A New Method of Detecting Network Traffic Anomalies

CAI Jun
School of Electronic and Information
Guang Dong Polytechnic Normal University
Guangzhou, China
e-mail: gzhcaijun@gmail.com

LIU Wai Xi
Department of Electronic and Information Engineering
Guang Zhou University, Guangzhou, P. R. China
Guangzhou, China
e-mail: liuwaixi@gmail.com

***Abstract*—As Internet communications and applications become more and more complex, accurately describing network traffic information and rapidly monitoring network traffic anomalies have become increasingly challenging tasks. In this paper, we present a framework and method for monitoring network traffic through measuring the dynamic changes of host communities. An unweighted and undirected host interaction network (HIN) is established through extracting the social-behavioral characteristics of network traffic. Based on social-behavior similarity in HIN, host community is defined, and then five features are proposed to capture host community changes. Finally, this method is evaluated through two real-world network traffic, and the experimental results show that the method presented in this paper can effectively capture the dynamic changes of host communities to monitor network traffic.**

Keywords-*Complex networks, Host community, Host community change, Network traffic monitoring, Host interaction network.*

I. INTRODUCTION

Complex systems such as large-scale enterprise networks are dynamic systems with tens of thousands of peer hosts connected. Measuring and monitoring their behaviors are fundamental for design, security, development and implementation of networks. Traffic flow is one of the most important information sources for host behavior monitoring and abnormal behavior detection. The traditional network traffic analysis was focused on scaling phenomena, such as self-similarity, long-range dependence, heavy-tailed, fraction and multi-fraction. Current traffic analysis for network security and management has become an active area of research. These methods can be grouped by their level of observation as follows.

(1) Packet level [1]. These approaches such as signature-based anomaly detection can detect anomalous traffic by matching traffic data with predefined signatures that are created by analyzing traffic generated by malware. However, these techniques can not handle traffic encryption and unknown attacks.

(2) Network level [2]. These techniques can detect anomalous traffic from unknown malware if it is statistically different from normal traffic. However, more and more intelligent attacks try to avoid detection by generating normal behaviors in flow statistics [3].

(3) Host level [4]. These approaches such as host profiling can capture and represent host activities and behaviors for a variety of diagnosis and security applications. However, it is difficult to devise a permanent model of legitimate or anomalous host behavior that can be applied to every Internet location, considering the heterogeneity of Internet hosts and the dynamics of their activities.

To fill the gap between host level and network-level traffic behavior monitoring, this paper proposes a new perspective of monitoring network traffic by identifying and analyzing host community change. Focusing on the changes of host community behavior not only reduces the number of behavior profiles for analysis compared with host level traffic profiling, but also reveals detailed pattern changes for a group of hosts sharing similar behaviors compared with network-level traffic monitoring. Our studies are mainly inspired by the works investigated in [5], [6], which found that community memberships evolve slowly over short time scales. Following these studies, we first consider network traffic as a whole to extract general social-behavioral characteristics and construct the host interaction model (HIN), which is an unweighted and unidirectional graph denoted by *G(V, E)*, with *V* being the set of nodes and *E* the set of edges. Here the social-behavior of a particular host refers to with whom it communicates. In the HIM, the IP addresses (hosts) are the set of vertices *V* and the logic links between the source IP address and the destination IP address are the set of edges *E*. Subsequently, based on social-behavior similarity in HIN, host community is defined. Each community consists of hosts that communicate with similar sets of servers, clients or peers, thus showing strong social-behavior similarity. In order to quantify host community changes, five features are defined. Four of them are proposed to represent host community changes across time windows, including community merge (*CM*), community split (*CS*), community form (*CF*), and community dissolve (*CD*). The last feature, the entropy of host distributing in communities (*EHDC*), captures host community changes at a particular time instance. Host community changes are usually caused by attacks and network environment changes (topology reconfiguration, network resources relocation, etc.). In order to differentiate abnormal changes of host community from normal changes, a baseline method is introduced to set up adaptive monitoring thresholds for the five features. Moreover, to capture the dynamics of network traffic, a sliding window mechanism is introduced with fixed time lengths. Finally, we apply our method to real-world traffic data and perform offline and online experiments. The results show that it can detect network traffic anomalies with the

Published by Atlantis Press, Paris, France.

group behavior characteristics, and it achieves high detection rates and low false positive rates.

