Journal Homepage: -www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI:10.21474/IJAR01/8569
DOI URL: http://dx.doi.org/10.21474/IJAR01/8569
ISSN NO. 2320-5407

RESEARCH ARTICLE

DETECTION OF BOTNETS USING INVARIANT REPRESENTATION.

Moinak Bhattacharya[1] and V.Bhattacharya[2].
1. SRM Institute of Science and Technology, Kattankulathur.
2. Birla Institute of Technology, Mesra, Ranchi.

Manuscript Info

Manuscript History
Received: 14 December 2018
Final Accepted: 16 January 2019
Published: February 2019

Key words:-
Botnets, Bot herder, C & C channels, Invariant Representation, Histogram Representation, Two-class SVM.

Abstract

Over the past few decades, botnets are known to be a serious threat to the cyber security. The botnets are the systems in a particular network environment that are commanded by the attacker also known as Bot herder through C & C channel and hence targets the neighbour systems. As a result, several anomalies(such as DDoS, spamming, key-logging etc) are detected which leads to failure of the systems, information breach and also threat to security. With the advancement of technology, botnets tend to change their feature and pattern of attack and tend to be indomitable. In the proposed architecture, we derieved a methodology to effectively detect problematic botnets irrespective of their variance in features and attack pattern. Invariant representation is implemented to effectively detect the botnets and keep in view the feature of invariance and the architechture is evaluated using bin histogram representations and two-class SVM(Support Vector Machine).

Introduction:-

Over last 15 years, botnets had been the most vexing cyber-security threats, which caused many devastating and costly threats to Internet Security. About 15 to 20 percent of the computers connected to the Internet are infected and are used by Botnets[7]. A victim host becomes a bot in the Botnetwork and controlled by a human(Bot herder) and numerous controllers(Botnets) through a Command and Control (C & C) communication channel. The attacker also known as ‘bot herder’, ‘botmaster’, or ‘controller’ commands the vulnerable victim host to perform attacks such as Distributed Denial of Service(DDoS) attacks with several fraudulent activities featuring spamming of other hosts in the system, security breach, information identity theft and exfiltration, malware dissemination, click-fraud and many more[3-6]. All these activities are performed in a specific manner through the C & C channel. Centralized C & C structures using the Internet Relay Chats(IRC) protocol is used by vast majority of Botnets[2] which is featured by more flexibility to the attackers, as it provides instant interaction with more number of bots and with more efficiency[20,26] and P2P(Point to Point) protocols, which does not have a central C & C server, and all the bots will be connected to each other to get controlled. That is why P2P botnet doesnot suffer from a single point failure[14].

Some of the botnets uses HTTP/HTTPS protocols[3,16,17,18], as HTTP-based C & C communications are allowed in most networks. It has been documented that a single botnet is capable of infecting about 4,00,000 systems and simultaneously keeping them under their control[8]. Several methodologies are proposed with an approach to detect the existance of botnets in monitored network. Most of the approaches are based on detection of botnets that uses IRC or HTTP-based C & C communication channel[9-13]. BotHunter[19] uses the concept of detecting the botnets

Corresponding Author:-Moinak Bhattacharya.
Address:-SRM Institute of Science and Technology, Kattankulathur.

by categorizing the bot behaviour, which follows a predefined infection Life-Cycle dialog model. According to recent studies, botnets can change its C & C server address using fast-flux service networks. Hence some more approaches may prove ineffective against these changes. Therefore, robustness and invariance of the features extracted from raw samples determines the change in botnet behaviour by implementing botnet samples (FEATURES) difference with training samples and test samples.
In this work, an invariant representation of sample traffic is implemented which is invariant under shifting and scaling of the features and permutation of the channelized representation. This is studied with Histogram representations with self-similarity matrix for each channel.

