Erdos – Renyi Random Graph and Machine Learning Based Botnet Detection

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

***Abstract:** Basically large networks are prone to attacks by bots and lead to complexity. When the complexity occurs then it is difficult to overcome the vulnerability in the network connections. In such a case, the complex network could be dealt with the help of probability theory and graph theory concepts like Erdos – Renyi random graphs, Scale free graph, highly connected graph sequences and so on. In this paper, Botnet detection using Erdos – Renyi random graphs whose patterns are recognized as the number of connections that the vertices and edges made in the network is proposed. This paper also presents the botnet detection concepts based on machine learning.*

***Keywords:** Botnet detection, machine learning, Erdos – Renyi random graph.*

I. INTRODUCTION

The advent of the digital age has prompted a rising attention towards the real world internet networks [11]. Due to the highly improved computational control, huge data sets can now effortlessly be stored and investigated, and this has a deep impact on the experimental studies of the huge networks [5]. Recently many researchers are showing quite attentions towards how the computers connected in the Internet behave under malicious attacks or accidental attacks [10, 18]. Parallelly the research is going on how this network is attracted and how these connections get attention towards the clouds [20]. In general Internet consists of various layers that incorporate the compatibility among the jumbled and diversified network topologies [1]. It may be divided into different layers in the clouds [21]. But the internal connectivity is totally invisible to trace the route and each route is having high in degrees [2]. The routers with high degree are most likely biased, so in such a case trace out the network route is quite difficult [19]. This Internet based study is designed to be aware of why many networks are affected by malwares by its fascinating features and what else the malwares can do with spread of routing information, ranking of the vertices and so on [15]. Botnet is a well-planned programmed compilation of zombies which is used for creating and developing DDoS (Distributed Denial-of-Service Attacks) attack as well as sending irrelevant data's to inbox or scattering the viruses [4] and botnets are not even conscious that they are used for malicious attacks [9]. The prevention techniques are identifying to recover the data and not to steal the data during the malicious attacks and to calculate the hitting time and total progeny of the process [14]. So the experts are introducing multi-level and multi-layer models for the internet based on the corresponding network topologies [17]. If the network is a complex network then the virtual source of malicious attacks and its related information are concerned massive attention [3]. Such a situation needs highly traceable to identify the critical vulnerability of the affected systems [8]. So the internet systems need some high degree deterministic and random routers to classify the breakdowns [6].

The progress in the digital technology shows the maximum of processing powers in their hardware equipment's like CPU, GPU and bandwidth capacity [7]. These digital techniques make the good and bad things make more powerful [13, 16]. When this powerful processer enters in to the computers and the internet, the vibration of botnet becomes more and more powerful [12]. The main influential stream of botnet attacks are:

- Partitioned the main source into the multiple sources
- Then combining this multiple sources and
- Started to attack according to its band width capacity

Hence attackers considered the botnet is a powerful source to spread malwares especially in the following activities:

- Swarming malicious software's to the computers which are connected in the internet
- Distributed Denial-of-Service Attacks (DDoS)
- E-mails Spam
- Creating some unnecessary Illegal Traffic in the cloud
- Network source stealing
- Hacking personal chats
- Developing more and more Click Fraud
- Remote Use of Personal Computers
- Attacking the financial sectors like Bank Computers and so on.

Now a day's many organizations are under security threats, not only financial related crisis but also related to its reputation. A botnet attack is an extensively known threat to these organizations. According to the U.S. Federal Bureau of Investigation, "Botnets caused over 9 billion $ losses to U.S. victims and over 110 billion

$ globally" [2017]. The most famous attack, Rustock, infected 1 million machines, sending up to 30 billion spam emails a day. More recently,

Revised Manuscript Received on March 30, 2020.
* Correspondence Author

Dr. K. Akilandeswari, Associate Professor, Department of Computer Science, Government Arts College (Autonomous), Salem.
Mr. L. Baranivel*, MCA, Department of M.Tech, VIT, Vellore, India.
Mr. G. Anwar Basha. Associate Professor, Department of Computer Science, Government Arts College (Autonomous), Salem.

