International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

A Three-Stage Botnet Detection Technique using Random and Obrazom Graphs

K. Akilandeswari, G.Anwar Basha, L.Baranivel

Abstract: ***This paper proposes three-stage botnet detection technique based on the anomaly and community detection. The first stage is a pragmatic node based distributed approach of sparse graph sequences. The second stage detects the bot from sparse matrix and correlations of interactions among the node. In the third stage, random graph is evaluating the performance of the bots and verified with both odd and even types of nodes. The same is extended and verified through Obrazom triple connected graphs. This verification is helpful to identify the aggressive bots through the optimized pivotal nodes. Machine Learning based Botnet Detection techniques are implemented in various levels like centralized and distributed level of networks. We can apply this three-stage bot detection in large-scale data.***

Key words: Botnets, Random graphs, Sparse Graph, social network and optimization.

I. INTRODUCTION

Generally Botnet means a well-organized automated collection of zombies which may use for creating a DDoS (Distributed Denial-of-Service Attacks) attack as well as spam actions of flooding any inbox or spreading the viruses [16]. Botnets are not even aware of that they are used for malicious purposes. Sometimes botnets are used explicitly to send spam mail. From the statistical data took from 2005 to till date, as estimation of 50 – 80 % of spam mails are sent by the collection of botnets [7] . Hence spammers are not allowed to find the bandwidth of the botnets, so botnets are having their own bandwidth [2]. Through the development of technology, every computer has the highest amount of processing powers in their CPU, GPU and bandwidth capacity [11]. Whenever these personal computers are joined into the internet and these made botnet is more and more powerful [18]. Majority of botnet attacks are subdivided the main source into the multiple sources and then combining this multiple sources [1]. After combining this multiple source the botnet started to attack based on its band width capacity and this becomes a powerful source. Hence attackers may use it in so many malicious purposes [17].

Few are listed:

- Distributed Denial-of-Service Attacks (DDoS)
- E-mails Spam
- Illegal Traffic
- Affecting New Hosts
- Network source theft
- Attacking Networks or chat networks
- Hosting of malicious software's
- Advertisement Addons
- Click Fraud
- Misusing Online Polls
- Remote Use of Personal Computers
- Attacking Bank Computers (ATM or any others since they are also networked)
- Manipulating Games and so on.
- Exploiting Private Documents and so on

Machine learning and even human is witnessed so many botnets and its attacks. Each botnets causes material damage for target firms [15]. Few botnets have grown extremely and caused very large damage across the world. More and more financial, entertainment and service oriented sites like PayPal, HBO, Netflix and some other services are experiencing the interruptions caused by botnets each and every second. Even we can witness some other IoT related botnets soon in the future too [13]. Hence botnet identification, detection and elimination are the important and challenging tasks in the cyber security domain. Few standard and big companies have started to ensure its security-concern and investing great effort to detect and eliminate botnets [8].

Basically anomaly detection did not fit in to the normal expected behavior, and it is useful to identify the exceptional networks as well as communication based patterns networks [4]. Technically, based on the existing, observed anomalous data the authors made a unique assumption for each category of anomaly detection techniques [14].

There are so many botnet detection techniques and tools are used worldwide in the literature. The general structure of botnet detection techniques are classified into two broad categories, IDSs and HoneyNets [12]. A honeynet is used to collect information from bots for further analysis to measure the intensity and vulnerability of the attack [9]. Moreover, the information collected from bots is used to discover the C&C system, unknown susceptibilities, techniques and tools used by the attacker, and the motivation of the attacker [10]. A honeynet is used to collect bot-binaries which penetrate the botnets. However, intruders developed novel methods to overcome honeynet traps [5].

As a outline, attacks which are detected in IDS/IPSs characterize an individual pattern and these attacks are functional from one precise source. Attacks which are produced by botnet are part of a gigantic network [3].

Revised Manuscript Received on March 30, 2020.
* Correspondence Author
Dr. K. Akilandeswari, Associate Professor, Dept of Computer Science, Govt Arts College (Autonomous), Salem -
Mr.G. Anwar Basha, an independent researcher with 10 years of teaching experience.
Mr.L.Baranivel, pursuing MCA, M.Tech, research at VIT, Vellore.