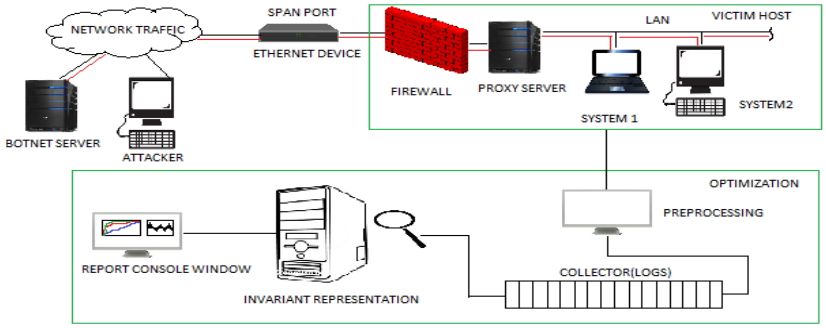

Figure 1:-The diagram of the proposed architecture. The collector logs are connected to the victim host through a preprocessor that channelizes the botnets and then transformed to $\tau(C, \delta, \Phi)$ for invariant representation and results are displayed in console window.

Related Works:-
A botnet is a group of targeted hosts that executes the specified commands of the botmaster. Botnets targets the nearby hosts by exploiting the vulnerability of the security if the victim host. Botnets tends to borrow several strategies from different types of malwares, including self- replicating worms, email-virus etc. and tends to diverge their features which makes botnets more dangerous. In this section we will discuss about the strategies attempted to detect botnets and efficiently detect their features.

Several researchers have proposed several approaches[9,10,11,12,13,19,20,21] to detect the existence of botnets in networks. Dewes et al.[22] introduced the concept for identifying the chat traffic. Sen et al.[23] used a signature based scheme to discern traffic produced by several well known P2P applications. Rishi[10] uses known IRC bot nickname patterns as signature to detect IRC botnets. BotSniffer[20] detects C & C activities with protocols such as HTTP and IRC[24].

Some of the other botnet detection methods have classified botnet detection into active and passive detections proposed by Daniel et al.[25] and characterization of behaviors namely Network Based Behaviors, Host Based Behavior and Global Correlated Behaviour based on Trends Micro's Report[26].

Explanation:-
According to the Figure 1, the congestion of the network traffic with the problematic elements refers to the fact that the botnets that simultaneously tends to attack the neighbour systems is commanded by any specific and channelized instruction sequence. This particular behaiviour of the botnets and their nature of attacks are organized and subsequently arranged into different channels.

The idea of channelizing the information leads to a different approach that fully determines the expression of features in specific format. The aspect kept in mind is that the system must be designed such that the features show invariant properties. This is attained by the representation of features in Invariant form. Invariant Representation specifies the response of features to the changes that can either be scale or shift variance. The representation also specifies invariance to changes in the arrangement of the channels.

Right from the initial status, when the attacker issues the commands to the vulnerable botnet to the classification of problematic and non-problematic botnets and thereby detection of the problematic botnets efficiently, the proposed architecture shows robustness in nature and exemplifies efficiency in working.

Experimental Characteristics:-

The proposed architecture in the figure deals with the problem of presenting a robust representation of network traffic communication that would be sufficiently invariant against modifications any Botmaster can deploy to evade the detection system. The invariant representation classifies the network into specified channels that will contain the malicious and Non-malicious botnet.

To evaluate the prototype, we have tested its performance on real network traffic traces for several times including normal data from BIT Mesra campus network and collected sample data simultaneously represented as an N-dimensional feature vector. Samples are then channelized into M-Channels. Here, Channels are represented as C. The channel is now represented as transformation of three stages of scaling and shifting Invariance of features and permutation(assumed that size of the channel is pre-determined) of the channel denoted as τ. The Invariant Representation is explained in three major heads:-

1. 1.Scale Invariance features Matrix Factorization(MF) in which an N feature vectors x M Channels is modeled by the product of an MxK channel factor matrix and the transpose of an NxK feature vector Factor Matrix[1,28].
2. Shift Invariance features Self-Similarity Matrix δr, represented as δ(Cαr,Cβr) where Cαr,Cβr are the feature vectors of the αth and βth sample from channelC.
3. Permutation Invariant representation features the invariance in size and the order of recording of the channels. For the representations of the histogram, edges of the bins are represented as Φ. The analysis is done by two representations, one will evaluate the values corresponding to the features and the other will evaluate on the basis of difference in the features of botnet samples. Subsequently they are compared and the detection of botnets are evaluated using classifiers.