Wannacry reported that over 230,000 computers in 150 countries are damaged through bots [8]. Attackers constantly find clever ways to interrupt networks using botnets, through zero-day attacks. The structure of such botnets are designed to affect the performance and to spread malwares in Social networks, Telecommunication networks, Collaborative and citation networks of scientists, Banking Sectors, Service Oriented networks and even hospitals. To overcome these barriers the blooming research works provides eminent solutions to these bots in a certain level.

This paper is organized as follows. Section II is dealt with the concepts of Erdos – Renyi random graph based networks. We took the monotonically increasing networks based on its edges and showing how the nodes or bots are interconnected and spread its connection. Section III focuses the machine learning concepts of bots. It is classified into four transition phases. Section IV concludes the paper.

II ERDOS – RENYI RANDOM GRAPH(ERRG) BASED NETWORKS

Consider the Erdos – Renyi random graph $\boldsymbol{R(G)}$ (monotonically increasing – based on its edges) whose vertex sets are { v_2, v_4, v_6, v_8, …, v_{2i}, v_{2i+2}, v_{2i+4}, v_{2i+6}, ... , v_{4i}, v_{4i+2}, v_{4i+4}, v_{4i+6}, …, v_{4i+k}}· It has edge set { e_1, e_3, e_5, e_7, …, e_i, e_{i+2}, e_{i+4}, ... , e_{3i}, e_{3i+2}, e_{3i+4}, …, e_{3i+k}}· If $\boldsymbol{i \equiv 0(mod\ 3)}$ then define, $\boldsymbol{g}: \boldsymbol{R(G)} \to \{\boldsymbol{1, 3, 5, \dots, 2q-1}\}$, by $\boldsymbol{f}$ (e_i) = $(\boldsymbol{2i-1})$, $\boldsymbol{i} = \boldsymbol{1}$ to $\boldsymbol{5n} + \boldsymbol{k}$.

To arrive the number of bonnets spread or infected in the particular domain or region, we defined the following function,

$$g^*\ (\boldsymbol{uv}) = \begin{cases} 0; & \text{if } \dfrac{f(u)+f(v)}{2} \text{ is an integer} \\ 1; & \text{otherwise} \end{cases}$$

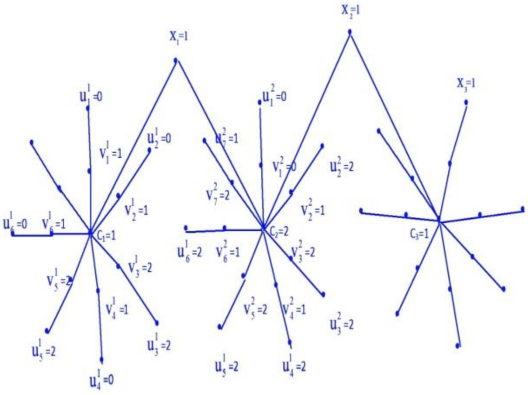

Figure 1: Erdos - Renyi graph nodes connected in the Internet

Figure 1shows how the Erdos - Renyi graph nodes connected in the internet. Again in order to get, how the botnets are interconnected, we define a map g* : E → {0, 1} as follows.

(i) for $\mathbf{i} = \mathbf{1}$ to $\mathbf{n}$ and $\mathbf{j} = \mathbf{1}$ to $\mathbf{k}$

$$g1^*\ (u_i^j\ v_i^j) = g^*\ (c1\ v_i^j) = \begin{cases} 0; & i \equiv 1\ (\text{mod}\ 2) \\ 1; & i \equiv 0\ (\text{mod}\ 2) \end{cases}$$

$$g2^*\ (v_i^j\ u_i^j) = g^*\ (v_i^j\ c1) = \begin{cases} 1; & i \equiv 0\ (\text{mod}\ 3) \\ 0; & i \equiv 1 (\text{mod}\ 3) \end{cases}$$