This paper is organized as follows. Section II dealt with Random and Obrazom graphs to detect the botnets via node interaction [6]. It's a triple connected graph, with this graph the maximum and minimum degree of edge connectivity and independent sets graph are considered to identify the bots. These edges are transmitted as nodes in the networks. Section III focused the Machine Learning Node Based Botnet Detection Techniques. This showed the basic structure distributed and centralized) of the botnets. Section IV concludes the paper.

II. RANDOM AND OBRAZOM GRAPH BASED BOTNET DETECTION (ROGBD)

In this section, we considered random graph, inverse random graph and node interaction graphs whose orders are l, m and n and respectively. Generally, for a graph G, we denote $V(G), E(G), \Delta(G), \delta(G), \lambda(G), i(G)$ and for its vertex set, edge set, maximum degree, minimum degree edge connectivity and independent set respectively and the degree of a vertex is denoted by $d_G(v)$. These random graph, inverse random graph and node interaction graphs related to hypothesis testing to select the appropriate graph to locate / identify the botnet.

Definition 2.1

A non-empty subset D of V of a random graph R_G if D is aconnected dominating set and the induced sub graph $< V - D >$ is triple connected. The minimum cardinality of R_G is denoted by $\gamma_c^{tc}(G)$.

Definition 2.2

A non empty subset D of vertices in a inverse random graph R'_G whose minimum cardinality is $\gamma_{c\leq 2}^{tc}(G)$.

In a random graph, with $n \geq m$, $\gamma_c'^{tc}[L(W_{1,n})] = \left\lfloor \frac{n+m}{2} \right\rfloor$.

Now let

$$V\left(L(W_{m,n})\right) = E(W_{m,n}) = \{e_i : 0 \leq i \leq m+n-1\} \cup \{e_i' : 0 \leq i \leq m-1\} \quad (1)$$

We know that $\gamma_c^{tc}\left(L(W_{1,n})\right) = \left\lfloor \frac{n+m}{2} \right\rfloor$. In $L(W_n)$, let us consider

$$B' = \begin{cases} e_p' : p = 1,2,5 \ldots n-1 \\ \text{otherwise} \end{cases} \quad (2)$$

(i) If p is odd then n is even : To Prove the above condition let $n = 6$, and Choose $\{e_1', e_3', e_5'\}$ are the vertices which are adjacent to $\{e_1, e_3, e_5\}$ in $L(W_n(V-D))$. The induced sub graph $< e_1', e_3', e_5' >$ is connected and $< e_0, e_1, e_2, e_3 e_4, e_5 >$ is triple connected. That is $L(W_{1,n}(V-E))$ a triple connected sub graph contains a blast dominating set. Hence $\gamma_c'^{tc}[L(W_{1,6})] = 3$.

(ii) If p is even then n is odd: To ensure the above the result let us consider $n = 7$, let us fix $\{e_2, e_2', e_4, e_4', e_6, e_6'\}$ be the vertices adjacent to $\{e_1, e_2, e_3, e_4, e_5, e_6\}$ in $L(W_{1,n}(V-D))$. Therefore $\{e_2', e_4', e_6'\}$ and $\{e_2, e_4, e_6\}$ whose induced sub graph $< D' >$ is connected dominating set and simultaneously the induced sub graph of its complement $< V - D' >$ is triple connected. That is $L(W_{1,n}(V-D))$ contains a blast dominating set. Hence

$$\gamma_c'^{tc}[L(W_{1,7})] = 4.$$

Thus in succession, the random graph and its dominating set of $L(W_{m,n})$ is the minimum random graph dominating set of $L(W_{1,n})$ and $L(W_{1,m})$.Thus,

$$\gamma_c'^{tc}[L(W_{1,n})] = \gamma_c^{tc}\left(L(W_{1,m})\right)\left\lfloor \frac{n+m}{2} \right\rfloor \quad (3)$$

Definition 2.3

A graph sequence $(G_{n,m})_{n,m \geq 1}$ is called sparse graph when $\lim_{n\to\infty}(P_k)^{(n)} = w_k$ where $k \leq 0$, and $\lim_{m\to\infty}(P_k)^{(m)} = w_k$ for some deterministic limiting probability distribution w_k where $k \leq 0$. These concepts are often used to define the random graphs and the limit is applied to correlates and interacts among the nodes.