The contributions of this paper are summarized as follows:

(1)We propose host interaction model (HIM) to represent communication patterns between source and destination IP addresses, define host community based on the social-behavior similarity in HIM and discover the stability of the host communities within relative short scale;

(2)We define four features to characterize and interpret host community change across time widows such as community merge (*CM*), community split (*CS*), community form (*CF*), and community dissolve (*CD*), and a feature at a particular time instance such as the entropy of host distributing in communities (*EHDC*);

(3) We introduce a baseline method to set up adaptive monitoring thresholds for five features, and a sliding window mechanism with fixed time lengths to online detect network traffic anomalies;

(4)We demonstrate the applications of exploring social behavior similarity of hosts in detecting anomalous traffic such as scanning activities, worms, denial of service attacks and centralized botnet attacks through real-world traffic traces.

The paper is structured as follows. In Section Ⅱ, we review the related work. In Section Ⅲ, we present the framework of the monitoring system. In Section Ⅳ, we apply our method to real-world traffic. Finally, we make some conclusions and the future work is discussed in SectionⅤ.

II. Type Style and Fonts

Analysis of complex network graphs has recently received considerable attention in the literature, mostly due to three particular developments: Watts and Strogatz's investigation of small-world networks [7], Barabasi and Albert's characterization of scale-free models [8], and Girvan and Newman's identification of the community structure in many networks [9].

Many approaches have been proposed to consider network traffic as a whole to extract general behavior characteristics, construct complex graphs by different means, and analyze their features, e.g. [10]-[14]. The profiling of "social" behavior of hosts was studied in BLINC [10] for the purpose of traffic classification. In BLINC, the notion of a "graphlet" was proposed to model a single host's flow patterns. TDG (Traffic Dispersion Graph) [11] was an aggregation of the graphlets of all hosts in a network for a particular key and modeled the social-behavior of hosts. Qi Liao et al. [12] proposed a visualization approach based on the hierarchical structure of similarity/difference visualization in the context of heterogeneous graphs. S. Nagaraja et al. [13] developed an inference algorithm to search botnet communication structures from the background communication graphs constructed from the collected network traffic. Kuai Xu et al. [14] used graph analysis to construct the bipartite graphs from host communication and then to generate the one-mode projection graphs for uncovering the social-behavior similarity among end hosts.

Various properties of complex network graphs have been widely studied. In particular, the community structure in network graphs has attracted the majority of research interest recently, which has included all kinds of applications in different research fields, such as: identifying terrorist organizations, social network analysis, metabolic network analysis, unknown protein function prediction, gene regulatory network analysis and gene identification, and web community mining based on keywords [15]. Currently, there are also some researches on network traffic from the perspective of the community structure [5], [16], [17]. The author in [5] defined community of interest (COI) for host communication in data network and evaluated the stability of COI. However, they statically defined 'server' and 'client' that were dynamic for some network traffic in practice. On the other hand, their work excluded the external network and focused on local network (intranet) hosts. In [16], COIs were used to automatically generate host-level firewall rules so as to suppress worm behavior within a LAN. PRIMED [17] was a proactive approach to DDoS mitigation, in which ISPs constructed a network-wide bad COI that contained network entities who exhibited unwanted behavior in the past, and per-customer good COIs containing remote network entities that have previously engaged in legitimate communication with the customer.

Enlightened by the above work, in this paper, we propose a new method for measuring and monitoring network traffic anomalies based on the social-behavior changes of host community.

III. Framework of the monitoring system

The framework for monitoring network traffic based on host community change from the observation of traffic flow to raising an alarm is divided into four steps as shown in Fig.3.

Step 1: Traffic data information collection. The network traffic data information is collected from capable source, such as routers, switches, firewalls, etc., through NetFlow records.

Step 2: Host interaction model and host community generation. The traffic packets are aggregated and converted into host interaction model. Based on the definition of host community, we construct a $hc_t \times h_t$ bit matrix to represent the relation between host-communities and hosts, where hc_t is the number of host communities and h_t is the number of hosts in the t^{th} ΔT_h.

Step 3: The measurement of dynamic changes of host communities. Based on the matrix time series, the feature values are calculated and applied to measure two categories of dynamic changes: the entropy of host distribution in communities at a particular time instant and the states of host communities between two neighboring time windows.

Step 4: The analysis of dynamic changes of host communities. Based on the measurement in Step 3 and the thresholds of different feature values, we can monitor network traffic.

