. Maciá-Fernández, and J. E. Díaz-Verdejo, "A model-based survey of alert correlation techniques," *Comput. Networks*, vol. 57, no. 5, pp. 1289–1317, 2013. Article (CrossRef Link)

[11] R. F. M. Dollah, M. A. Faizal, F. Arif, M. Z. Mas'ud, and L. K. Xin, "Machine learning for HTTP botnet detection using classifier algorithms," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 1–7, pp. 27–30, 2018.

[12] A. Ahmadian, M. Amini, and R. Ebrahimi, "RTECA : Real time episode correlation algorithm for multi-step attack scenarios detection," *Comput. Secur.*, vol. 49, pp. 206–219, 2015. Article (CrossRef Link)

[13] M. Ghasemigol and A. Ghaemi-bafghi, "E-correlator : an entropy-based alert correlation system," *security and communication networks*, vol. 8, no. 5, pp. 822–836, 2015. Article (CrossRef Link)

[14] J. Kwon, J. Kim, J. Lee, H. Lee, and A. Perrig, "PsyBoG: Power spectral density analysis for detecting botnet groups," in *Proc. of 9th IEEE Int. Conf. Malicious Unwanted Software, MALCON 2014*, pp. 85–92, 2014. Article (CrossRef Link)

[15] H. Choi, H. Lee, H. Lee, and H. Kim, "Botnet detection by monitoring group activities in DNS traffic," in *Proc. of CIT 2007 7th IEEE Int. Conf. Comput. Inf. Technol.*, pp. 715–720, 2007. Article (CrossRef Link)

[16] H. Choi, H. Lee, and H. Kim, "BotGAD: detecting botnets by capturing group activities in network traffic," in *Proc. of Fourth Int. ICST Conf. Commun. Syst. Softw. Middlew.*, pp. 1–8, 2009. Article (CrossRef Link)

[17] B. Zhu and A. A. Ghorbani, "Alert correlation for extracting attack strategies," *Int. J. Netw. Secur.*, vol. 3, no. 3, pp. 244–258, 2006.

[18] G. Khehra, "BotScoop : Scalable detection of DGA based botnets using DNS traffic," in *Proc. of 2018 9th Int. Conf. Comput. Commun. Netw. Technol.*, pp. 1–6, 2018. Article (CrossRef Link)

[19] M. Rigaki and S. Garcia, "Bringing a GAN to a knife-fight: Adapting malware communication to avoid detection," in *Proc. of - 2018 IEEE Symp. Secur. Priv. Work. SPW 2018*, pp. 70–75, 2018. Article (CrossRef Link)

[20] S. Garcia, "Modelling the Network Behaviour of Malware To Block Malicious Patterns. The Stratosphere Project : a Behavioural Ips," in *Proc. of Virus Bull.*, pp. 1–8, 2015. Article (CrossRef Link)

[21] S. García, "Identifying, Modeling and Detecting Botnet Behaviors in the Network Universidad Nacional del Centro de la Provincia de Buenos Aires Doctoral Thesis Identifying, Modeling and Detecting Botnet Behaviors in the Network," 2015.

[22] S. Garcia, A. Zunino, and M. Campo, "Survey on Network-based Botnet Detection Methods," *Sec. Commun. Netw.*, vol. 7, no. 5, pp. 878–903, May 2014. Article (CrossRef Link)

[23] S. Garc, M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," *Comput. Secur.*, vol. 45, pp. 100-123, 2014. Article (CrossRef Link)

[24] E. B. Beigi, H. H. Jazi, N. Stakhanova, and A. A. Ghorbani, “Towards Effective Feature Selection in Machine Learning-Based Botnet Detection Approaches,” in *Proc. of 2014 IEEE Conf. Commun. Netw. Secur.*, pp. 247–255, 2014. Article (CrossRef Link)

Dandy Pramana Hostiadi is a Doctoral Student in the Department of Informatics, Institut Teknologi Sepuluh Nopember (ITS), Indonesia. He also as Lecturer in Institut Teknologi dan Bisnis STIKOM Bali, Indonesia.
He received the bachelor’s degree from STMIK STIKOM Bali, master’s degree from Udayana University, all in Computer Science. His scientific interests include network security and network forensic. Dandy Pramana Hostiadi is the corresponding author and can be contacted at: dandy.17051@mhs.its.ac.id

Waskitho Wibisono is an associate professor in the Department of Informatics, Institut Teknologi Sepuluh Nopember (ITS), Indonesia. He received his bachelor's degree from ITS, a master's degree from Ritsumeikan University, and a Ph.D. degree from Monash University, all in Computer Science.
He is also a reviewer of some journals, including the International Journal of Communication Systems, IEEE Transactions on Industrial Informatics, Journal of King Saud University - Computer and Information Sciences, and IEEE Access. His scientific interests include wireless sensor networks, context-aware systems, pervasive and mobile computing, and distributed systems.

Tohari Ahmad received the bachelor degree in computer science from Institut Teknologi Sepuluh Nopember (ITS), Indonesia, MIT degree in information technology from Monash University, Australia, and the Ph.D degree in computer science from RMIT University, Australia, in 2012.
He is now an Associate Professor in ITS. His research interests include network security, information security, data hiding and computer network. He is a reviewer of some journals, including IEEE Access, Information Sciences, and Signal Processing. Some of his research can be found in Pattern Recognition, Engineering Letters, Journal of King Saud University - Computer and Information Sciences.