Figures 1 and 2 shows the distributed and centralized botnet structures.

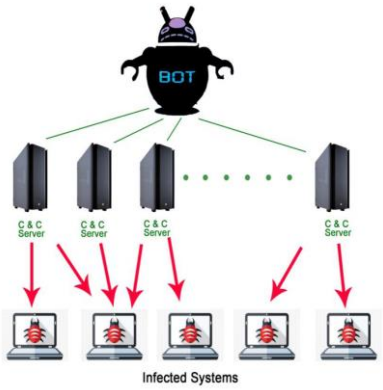

Figure 1: Distributed Botnet Structure

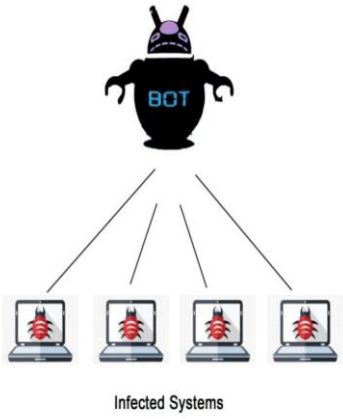

Figure 2: Centralized Botnet Structure

In the above definition when n and m are embedded then the sparse cluster will exists.

Definition 2.4

A graph sequence $(G_{n,m})_{n,m \geq 1}$ is called scale free with exponent τ when it is sparse and

$$\lim_{n\to\infty} \frac{\log[1-F(k)]}{\log(\frac{1}{k})} = \tau - 1 \qquad (4)$$

where $F(k) = \sum_{n,m\leq k} p_{n,m}$ denotes the cumulative distribution function corresponding to the probability mass function $p_{n,m\leq k}$ defined in (1). Thus for a scale free random graph process its degree sequence converges to a limiting probability distribution as in (1), and the limiting distribution has asymptotic power law tails described in (1).

Consider for any sparse graph with $m, n \geq 3, 2 \leq \gamma_{c\leq 2}^{'tc}(L(W_{1,n})) \leq \left\lfloor \frac{\Delta}{2} \right\rfloor$. Now consider a ϑ −obrazom of $W_{m,n}$ with n vertices and maximum degree Δ, $\left\lceil \frac{\Delta-1}{n} \right\rceil \leq \gamma_c^{'tc}(L(W_{m,n}))$.

If we can take a minimal dominating set in $L(W_{m,n})$ then the undirected and induced sub graph $< D >$ such that for every two distinct vertices in the clique are adjacent. Let us presume $D = \{e_i' : 0 \leq i \leq m+n-1\}$ .Now choose the vertices $\{e_i' : 0 \leq i \leq m+n-1\}$ is adjacent to $\{e_i : 0 \leq i \leq n-1\}$in $L(W_{m,n})$, whose induced sub graph is connected and a complete graph graph. It's complement $< V - D >$ is triple connected.

If we take a minimal blast dominating set in $L(W_{m,n})$ then the induced sub graph $< V - D >$ is cycle of length $m+n-1$ and consider a ϑ −obrazom of $W_{m,n}$,

$$\frac{n}{\Delta+1} + 1 \leq log\ i(G) \leq \frac{n\Delta}{\Delta+1} - 1.$$

Every scale free dominating set is a random dominating set in $W_{m,n}$. Now for a ϑ −obrazom of $W_{m,n}$, we have the following conditions.