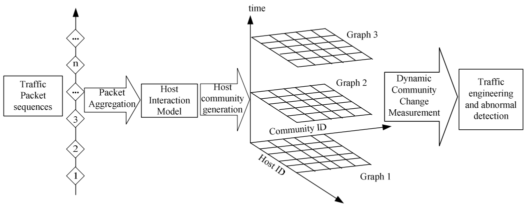

Figure 1. Framework of monitoring system

IV. PERFORMANCE EVALUATION

A. *Traffic Data*

In this study, we use network trafces from two different locations. One is Auckland IX trace set from University of Auckland [22]. This is a mostly continuous 2 day packet header trace taken in March 2008 at the link between the University of Auckland and the rest of the Internet. The traces used in our analysis were captured at March 27, 2008, which is denoted as Auckland- IX in this paper. IP addresses in the traffic traces are anonymized using one-to-one mapping anonymization for privacy reasons. However, the anonymization process does not affect our analysis. The other is collected from a single site in a large education environment, which serves a total host population in excess of 16000 hosts. The traces we use were captured in January 2008, which is denoted as Zsdx in this paper. The duration is two weeks.

B. *Online Validation of Abnormal Network Found*

To validate the abnormal network traffic found by the above method, we need to know the anomalies commonly encountered in the intranet networks as listed in Table Ⅰ [23] [24]. A software prototype is implemented to measure these feature values. We monitor the link between a large education network and the rest of the Internet for two weeks. Throughout the two-week duration, these associated anomalies are validated through manual analysis and other security devices already deployed. These anomalies generally affect certain features of host community change and can be captured by adaptive monitoring thresholds. Throughout the monitoring process, the detection ratio (DR) and the false positive ratio (FPR) of the methods are listed in Table Ⅱ together with the effective features that can detect these anomalies. The results show that the methods proposed in this paper can achieve high detection rates and low false positive rates in detecting the attacks with the group behavior characteristics. In addition, the total number of abnormal hosts (TNAH) and the abnormal percentage of the total hosts (APTH) from detection results are further observed. The minimum values of the TNAH and APTH are listed in Table Ⅲ. It is seen that these feature values are effective even if the abnormal percentage of the total hosts is small.

TABLE I. THE DR AND FPR OF THE METHOD AND THE EFFECTIVE FEATURES

Anomaly Type	**Effective Features**	**DR(%)**	**FPR(%)**
NS	f_1， f_4， f_5	0.94	0.61
Worms	f_2 , f_3 , f_5	0.92	0.33
DDoS	f_1， f_4， f_5	0.97	0.42
Centered Botnet	f_1， f_4	0.89	0.25

TABLE II. THE MIMUM VALUES OF TNAH AND APTH FROM THE DETECTION RESULTS

Anomaly Type	**Effective Features**	**TNAH**	**APTH**
NS	f_1， f_4， f_5	301	2.01%
Worms	f_2 , f_3 , f_5	261	1.92%
DDoS	f_1， f_4， f_5	552	3.46%
Centered Botnet	f_1， f_4	360	2.43%

V. CONCLUSIONS

In this paper, we present a framework and method from the perspective of social-behavioral characteristics of network traffic to detect possible network traffic anomalies with the group behavior characteristics in IP flow data collected from university's or enterprise border router. Host community is defined based on host social-behavioral characteristics. Five features are proposed to quantify host community changes and are found to be capable of capturing the dynamic changes of host communities, especially the changes caused by attacks, where DDoS and network scan attacks lead to the changes in *CM*, *CD*, *EHDC*, worm attacks adopting random scanning policy lead to changes in *CS*, *CF*, and *EHDC*, and centered Botnet leads to changes in *CM* and *CD*. We performed offline and online experiments in real-world networks to evaluate the method proposed in this paper. Experimental results show that the method is effective.

We successfully monitor the network traffic anomaly through the host community change. However, we can not accurately locate those communities lead to be abnormal. The future work is to define some parameters to differentiate abnormal communities from normal communities. In addition, the host interaction model is unweighted and undirected. Thus, our ongoing work is to further construct some directed and weighted host interaction models, for example, the edge may present the number of packets/bytes exchanged and its direction between hosts.

ACKNOWLEDGEMENT

We would like to thank all the members in our research group, for the valuable discussions about the ideas presented in this paper. Funding from Project supported by the National High Technology Research and from Development Program of China (Grant No.2007AA01Z449), the Key Program of NSFC-Guangdong Joint Funds under Grant

No.U0735002 and the National Natural Science Foundation of China under Grant No. 61202271.

REFERENCES

[1] P.Haffner, S.Sen, O.Spatscheck, and D.Wang, “ACAS: Automated Construction of Application Signatures,” In ACM SIGCOMM MineNet Workshop, Philadelphia, USA, 2005, pp. 197-202.

[2] K.Xu, Z.Zhang, S.Bhattacharyya, “Profiling Internet Backbone Traffic: Behavior Models and Applications,” In ACM SIGCOMM, Philadelphia, USA, 2005, pp. 169-180.

[3] G. Giorgi, C. Narduzzi, “Detection of Anomalous Behaviors in Networks from Traffic Measurements,” IEEE Trans On Instrumentation And Measurement , vol. 57,no. 12, pp. 2782-2791,2008.

