International Journal on Applications in Information and Communication Engineering
Volume 1: Issue 10: October 2015, pp 11-14. www.aetsjournal.com ISSN (Online) : 2394-6237

BOTNET Detection Based on Multivariate Correlation Analysis

S.Revathi, V.Nagaraj

Abstract— A botnet is a collection of compromised hosts that are remotely controlled by an attacker (the botmaster) through a command and control (C&C) channel. Peer-to-peer (P2P) botnets have recently been adopted by bot masters for their resiliency against take-down efforts. Besides being harder to take down, modern botnets tend to be stealthier in the way they perform malicious activities, making current detection approaches ineffective. In addition, the rapidly growing volume of network traffic calls for high scalability of detection systems. In this paper, propose a novel scalable botnet detection system capable of detecting stealthy P2P botnets. The system first identifies all hosts that are likely engaged in P2P communications. It then derives statistical fingerprints to profile P2P traffic and further distinguish between P2P botnet traffic and legitimate P2P traffic. The parallelized computation with bounded complexity makes scalability a built-in feature of system. During this paper, tend to present a DOS attack detection system that uses variable Correlation Analysis (MCA) for accurate network traffic characterization by extracting the geometrical correlations between network traffic options. MCA-based DOS attack detection system employs the principle of anomaly-based detection in attack recognition. Extensive evaluation has demonstrated both high detection accuracy and great scalability of the proposed system.

Keywords: Botnet, Denial of Service attack, Multivariate correlation analysis, security, accuracy.

I. INTRODUCTION

A modern definition of bot is a concept of advanced malicious software that incorporates usually one or more aspects of the aforementioned techniques introduced by viruses, worms, Trojan horses and root kits for propagation and hostile integration into a foreign system, providing the functionality of the compromised system to the attacker. A defining characteristic of bots is that they connect back to a central server or other infected machines after successfully compromising the host system, thus forming a network. This network is the so-called botnet. The bots provide a range of implemented features to a corresponding controlling entity. This entity is commonly a command-and-control server under the control of one or more persons called the bot masters or bothered, who relay commands through this server. Depending on the network infrastructure, the bots may be connected with each other to enable this desired control structure. Alternatively, they can exist completely independently, not knowing of the existence of other bots.

S.Revathi, PG Scholar, Department of Computer Science and Engineering, Shanmuganathan Engineering College, Arasampatti-622507.

V.Nagaraj, Assistant Professor, Department of Computer Science and Engineering, Shanmuganathan Engineering College, Arasampatti-622507

A BOTNET is a collection of Internet-connected programs communicating with other similar programs in order to perform tasks. This can be as mundane as keeping control of an Internet Relay Chat (IRC) channel, or it could be used to send spam email or participate in distributed denial-of-service attacks. The word botnet is a combination of the words robot and network. The term is usually used with a negative or malicious connotation.

A BOTNET is a collection of compromised hosts (a.k.a. bots) that are remotely controlled by an attacker (the bot master) through a command and control (C&C) channel. Botnets serve as the infrastructures responsible for a variety of cyber-crimes, such as spamming, distributed denial of Service (DDoS) attacks, identity theft, click fraud, etc. The C&C channel is an essential component of a botnet because bot masters rely on the C&C channel to issue commands to their bots and receive information from the compromised machines. Botnets may structure their C&C channels in different ways. In a centralized architecture, all bots in a botnet contact one (or a few) C&C server(s) owned by the bot master. However, a fundamental disadvantage of centralized C&C servers is that they represent a single point of failure. In order to overcome this problem, bot masters have recently started to build botnets with a more resilient C&C architecture, using a peer-to-peer (P2P) structure or hybrid P2P/centralized C&C structures

DENIAL OF SERVICE (DOS) attacks square measure one form of aggressive and ugly intrusive behavior to online servers. DOS attacks severely degrade the supply of a victim, which might be a number, a router, or a complete network. They impose intensive computation tasks to the victim by exploiting its system vulnerability or flooding it with quantity of useless packets. The victim may be forced out of service from a couple of minutes to even many days. This causes serious damages to the services running on the victim. Therefore, effective detection of DOS attacks is crucial to the protection of on-line services. Work on DOS attack detection primarily focuses on the event of network-based detection mechanisms. Detection systems supported these mechanisms monitor traffic transmittal over the protected networks.