(ii) for $\mathbf{i} = \mathbf{1}$ **to** $\mathbf{n+1}$ and $\mathbf{j} = \mathbf{1}$ to $\mathbf{k+1}$

$$\boldsymbol{g_1}^*\ (c2\ xj) = \begin{cases} 1; & i \equiv 1\ (\text{mod}\ 3) \\ 0; & i \equiv 0 (\text{mod}\ 3) \end{cases}$$

$$\boldsymbol{g_2}^*\ (c2\ xj) = \begin{cases} 0; & i \equiv 1\ (\text{mod}\ 2) \\ 1; & i \equiv 0 (\text{mod}\ 2) \end{cases}$$

(iii) for $\mathbf{i} = \mathbf{1}$ **to** $\mathbf{n+k}$ and $\mathbf{j} = \mathbf{1}$ to $\mathbf{n+k}$

$$\boldsymbol{g_1}^*\ (c3\ xn{+}k) = \begin{cases} 1; & i \equiv 1\ (\text{mod}\ 2) \\ 0; & i \equiv 0 (\text{mod}\ 2) \end{cases}$$

$$\boldsymbol{g_2}^*\ (c3\ xn{+}k) = \begin{cases} 0; & i \equiv 1\ (\text{mod}\ 2) \\ 1; & i \equiv 0 (\text{mod}\ 2) \end{cases}$$

Then proceed with the following algorithm:

Step 1: Construct a randomized study towards the efficiency of odd and even edges and odd and even vertices.

Step 2: When the vertices are even:

for $\boldsymbol{i} = \mathbf{1}$ to $\dfrac{n+1}{2}$ and n ≡ 1 (mod 2) and for $\boldsymbol{j} = \mathbf{1}$ to $\dfrac{m+1}{2}, m \equiv 1\ (\text{mod}\ 2)$

$$\boldsymbol{g_1}^*\ (u_i^j\ v_i^j) = 0,\ \ \boldsymbol{g_2}^*\ (u_i^j\ v_i^j) = 1$$

$$\boldsymbol{g_3}^*\ (u_{\frac{n+1}{2}+i},\ v_{\frac{m+1}{2}+j}) = 0;\ \ \boldsymbol{g_4}^*\ (u_{\frac{n+1}{2}+i}\quad v_{\frac{m+1}{2}+j}) = 1$$

Step 3:

When the vertices are odd

for $\boldsymbol{i} = \mathbf{1}$ to $\dfrac{n-1}{2}$ and n ≡ 1 (mod 3) and for $\boldsymbol{j} = \mathbf{1}$ to $\dfrac{m-1}{2}, m \equiv 1\ (\text{mod}\ 3)$

$$\boldsymbol{g_1}^*\ (u_i\ v_i^j) = 0,\ \ \boldsymbol{g_2}\ (u_i\ v_i^j) = 1$$

$$\boldsymbol{g_3}\ (u_{\frac{n+3}{2}+i}\quad v_{\frac{n+3}{2}+i}) = 0;\ \boldsymbol{g_4}\ (u_{\frac{n+3}{2}+i}\quad v_{\frac{n+3}{2}+i}) = 1$$

Step 4: When the edges are even, we have E_a ={(e_i, $v_{(a+i)(mod\ n)}$ | $0 \le i \le n$ }

Step 5: When the edges are odd, we have E_b = {(e_j,$v_{(b+i+1)(mod\ n)}$ | $0 \le j \le n + 1$}

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

If $\boldsymbol{e} < \frac{(1-m-n)\ln n}{m+n}$, then the graph $\boldsymbol{R(G)}$ contains isolated vertices. This leads to the disconnected graph (bounded bandwidth networks). If $\boldsymbol{e} > \frac{(1-m-n)\ln n}{m+n}$, then a graph $\boldsymbol{R(G)}$ contains no isolated vertices, which is connected and inter connected, since $\boldsymbol{R(G)}$ is monotonically increasing based on its edges. The edges in the graph are independent and whose probability is P (the connection between one node to another node) and n is the number nodes connected or inter connected. Then the degree of the corresponding graph is calculated based on,

$$P(deg(v) = k) = \binom{n-1}{k} p^k - p^{n-k-1}$$

This shows the existing connection between the nodes.