[4] Thomas K, Konstantina P, Nina T., “Profiling the End Host,” In Proc. of the 8th international conference on Passive and active network measurement, Louvain-la-Neuve, Belgium, 2007, pp. 186-196.

[5] W.Aiello, C.Kalmanek, P.McDaniel. “Analysis of Communities of Interest in Data Networks,” In Proc. of the Passive and Active Network Measurement, Boston, MA, USA, 2005, pp. 83-96.

[6] Wei, S., J. Mirkovic, E. Kissel, “Profiling and Clustering Internet Hosts,” In Proc. of the International Conference on Data Mining, Las Vegas, Nevada, USA, 2006, pp.11-17.

[7] D.J. Watts and S.H. Strogatz, “Collective dynamics of small-world networks,” Nature,vol.393, no.664,pp. 440-442, 1998.

[8] Albert-László Barabási, Réka Albert, “Emergence of Scaling in Random Networks”. Science,vol. 286, no.5434, pp.509-512, 1999.

[9] M. Girvan and M.E.J., Newman, “Community structure in social and biological networks,” PNAS,vol.99,no.12,pp. 7812-7826, 2002.

[10] T.Karagiannis. K.Papagiannaki, and M.Faloutsons, “BLINC: multilevel traffic classification in the dark,” In ACM SIGCOMM, Philadelphia, PA, 2005, pp. 229-240.

[11] M. Iliofotou, P. Pappu, M. Faloutsos,S. Singh, and G. Varghese, “Network monitoring using traffic dispersion graphs (tdgs),” In Proc. of the 7th ACM SIGCOMM conference on Internet measurement, San Diego, USA, 2007, pp. 315-320.

[12] Qi Liao, Aaron Striegel and Nitesh Chawla, “Visualizing graph dynamics and similarity for enterprise network security and management,” In Proc. of the Seventh International Symposium on Visualization for Cyber Security, Ottawa, Ontario, 2010, pp. 34-45.

[13] S. Nagaraja, P.Mittal, C.-Y. Hong, M.Caesar, N.Borisov, “ BotGrep: Finding P2P Bots with Structured Graph Analysis,” In Proc. of USENIX Security Symposium, Washington, DC, 2010, pp. 1-16.

[14] Kuai Xu, Feng Wang, Lin Gu, “Network-Aware Behavioral Clustering of Internet end hosts,” In IEEE International Conference on Computer Communications (INFOCOM), Shanghai, China, 2011, pp. 2078-2086.

[15] M.E.J.Newman, “Detecting community structure in networks,” The European Physical Journal B - Condensed Matter and Complex Systems , vol.38, no.2, pp. 321-330, 2004.

[16] P. McDaniel, S. Sen, O. Spatscheck, “Enterprise security: a community of interest based approach,” In Proc. of Network and Distributed System Security (NDSS), San Diego, California USA, 2006, pp. 1-15.

[17] P.Verkaik, O.Spatscheck, “PRIMED: community-of-interest-based DDoS Mitigation,” In ACM LSAD Workshop, Pisa, Italy, 2006, pp.147-154.

[18] S.Asur, S.Parthasarathy, D.Uca, “An event-based framework for characterizing the evolutionary behavior of interaction graphs,” In Proc. of 13th ACM SIGKDD international conference on Knowledge Discovery and Data minning, San Jose, California, 2007, pp. 913-921.

[19] Greene.D, Doyle.D, Conningham.P., “Tracking the evolution of communities in dynamic social networks,” In Adcances in Social Networks Analysis and Mining, Odense, Denmark, 2010, pp. 44-54.

[20] S.Staniford, V.Paxson, and N.Weaver, “How to own the Internet in your spare time,” In Proceeding of the 11th USENIX Security Symposium, 2002, pp. 149-167.

[21] K.Ishibashi, T.Mori, R.Kawahara, “2-D Bitmap for Summarizing Inter-Host Communication Patterns,” In International Symposium on Applications and the Internet Workshops, Hiroshima, Japan, 2007, pp.83-87.

[22] http://www.wand.net.nz/wits/auck/9/auckland_ix.php.

[23] A.Lakhina, M.Crovella, and C.Diot, “Characterization of network-wide anomalies in traffic flows,” In Proc.4th ACM SIGCOMM Internet Measurement of Conf., Taormina, Italy, 2004, pp.201-206.

[24] R. Pang, M.Allman, M.Bennett, “A first at modern enterprise traffic,” In Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement, Berkeley, California, USA, 2005, pp.15-28.

Published by Atlantis Press, Paris, France.