(i) $\gamma_{c\leq 2}^{tc}(L(W_{1,n})) < \gamma_c^{tc}(L(W_{m,n}))$

(ii) $\gamma_c^{'tc}\left(L(W_{1,m})\right) \leq \gamma_c^{tc}\left(L(W_{m,n})\right)$

(iii) $\gamma(L(W_{1,n})) \leq \gamma_c^{tc}(L(W_{m,n}))$

$$(iv)\gamma'\left(L(W_{1,n})\right)\gamma_c^{'tc}\left(L(W_{m,n})\right) \qquad (5)$$

The following are the proposed algorithm shows the dominating set of a graph G which is useful to identity the botnets.

Step 1: All the vertices in V are initialized.

Step 2: Select any one vertex randomly called as v_1 which has the maximum degree, comparing to its neighbor vertices. Allow this vertex v_1 to send a notification to all its neighbors within the network. Note down the reaction of the neighbor networks after receiving the notification from the vertex v_1.

Step 3: Select any one vertex randomly called as v_2 which has the minimum degree, comparing to its neighbor vertices. Allow this vertex v_2 to send a notification to all its neighbors within the network. Note down the reaction of the neighbor networks after receiving the notification from the vertex v_2.

Step 4: Identify the Region Of Convergence (ROC) of v_1. Check the adjacent vertices about its anomaly and compute it.

Case 4.1: If the ROC reaches its maximum likelihood level then do the parity check between the neighbor vertices based on its eigen values.

Case 4.2: If the ROC reaches its minimum likelihood level then do the parity check between the neighbor vertices based on its eigen values.

Step 5: Identify the Region Of Convergence (ROC) of v_2. Check the adjacent vertices about its anomaly and compute it.

Case 5.1: If the ROC reaches its maximum likelihood level then do the parity check between the neighbor vertices based on its eigen values and maximum cardinality vertex covering.

Case 5.2: If the ROC reaches its minimum likelihood level then do the parity check between the neighbor vertices based on its eigen values and minimum cardinality vertex covering.

This minimum and maximum covering will help to find the maximum and minimum botnets in the network.

Step 6: Identify / locate the appropriate dominant graphs.

Step 7: Identify the illegal anomaly vertex called as pivotal node. After identifying the pivotal node, then does the parity check or intermediate communication of the pivotal nodes and the other nodes.

Consider the dominating set and its subset $D \subseteq V$ in a graph $G = (V, E)$ which is an independent distance dominating set, provided no two vertices in D are adjacent. If the adjacencies were occurred then the nodes are collinear. To overcome this we took a dominant node. An independent dominating set D is called a minimal independent (to reduce the aggressiveness of the bots) dominating set if no proper subset of D is an independent dominating set of G.

Now we consider the Obrazom triple connected graph say $L(K_{m,n})$ with $n, m > 4, \gamma_c^{'tc}\left(L(K_{m,n})\right) = 2$.

The $L(K_{m,n})$ graph hold the following (even)

$$\tau_{c\leq 2}^{'tc}(L(K_{m,n})) \leq \gamma_{c\leq 4}^{'tc}(L(K_{m,n}) \leq \gamma_{\leq 6}'\left(L(K_{m,n})\right) \qquad (6)$$

and so on. In a similar way we can have an odd based obrazom triple connected graph as

$$\tau_{c\leq 1}^{'tc}(L(K_{m,n})) \leq \gamma_{c\leq 3}^{'tc}(L(K_{m,n}) \leq \gamma_{\leq 5}'\left(L(K_{m,n})\right) \qquad (7)$$

and so on.

For any random graph $G_{m,n}$, we have

$\gamma_{s\leq 2}(G) \leq \log \frac{n.\Delta(G_{m,n})}{(\Delta(G)+1)}$.

Since $\gamma_{s\leq 2}(G_{m,n}) \leq \gamma_s(G_{m,n})$ and, we have

$$\gamma_s(G_{m,n}) \leq \log \frac{n.\Delta(G_{m,n})}{(\Delta(G_{m,n})(m+n+1)} \quad (8)$$

We have,

$$\gamma_s(G_{m,n}) \leq \log \frac{n.\Delta(G_{m,n})}{(\Delta(G_{m,n})(m+n-1)} \quad (9)$$

This shows the expected number of large connected components in the graph. We transmit the same to the bots and conclude about the connectedness and inter connectedness of the networks. Figure 3- 5 shows the various stages of botnet detection techniques.