III PHASE TRANSITIONS FOR THE LARGEST CONNECTED COMPONENTS

Botnet detection concepts based on machine learning is classified into the following four transition phases for the largest connected components:

- Classification Phase
- Implementation Phase
- Testing Phase
- Concluding Phase

These phases are applicable to the weakly connected or strongly connected graphs. During the classification phase, we showed how the botmaster is spreading the bots through C & C Servers and Peer to Peer to networks.

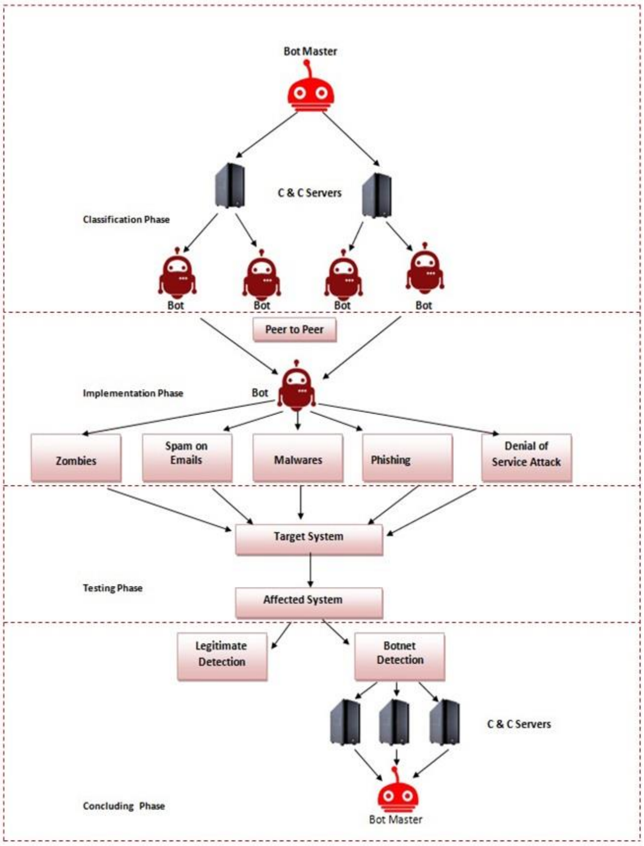

Figure 2 : Phase Transitions of Bots

In the implementation phase how the bots attacked its targeted system through zombies, spam on emails, malwares, phishing, denial of service attack and so on. In the testing phase, the affected systems are classified into legitimate users and attackers based system. In the evaluation phase we focused how the systems are affected.

This is the processing features of our machine learning algorithm and we implemented the same through Erdos – Renyi Random graph. Figure 2 explains the various phase transitions of Bots.

IV. RESULT AND DISCUSSION

This section illustrates the comparison of the performance for existing Virtual Honeynet Based Botnet Detection (VHBD) architecture and proposed technique ERRG. The metrics that are used for validating the performance of the proposed system are as follows,

- False positive rate
- False negative rate

A. False positive rate

The false positive rate is the proportion of valid IP address detected as bots. Increased false positive rate minimizes the overall performance of the system. The comparison of false positive rate for the existing VHBD architecture and proposed technique ERRG is depicted in Fig. 5.1.

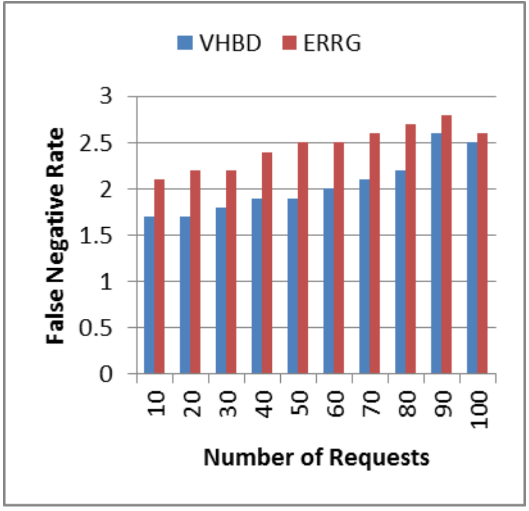

Fig 5.1 : Comparison of False Positive rate for the existing VHBD and the proposed ERRG

B. False negative rate

The false negative rate is defined as the proportion of malicious IP address not detected as the botnet. It is mandatory to minimize the false negative rate because the increase in this rate minimizes the overall performance of the botnet detection system.