These mechanisms unleash the protected on-line servers from observation attacks and make sure that the servers will dedicate themselves to supply quality services with minimum delay in response. Moreover, network-based detection systems square measure loosely including operative systems running on the host machines that they're protective. As a result, the configurations of network primarily based detection systems square measure easier than that of host-based detection

systems. Analysis community, therefore, began to explore the simplest way to realize novelty-tolerant detection systems and developed a lot of advanced thought, specifically anomaly primarily based detection. Because of the principle of detection, that monitors and flags any network activities presenting vital deviation from legitimate traffic profiles as suspicious objects, anomaly-based detection techniques show less dimmed in police work zero-day intrusions that exploit previous unknown system vulnerabilities.

Moreover, it's not strained by the experience in network security, because of the actual fact that the profiles of legitimate behaviors square measure developed supported techniques, like data processing machine learning and applied math analysis. However, these projected systems unremarkably suffer from high false positive rates as a result of the correlations between features/attributes square measure as such neglected or the techniques don't manage to totally exploit these correlations. The DOS attack detection system bestowed during this paper employs the principles of MCA and anomaly based detection. They equip our detection system with capabilities of correct characterization for traffic behaviors and detection of celebrated and unknown attacks severally. A triangle space technique is developed to reinforce and to hurry up the method of MCA. A applied math standardization technique is employed to eliminate the bias from the data.

DOS detection system is evaluated victimization KDD Cup ninety nine dataset and outperforms the state-of the- art systems. The remainder of this paper is organized as follows. The summary of the system design in presents a completely unique MCA technique. It describes our MCA-based detection mechanism. To evaluates the performance of our projected detection system victimization KDD Cup ninety nine dataset. Support Vector Machines (SVM) could be a powerful, progressive algorithmic rule with sturdy theoretical foundations. SVM supports each regression and classification tasks and may handle multiple continuous and categorical variables. To construct Associate in Nursing optimum hyper plane, SVM employs Associate in Nursing unvarying coaching algorithmic rule, that is employed to attenuate a slip operate.

II. RELATED WORKS

P. Garca-Teodoro, J. Daz-Verdejo, G. Maci-Fernndez, and E. Vzquez [3] This paper begins with a review of the foremost well-known anomaly-based intrusion detection techniques. Then, offered platforms, a system below development and analysis comes within the space square measure conferred. Finally, the foremost vital open problems concerning A-NIDS square measure known, among that that of assessment is given specific stress.

A.Tajbakhsh, M.Rahmati, and A. Mirzaei [1] A GNP-based fuzzy class-association-rule mining with sub attribute utilization and also the classifiers supported the extracted rules are planned, which might systematically use and mix distinct and continuous attributes in an exceedingly rule and with efficiency extract several sensible rules for classification. As associate degree application, intrusion-detection classifiers for each misuse detection and anomaly detection are developed and their effectiveness is confirmed victimization KDD99Cup and DARPA98 knowledge. The experimental leads to the misuse detection show that the planned technique shows high DR and low PFR, those square measure 2 vital criteria for security systems. within the anomaly detection, the results show high DR and affordable PFR even while not pre knowledgeable information, that is a vital advantage of the planned technique.

W. Wang, X. Zhang, S. Gombault, and S. J. Knapskog [5] Anomaly intrusion detection is a vital issue in electronic network security. As a step of information preprocessing, attribute normalization is important to detection performance. However, several anomaly detection ways don't normalize attributes before coaching and detection. Few ways bear in mind to normalize the attributes however the question of that normalization technique is more practical still remains. During this paper to introduce four completely different schemes of attribute normalization to preprocess the info for anomaly intrusion detection. 3 ways, k-NN, PCA also as SVM, square measure then utilized on the normalized knowledge also as on the first knowledge for comparison of the detection results. KDD Cup 1999 knowledge also as a true knowledge set collected in our department square measure wont to assess the normalization schemes and also the detection ways. The systematical analysis results show that the method of attribute normalization improves lots the detection performance.