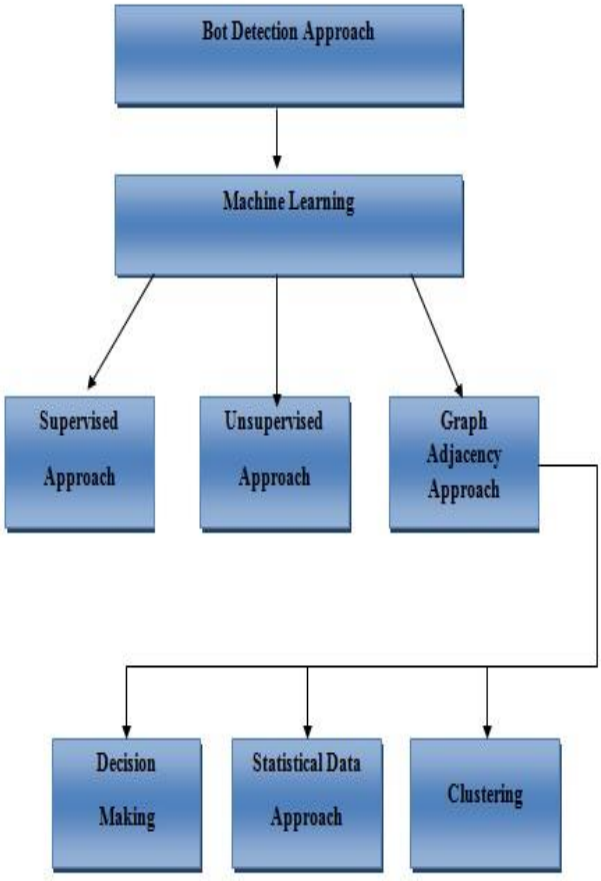

Figure 3: Stage 1- Pragmatic Node based distributed approach of botnets

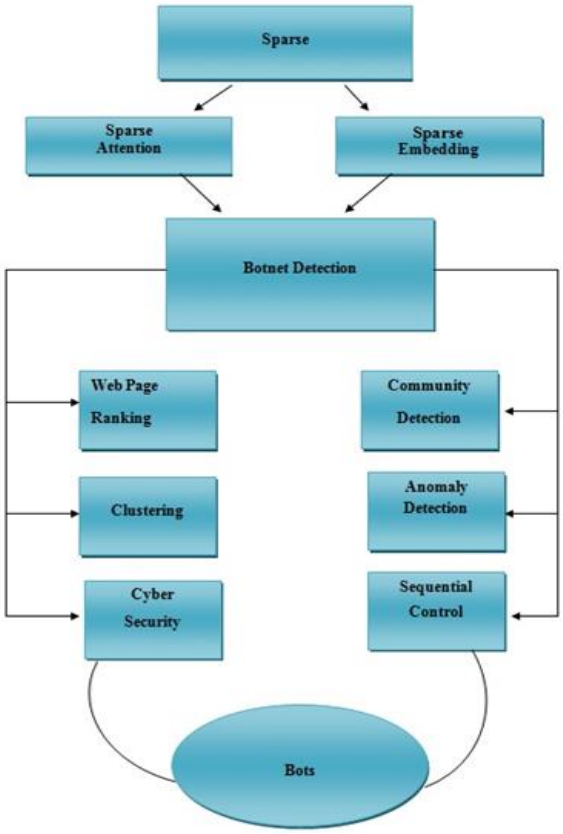

Figure 4: Stage 2- Bot detection from Sparse Graph

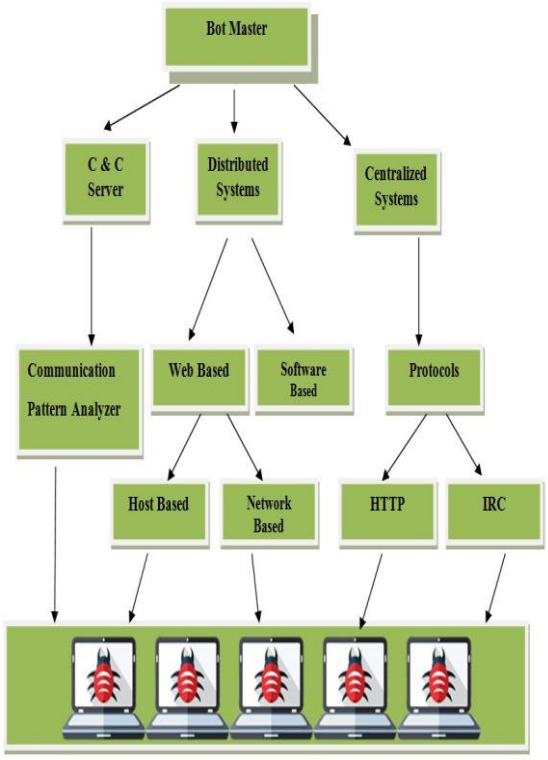

Figure 5: Stage 3 – Identification of Regressive bots through Machine Learning

III. MACHINE LEARNING BASED BOTNET DETECTION (MVBD)

In this section we introduced the Graph based clustering approach related to stochastic model which involve Fast Fourier Transform. Initially, identify the highly interacted nodes both in maximum likelihood level and minimum likelihood level. If the systems are interconnected then the botnet detection approach, related to the malicious traffic is noted by observing its network traffic in a certain domain i.e., within different coordinates, which includes the traffic passage in the set of connections, traffic patterns, response and react time, network load balance and management and uniqueness in the link. These are classified into active monitoring type, and passive monitoring type.