The comparison of false negative rate for the existing VHBD architecture and proposed technique ERRG is depicted in Fig. 5.2. The analysis results show that the proposed architecture provides minimal rates than the existing algorithm.

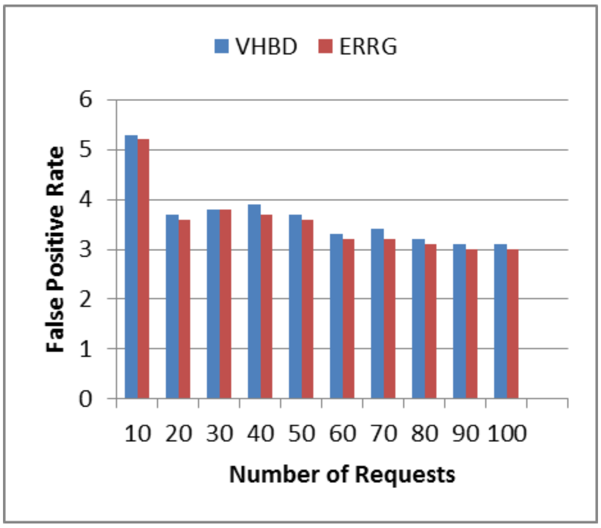

Fig 5.2 : Comparison of False Negative rate for the existing VHBD and the proposed ERRG

V CONCLUSION

Erdos – Renyi random graph and machine learning based approach with four transition phases are initialized to evaluate bots and its behavior and are arrived in this paper. The connected nodes (bots) are extended and are shown based on Erdos – Renyi random graph. The legitimate and regressive attackers are identified in the testing phase through machine learning algorithm. Hence in the concluding phase the bots spread as malware through its C & C servers and peer – to – peer networks are identified.

REFERENCES

1. Antonakakis M, Bailey M, Bernhard E, Bursztein Cochran Z., Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, Sullivan, K. Thomas, and Y. Zhou, “Understanding the Mirai Botnet,” in Proceedings of USENIX Security Symposium, pp. 1093–1110, 2017.
2. 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.
3. 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.
4. 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.
5. Boutaba R , M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, “A comprehensive survey on machine learning for networking: evolution, applications and research opportunities,” Journal of Internet Services and Applications, vol. 9, no. 1, pp. 1 – 99, 2018.
6. 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.
7. 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.
8. Ehrenfeld J M, “Wannacry, Cybersecurity and Health Information Technology: A Time to Act,” Journal of Medical Systems, vol. 41, no. 7, p. 104, 2017.
9. Garcia S, M. Grill, J. Stiborek, and A. Zunino, “An empirical comparison of botnet detection methods,” Computers & Security, vol. 45, pp. 100–123, 2014.
10. 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.
11. 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, pp. 1–4. August 2010.
12. 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.
13. K Karasaridis, B. Rexroad, D. A. Hoeflin et al., “Wide-scale botnet detection and characterization” HotBots, vol. 7, pp. 7–7, 2007.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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, pp. 393–400, 2014.
19. 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.
20. 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.
21. 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.

AUTHORS PROFILE

Dr. K. Akilandeswari, Associate Professor, Dept of Computer Science, Govt Arts College (Autonomous), Salem - 7, with 20 years of teaching experience and 15 publications in peer reviewed journals.

Mr. L. Baranivel, MCA, pursuing his M.Tech by research at VIT, Vellore.

Mr. G. Anwar Basha. an independent researcher with 10 years of teaching experience.