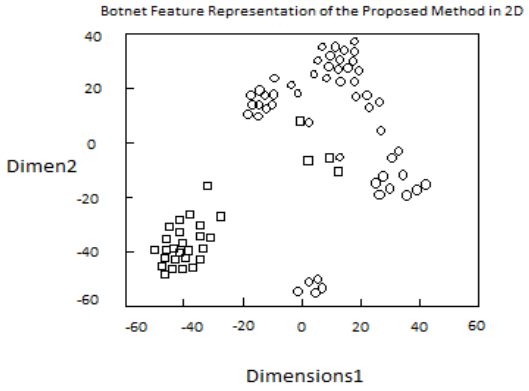

Figure 2:-Graphical representation of the botnet features in two dimensional using t-SNE transformations. The non-problematic botnets are concentrated in the bottom-left and at a considerable distance from the problematic botnets. This represents that the proposed model will achieve a higher efficiency by properly training the classifier. Here square represents non-malicious samples and circles represent malicious samples.

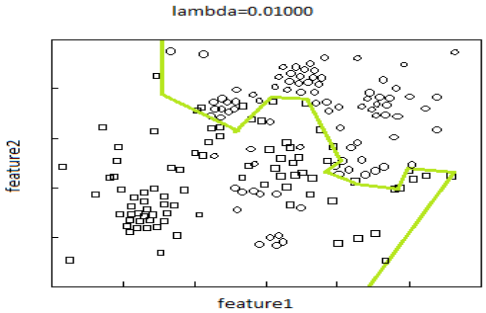

Figure 3:-Implementation of Invariant representation for the collected samples and projected in 2D. The decision boundary(green) is represented of two-class classifier with λ=0.01000. The value of λ determines the histogram representation and the smoothness of the decision boundary.

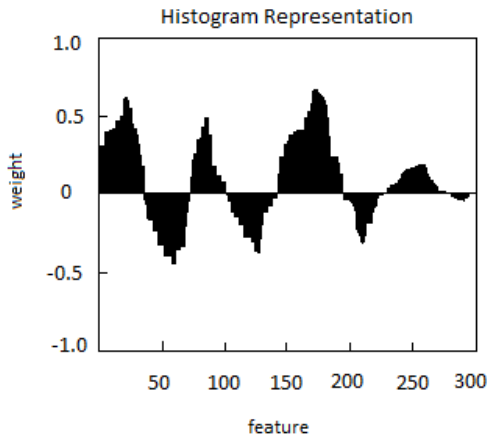

Figure 4:-Histogram Representation of the proposed architecture. The parameters of the bins are changed and evaluated to get a smoother decision boundary and robust representation. The change in value of λ signifies the characteristics of the histogram.

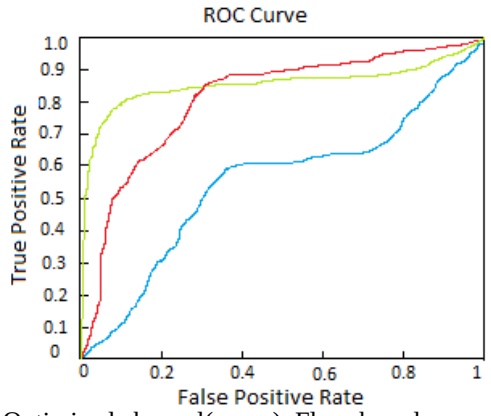

Figure 5:-Representation of Optimized channel(green), Flow-based representation(blue) and Collaborated-Channel(red) flow using ROC(Receiver Operating Characteristics) curve of SVM classifier. Here the initial results of flow-based differs from the final and proved to be an inefficient method. Optimized Channel method shows the efficient result and by changing the parameters of Invariant Representation, the results improved further.