Active monitoring is also known as synthetic monitoring. In this active monitoring, in order to detect the malwares and malware affected system new packets are injected in the network. Passive monitoring techniques employing various application modules which includes various statistical approach based technologies. This machine learning based technique detects the communication pattern of the bots.

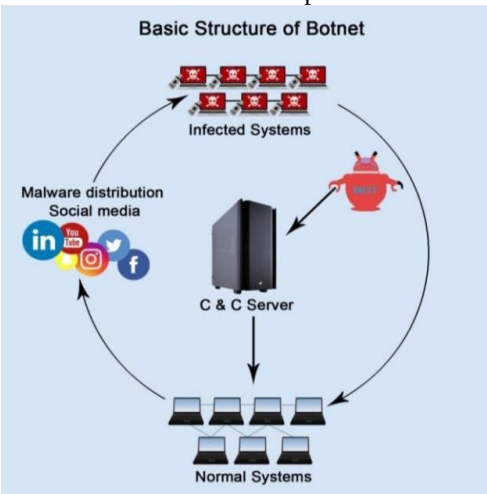

Figure 6: Basic Structure of Botnet

Once we identify the communication pattern of the bots, then we needs to observe the behavioral characteristics of the bots. Figure 6 shows the basic functionality of botnets. Figure 7 shows the odd and even number of dominating pivotal value of the sparse graph.

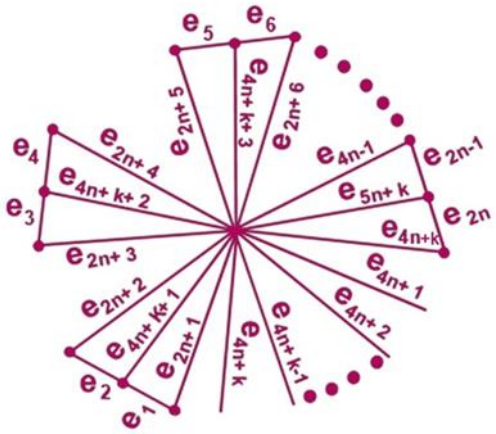

Figure 7: Odd and even nodes of the Bots

If $\gamma_{ct}^{-1}[J(T)] \leq \log(q - \Delta'(G))$, where G is a connected graph with q edges and $\Delta'(G)$ is the maximum edge degree of G. From this concept we can identify the maximum connected nodes in the bots. This also helps us to know how the edges are connected via bots. This learning technique was very useful to be aware from the bots and infects. Hence the cardinality of maximum dominating set will be arrived.