M. Tavallaee, E. Bagheri, L. Wei, and A. A. Ghorbani [6] proposed a replacement knowledge set, NSL-KDD that consists of hand-picked records of the entire KDD knowledge set. This knowledge set is publically offered for researchers through our web site and has the subsequent blessings over the first KDD knowledge set: It doesn't embrace redundant records within the plaything, therefore the classifiers won't be biased towards additional frequent records. There are not any duplicate records within the planned check sets; so, the performances of the learners don't seem to be biased by the ways that have higher detection rates on the frequent records. the quantity of hand-picked records from every problem level cluster is reciprocally proportional to the proportion of records within the original KDD knowledge set. As a result, the classification rates of distinct machine learning ways vary in an exceedingly wider vary, that makes it additional economical to own associate degree correct analysis of various learning techniques.

G. Thatte, U. Mitra, and J. Heidemann [8] This paper develops constant ways to observe network anomalies victimization solely combination traffic statistics, in distinction to alternative works requiring flow separation, even once the anomaly could be a tiny fraction of the overall traffic. By adopting straightforward applied mathematics models for abnormal and background traffic within the time domain, one will estimate model parameters in real time, therefore preventive the requirement for a protracted coaching section

or manual parameter calibration. The planned quantity constant Detection Mechanism (BPDM) uses a ordered chance quantitative relation check, giving management over the false positive rate whereas examining the trade-off between detection time and also the strength of associate degree anomaly. in addition, it uses each traffic-rate and packet-size statistics, yielding a quantity model that eliminates most false positives.

III. DESCRIPTION OF THE PROPOSED SCHEME:

The basic features are generated from ingress network traffic to the internal network where protected servers reside in and are used to form traffic records for a well-defined time interval. Monitoring and analyzing at the destination network reduce the overhead of detecting malicious activities by concentrating only on relevant inbound traffic. This also enables our detector to provide protection which is the best fit for the targeted internal network because legitimate traffic profiles used by the detectors are developed for a smaller number of network services.

A. Multivariate Correlation Analysis

DOS attack traffic behaves differently from the legitimate network traffic, and the behavior of network traffic is reflected by its statistical properties. To well describe these statistical properties, a novel Multivariate Correlation Analysis (MCA) approach employs triangle area for extracting the correlative information between the features within an observed data object (i.e., a traffic record). Multivariate Correlation Analysis, in which the "Triangle Area Map Generation" module is applied to extract the correlations between two distinct features within each traffic record coming from the first step or the traffic record normalized by the "Feature Normalization" module. The occurrence of network intrusions cause changes to these correlations so that the changes can be used as indicators to identify the intrusive activities. All the extracted correlations, namely triangle areas stored in Triangle Area Maps (TAMs), are then used to replace the original basic features or the normalized features to represent the traffic records. This provides higher discriminative information to differentiate between legitimate and illegitimate traffic records. Individual network traffic records rather than model network traffic behavior of a group of network traffic records. These results in lower latency in decision making and enable sample-by-sample detection.

B. Anomaly-Based Detection Mechanism

The frequent update of the attack signature database in the case of misuse-based detection are avoided. Meanwhile, the mechanism enhances the robustness of the proposed detectors and makes them harder to be evaded because attackers need to generate attacks that match the normal traffic profiles built by a specific detection algorithm. This, however, is a labor-intensive task and requires expertise in the targeted detection algorithm. Specifically, two phases (i.e., the "Training Phase" and the "Test Phase") are involved in Decision Marking. The "Normal Profile Generation" module is operated in the "Training Phase" to generate profiles for various types of legitimate traffic records, and the generated normal profiles are stored in a database. The "Tested Profile Generation" module is used in the "Test Phase" to build profiles for individual observed traffic records. Then, the tested profiles are handed over to the "Attack Detection" module, which compares the individual tested profiles with the respective stored normal profiles. A threshold-based classifier is employed in the "Attack Detection" module to distinguish DOS attacks from legitimate traffic.