Evaluation And Result:-

The proposed architecture was implemented in real network at BIT Mesra Campus to detect botnet samples at different interval of time. Two-class classification method is deployed that transforms the features as $\tau(C, \delta, \Phi)$. Different studies are made by regularizing the specified bins of the histogram and at different intervals as shown in Figure 4 and Figure 6 . The training data and test data was evaluated using a two-class SVM classifier. The result on the test data is shown in Figure 5 and Figure 7. With these specifications, the proposed method achieves 90% efficiency in detection botnets besides the change in features of the same at different interval of time.

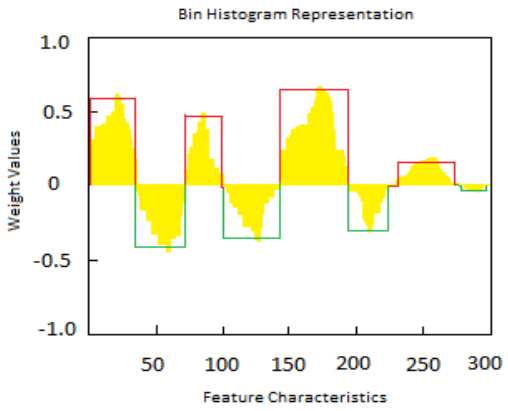

Figure 6:-The representation describes the proposed architecture transformation with Yellow Bars as weights and Red and Green lines as representation of bins. This figure shows the robustness of the architecture and smooth decision boundary.

Conclusion:-

Future Works:-

The proposed architecture is capable of detecting botnets in congested network traffic and is able to classify the activities of the botnets irrespective of the change in features and behavior of the mentioned. It works with the methodology of channelizing the flows and representing them to sustain the invariance in properties. Subsequently the results are analyzed and the method was found to be 90% efficient.

Furthermore, with the gradual advent of technology, the threats to cyber security increases creating a path for deep research in the malicious behavior of the botnets and the study of change in features of botnets may lead to the early detection of any problem related to security breach, leading to the eradication of any kind of potential threat to network.

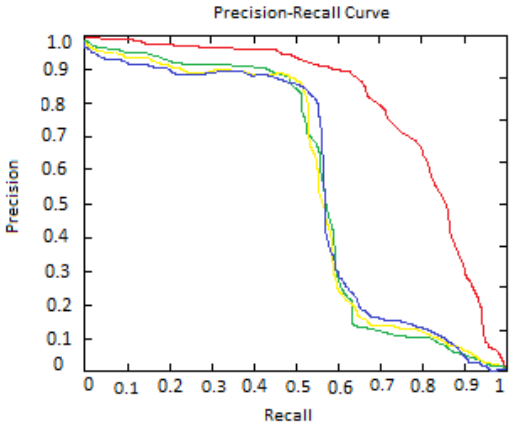

Figure 7:-The figure shows the precision recall curve of SVM classifier on the variance in features. The classifier worked with sufficient efficiency of 90% and achieved near about 80%recall of the undetected malware. Combined-Red,Bins-others.

Acknowledgements:-

This paper was done during Research Internship at BIT Mesra, Ranchi under Prof.Dr.V.Bhattacharya, Head of the Department, Department of Computer Science and Engineering, BIT Mesra. I would like to thank Arindam Chakraborty, King Abdullah University of Science and Technology (KAUST) for his insightful comments of drafting and helping with the technical aspects. I would also like to thank the entire Department of Computer Science, BIT Mesra and SRM IST, Kattankulathur for developing the specific environment for Research & Development.