IV. RESULTS AND DISCUSSION

This section compares the performance of the existing Virtual Honeypot Botnet Detection(VHBD) architecture and the proposed techniques. The metrics that are used for validating the performance of the proposed system are as follows,

- Detection rate
- Precision
- Recall

A. Detection rate

The detection rate is defined as the amount of time consumed between successive packet transmissions. The comparison of detection rate for the existing VHBD architecture and the proposed techniques is depicted in Fig. 4.1. From the figure, it is analyzed that the suggested techniques provide increased detection rate than the existing algorithms.

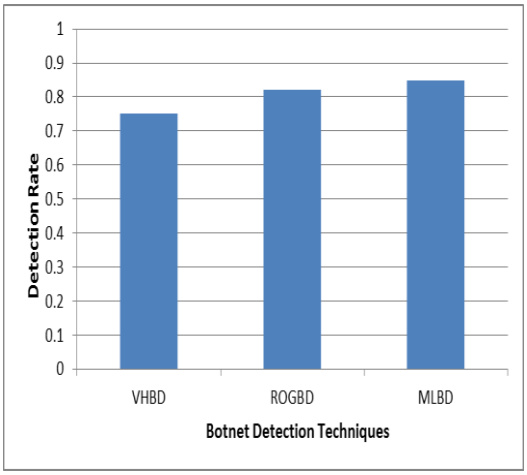

Fig. 8: Comparison of detection rate for the existing and the proposed algorithms

B. Precision and Recall

The precision measure estimates the purity of the cloud environment by considering the fraction of botnet nodes to the total number of nodes in the cloud.

$$Precision = \frac{|b \cap c|}{|c|} \quad (1)$$

The recall measure estimates the total fraction of bots being identified.

$$Recall = \frac{|b \cap c|}{|b|} \quad (2)$$

Where,

b represents the number of malicious IP address

c denotes the total number of IP addresses that enter.

The comparison of precision and recall for existing VHBD architecture and the proposed techniques is depicted in Fig. 4.2. The analysis results prove that the proposed techniques provide increased precision values than the existing algorithms.

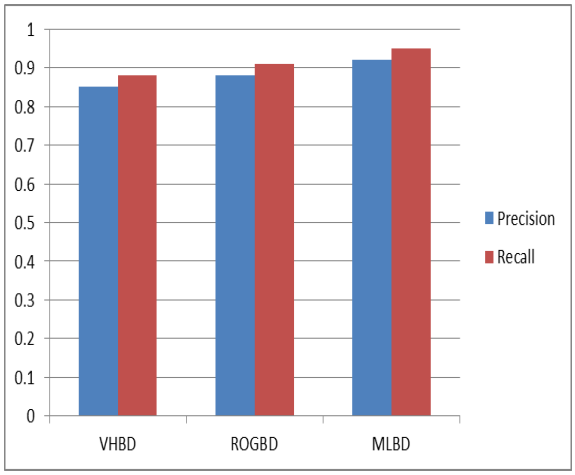

Fig. 9: Comparison of Recall and Precision for the existing and the proposed methods

V CONCLUSION

The three-stage botnet detection based on Graph and Network approaches are arrived in this paper. Sparse graph based approach detects the bots that rely on any protocols which is connected in the Internet that is large scale data. The aggressive nodes are identified through Random and Obrazom graphs. Both odd and even edges are verified for the connected graphs. Finally this study applied the advantages of three stages of detecting the bots (pragmatic node based distributed approach, bot detection from sparse graph and identification of regressive bots through machine learning). However according to the C & C Server, HTTP, P2P and the other network protocols, this graph based concepts are optimizing and evolving the bots and used to predict the infected level of the normal system.