C. Performance Evaluation

The internal network where protected servers reside in and are used to form traffic records for a

well-defined time interval. Monitoring and analyzing at the destination network reduce the overhead of detecting malicious activities by concentrating only on relevant inbound traffic. Multivariate Correlation Analysis, in which the "Triangle Area Map Generation" module is applied to extract the correlations between two distinct features within each traffic record coming from the first step or the traffic record normalized by the "Feature Normalization" module in this step. The occurrence of network intrusions cause changes to these correlations so that the changes can be used as indicators to identify the intrusive activities. Triangle Area Maps (TAMs) are then used to replace the original basic features or the normalized features to represent the traffic records. This provides higher discriminative information to differentiate between legitimate and illegitimate traffic records. The anomaly-based detection mechanism is adopted in Decision Making. Specifically, two phases are involved in Decision Making. The "Normal Profile Generation" module is operated in the "Training Phase" to generate profiles for various types of legitimate traffic records, and the generated normal profiles are stored in a database. The "Tested Profile Generation" module is used in the "Test Phase" to build profiles for individual observed traffic records. Then, the tested profiles are handed over to the "Attack Detection" module, which compares the individual tested profiles with the respective stored normal profiles. A threshold-based classifier is employed in the "Attack Detection" module to distinguish DOS attacks from legitimate traffic.

IV. IMPLEMENTATION RESULTS

A. Computational complexity:

Computational complexity theory is a branch of the theory of calculation in mathematics focuses on classifying computational problems according to their inherent difficulty, and relating those classes to each other.

B. Communication overhead:

Communication Overhead is the proportion of time you spend communicating with your team instead of getting productive work done. Communication Overhead is the time spent waiting for an event to occur on a new task. In certain

modes, the sender must wait for the receive to be executed and for the handshake to arrive before the message can be transferred.

C. Message Integrity

The message receiver should be able to verify whether the message has been modified en-route by the adversaries. In other words, the adversaries cannot modify the message content without being detected.

D. Detection accuracy:

The approach improves detection accuracy; it is vulnerable to attacks that linearly change all monitored features. Proposed detection system is required to achieve high detection accuracy.

V. CONCLUSION

This paper has presented a MCA-based DOS attack detection system which is powered by the triangle-area based MCA technique and the anomaly-based detection technique. The former technique extracts the geometrical correlations hidden in individual pairs of two distinct features within each network traffic our system to be able to distinguish both known and unknown DOS attacks from legitimate network traffic. The DOS attack detection system presented in this paper employs the principles of MCA and anomaly-based detection. They equip our detection system with capabilities of accurate characterization for traffic behaviors and detection of known and unknown attacks respectively. The rest of this paper is organized as give the overview of the system architecture a novel MCA technique. To describes our MCA-based detection mechanism. To evaluates the performance of our proposed detection system using KDD Cup 99 dataset. The systematic analysis on the computational complexity and the time cost of the proposed system.

REFERENCES

[1] "Intrusion detection using fuzzy association rules," A.Tajbakhsh, M.Rahmati, and A. Mirzaei, Year-2009
[2] G. V. Moustakides, "Quickest detection of abrupt changes for a class of random processes," Information Theory, IEEE Transactions on, vol. 44, pp. 1965-1968, 1998.
[3] "Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges," P. Garca-Teodoro, J. Daz-Verdejo, G. Maci-Fernndez, and E. Vzquez, Year-2009.
[4] A. A. Cardenas, J. S. Baras, and V. Ramezani, "Distributed change detection for worms, DDoS and other network attacks," The American Control Conference, Vol.2, pp. 1008-1013, 2004.
[5] "Attribute Normalization in Network Intrusion Detection", W. Wang, X. Zhang, S. Gombault, and S. J. Knapskog, Year-2009.
[6] M. Tavallaee, E. Bagheri, L. Wei, and A. A. Ghorbani, "A Detailed Analysis of the KDD Cup 99 Data Set," The The Second IEEE International Conference on Computational Intelligence for Security and Defense Applications, 2009, pp. 1-6.
[7] "A Detailed Analysis of the KDD Cup 99 Data Set," M. Tavallaee, E. Bagheri, L. Wei, and A. A. Ghorbani,Year-2009.
[8] "Parametric Methods for Anomaly Detection in Aggregate Traffic," G. Thatte, U. Mitra, and J. Heidemann, Year-2011.
[9] C. Yu, H. Kai, and K. Wei-Shinn, "Collaborative Detection of DDoS Attacks over Multiple Network Domains," Parallel and Distributed Systems, IEEE Transactions on, vol. 18, pp. 1649-1662, 2007.
[10] W. Wang, X. Zhang, S. Gombault, and S. J. Knapskog, "Attribute Normalization in Network Intrusion Detection," The 10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN), 2009, pp. 448-453.