References:-

1. K. Bartos, M. Sofka, and V. Franc. Optimized invariant representation of network traffic for detecting unseen malware variants. In 25thUSENIX Security Symposium (USENIX Security 16), pages 807–822,Austin, TX, Aug. 2016. USENIX Association.
2. C. Kalt. Internet Relay Chat: Client Protocol. RFC 2812(Informational), April 2000.
3. Ianelli N, Hackworth A. Botnets as a vehicle for online crime; 2005. https://resources.sei.cmu.edu/asset_files/White-Paper/2005_019_001_51249.pdf.
4. Bacher P, Holz T, Kotter M, Wicherski G. Know your enemy: tracking botnets; 2008. https://www.honeynet.org/papers/bots/.
5. Kaspersky. What is botnet attack?2016. https://usa.kaspersky.com/internet-security-center/threats/botnet-attacks#.V1du3TUrIdV.
6. Sonawane SR. A review on botnet and botnet detection methods. Int J Comput Sci Innov. 2016.
7. Emerging cyber threats report for 2009, Georgia Tech Information Security Center, Oct. 2008.
8. D. Ramsbrock, X. Wang, and X. Jiang, A first step towards live botmaster traceback," in Recent Advances in Intrusion Detection, vol. LNCS 5230, pp. 59-77, Springer-Verlag, 2008.
9. J. R. Binkley and S. Singh. An algorithm for anomaly-based botnet detection in Proceedings of USENIX SRUTI'06, pages 43-48, July 2006.
10. J. Goebel and T. Holz.Rishi: Identify bot cont aminated hosts by irc nickname evaluation.In Proceedings of USENIX HotBots'07, 2007.
11. Karasaridis, B. Rexroad, and D. Hoeflin. Wide-scale botnet detection and characterization.In Proceedings of USENIX HotBots'07, 2007.
12. Livadas, R. Walsh, D. Lapsley, and W. T. Strayer.Using machine learning techniques to identify botnet traffic.In Proceedings of the 2nd IEEE LCN Workshop on Network Security (WoNS'2006), 2006.
13. W. T. Strayer, R. Walsh, C. Livadas, and D. Lapsley. Detecting botnets with tight command and control.In Proceedings of the 31st IEEE Conference on Local Computer Networks (LCN'06), 2006.
14. Grizzard, J. B., Sharma, V., Nunnery, C., Kang, B. B. and Dagon, D. 2007. Peer-to-peer botnets: overview and case study, Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets, USENIX Association, Berkeley, CA, USA.
15. M. Rajab, J. Zarfoss, F. Monrose, and A. Terzis.A multi-faceted approach to understanding the botnet phenomenon.In Proceedings of ACM SIGCOMM/USENIX Internet Measurement Conference (IMC'06), Brazil, October 2006.
16. K. Chiang and L. Lloyd. A case study of the rustoc k rootkit and spam bot. In Proceedings of USENIX HotBots'07, 2007.
17. N. Daswani and M. Stoppelman.The anatomy of clickbot.A.In Proceedings of USENIX HotBots'07, 2007.
18. SecureWorks.Bobax trojan analysis.
19. http://www.securework s.com/research/threats/bobax/, 2004.
20. G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee. BotHunter: Detecting malware infection through ids-driv en dialog correlation.In Proceedings of the 16th USENIX Security Symposium (Security'07), 2007.
21. G. Gu, J. Zhang, and W. Lee. BotSniffer: Detecting botnet c ommand and control channels in network traffic.In Proceedings of the 15th Annual Network and Distributed System Security Symposium (NDSS'08), 2008.
22. M. K. Reiter and T.-F. Yen.Traffic aggregation for malw are detection.In Proceedings of the Fifth GI International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA'08),2008.
23. C. Dewes, A. Wichmann, and A. Feldmann. An analysis of internet chat systems.In IMC '03: Proceedings of the3rd ACM SIGCOMM conference on Internet measurement,pages 51–64, New York, NY, USA, 2003. ACM Press.
24. S. Sen, O. Spatscheck, and D. Wang. Accurate, scalable in-network identification of p2p traffic using application signatures.In WWW '04: Proceedings of the 13th international conference on World Wide Web, pages 512–521, New York,NY, USA, 2004. ACM Press.
25. A guide to understanding covert channel analysis of trusted systems, version1.NCSC-TG-030,Library No. S-240,572, National Computer Security Center, November 1993.
26. D. Plohmann, E. Gerhards-Padilla, and F. Leder, Botnets: measurement, detection, disinfection and defence," in ENISA workshop on (Giles Hogben, ed.),Mar. 2011
27. Taxonomy of botnet threats, white paper, Trend Micro Incorporated, Nov. 2006.White Paper.
28. M. Mahmoud, M. Nir, and A. Matrawy,Survey on botnet architectures, detection and defences.In International Journal of Network Security.
29. Andrea Vedaldi, "Invariant Representation and Learning for Computer Vision," Doctoral thesis, 2008.