REFERENCES

1. Arora, A., Yadav S.K., Sharma K, Denial-of-Service (DoS) Attack and Botnet: Network Analysis, Research Tactics, and Mitigation. In Handbook of Research on Network Forensics and Analysis Techniques; IGI Global: Hershey, PA, USA, pp. 117–141, 2018.
2. Azab A., Alazab M., Aiash M, Machine Learning Based Botnet Identification Traffic. In Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, China, 23–26 August, pp. 1788–1794, 2016.
3. Chen C M., Lai G.H, Young P Y. , Defense Joint Attacks Based on Stochastic Discrete Sequence Anomaly Detection. In Proceedings of the 2016 11th Asia Joint Conference on Information Security (Asia JCIS), Fukuoka, Japan, 4–5, pp. 74–79, August 2016.
4. Chowdhury S, Khanzadeh M, Akula R, Zhang F, S. Zhang, H. Medal, Marufuzzaman, and L. Bian, "Botnet detection using graph based feature clustering," Journal of Big Data, vol. 4, no. 1, p. 14, 2017.
5. Collins M P and M. K. Reiter, "Hit-list worm detection and bot identification in large networks using protocol graphs," in International Workshop on Recent Advances in Intrusion Detection. Springer, pp. 276–295, 2007.
6. Garcia S , M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," Computers & Security, vol. 45, pp. 100–123, 2014.
7. Haddadi F., Zincir Heywood A N, Botnet behaviour analysis: How would a data analytics-based system with minimum a priori information perform, Int. J. Netw. Manage. 2017, 27, 1977.
8. Liao, W.H.; Chang, C.C. Peer to peer botnet detection using data mining scheme. In Proceedings of the 2010 International Conference on Internet Technology and Applications, Wuhan, China, 20–22 August 2010; pp. 1–4.
9. Jaikumar P and A. C. Kak, "A graph-theoretic framework for isolating botnets in a network," Security and communication networks, vol. 8, no. 16, pp. 2605–2623, 2015.
10. K Karasaridis, B. Rexroad, D. A. Hoeflin et al., "Wide-scale botnet detection and characterization" HotBots, vol. 7, pp. 7–7, 2007.
11. Krueger T., Gascon, H. Kramer, N.; Rieck, K. Learning stateful models for network honeypots. In Proceedings of the 5th ACM Workshop on Security and Artificial Intelligence (AISec '12), Raleigh, NC, USA,; p. 37, 19 October 2012.
12. Ma X., Guan X., Tao J, Zheng Q, Guo, Y., Liu L., Zhao S., A novel IRC botnet detection method based on packet size sequence, In Proceedings of the 2010 IEEE International Conference on Communications (ICC), Cape Town, South Africa, 23–27, pp. 1–5, May 2010.
13. Saad S., Traore I., Ghorbani, A., Sayed B., Zhao D., Lu W., Felix J., Hakimian P, Detecting P2P botnets through network behavior analysis and machine learning, In Proceedings of the 2011 - 9th Annual International Conference on Privacy, Security and Trust (PST 2011), Montreal, QC, Canada, 19–21; pp. 174–180, July 2011.
14. Venkatesh B, S. H. Choudhury, S. Nagaraja, and N. Balakrishnan, "Botspot: fast graph based identification of structured p2p bots," Journal of Computer Virology and Hacking Techniques, vol. 11, no. 4, pp. 247– 261, 2015.
15. Wang J and I. C. Paschalidis, "Botnet detection using social graph analysis," in Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on. IEEE, 2014, pp. 393–400.
16. Zeidanloo H R, A. B. Manaf, P. Vahdani, F. Tabatabaei, and M. Zamani, "Botnet detection based on traffic monitoring," in Networking and Information Technology (ICNIT), 2010 International Conference on. IEEE, pp. 97–101, 2010.
17. Zhao S., Lee P P., Lui J., Guan X., Ma X., Tao J. , Cloud based push-styled mobile botnets, A case study of exploiting the cloud to device messaging service, In Proceedings of the 28th Annual Computer Security Applications Conference, Orlando, FL, USA, 3–7; pp. 119–128, December 2012.
18. D. Zhuang and J. M. Chang, "Peer hunter: Detecting peer-to-peer botnets through community behavior analysis," in Dependable and Secure Computing, 2017 IEEE Conference on IEEE, , pp. 493–500, 2017.