Distributed Intrusion Alert Fusion Based on Multi Keyword

Ming XU
Institute of Computer Application Technology, Hangzhou Dianzi University, P. R. China
mxu@hdu.edu.cn

Wei HAN
Colleage of Science, Zhejiang SCI-TECH University, P. R. China
hanwei@ zist.edu.cn

Abstract

Intrusion alert fusion is a key problem in distributed intrusion detection system (DIDS). In this paper, we propose a distributed intrusion alert fusion scheme based on Multi Keywords. All the related alarms produced by local sensor can be evenly routed and fused to its corresponding sensor fusion centers (SFCs) by multi keywords, while evenly distributing unrelated alarms to different SFCs. We use DShield data collected from worldwide providers to evaluate feasibility of our scheme.

1. Introduction

Recently, P2P IDS is an active research field[1-3]. Chen[4] embed the attack symptoms into the DHT dimensions so that alarms related to the same intrusion will be routed to the same SFC, while evenly distributing unrelated alarms to different SFCs. Their scheme resembles our, but they use single keyword to route and fuse while we use multi keywords. Current DIDSs mainly deal with distributed audit collecting, and have a central coordinator or static hierarchical architecture. Most previous works presume that the local alert classification and identification is precise. We argue that the local classification to some alert may be imprecise because local sensor has limited view and detecting methods. So we should use multi potential usefulness keywords to alleviate disadvantage of the local imprecise classification, and implement multi points of view fusion.

2. The DHT network architecture

The network architecture of the distributed intrusion alert fusion system is a DHT P2P overlay. There are two types of nodes in the system: multiple heterogeneous IDS sensors and SFCs. We select some node as SFC, because it has more resources (e.g., more CPU and bandwidth) and better security measures as the SFCs. Past researches have shown that information sharing between these networks is an effective way to detect intrusion. We use DHT mechanism based on multi keywords to share and fuse alert information (see Fig 1). When a attack alert reporting, the multi keywords about intrusion symptoms are embed into the DHT dimensions so that alarms related to the similitude intrusion (with a same keyword) will be routed to the same peer to fuse, evenly distributing unrelated alarms to different peers.

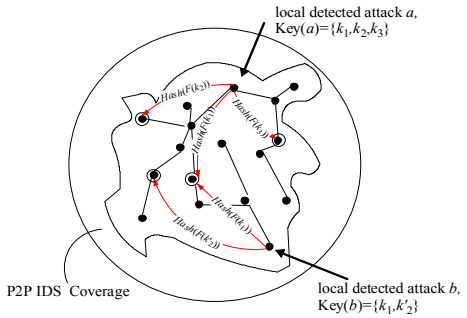

Figure 1. **DHT network of P2P IDS**

All DHT systems provide two basic interfaces: *put*( ) and *get*( ). The interface for insertion, *put*(*keyword*, *object*), causes the DHT to route the given object to the node with a node identifier closest to the keyword. The interface for retrieval, object=*get*(*keyword)*, causes the DHT to obtain the object from the node with a node identifier closest to the keyword. DHT systems can guarantee a deterministic routing in $O(\log n)$ hops for a DIDS system of n nodes. This implies that the system built on top of the DHT can be scalable to very large networks. When a locally sensor detect an attack, it generates a alert report that will be routed to the appropriate SFC with the *put*(k_i, *alert*). The keywords set of the alert, $\{k_1, k_2, \ldots, k_n\}$, will route the alert to the appropriate SFCs, thus the attacks which have different keyword are routed to different SFC. Based on the deterministic nature of the keywords routing, the attack which have the same keyword will route and reports to the same SFC. For an attack alert with a keywords set,

0-7695-3016-8/07 $25.00 © 2007 IEEE
DOI 10.1109/ISDPE.2007.20

every local sensor can query its prevalence with the *get*(*keyword*) operation.

3. Multi keywords

As intrusions are detected locally, the alert must be reported and routed to SFCs which can perform data fusion and inferences about such an attack. The challenge is come from two sides. One side, the local classifying and identifying to some alert may be imprecise because local sensor has limited view and detecting methods, For example, alert generated by a new unknown attack or statistics anomaly detection method. So the really meaning of alert may not be comprehended before global fusing. Moreover, an attack alert could denote different meanings from differently point of views. So the multi keywords, which could potential embody connotation of the alert, should be adopted to route and fusion. On the other hand, for a single intrusion, diverse symptoms are perceived from many heterogeneous IDSs.

To ensure that related event alert information will be routed to the appropriate SFCs, we must use the potentially intrinsic characteristics of each type of attacks alert as routing information. We use alerts of prevailing IDS detected intrusion by TCP/IP protocol as basic reference to construct keywords set. I.e. the source and destination IP address IP_s, IP_d, the source and destination port number P_s, P_d, are selected. When the attack can be identified or classified accurately, for example, the attack was detected by Snort used rule, intrinsic feature of the uniquely attack identification, such as some attack identification systems Bugtraq, CVE and Nessus, clearly should be a keyword. The Snort message ID S_{id} in sid-msg file can be used as a keyword, because the Snort message ID identify a known attack, and often give corresponding ID of Bugtraq, CVE, and Nessus.

4. Load balancing

In this section, we discuss load balancing problem. Internet attacks are increasing in frequency, severity and sophistication. All the related alarms produced by local sensor can be evenly routed and fused to its corresponding SFCs by multi keywords IP_s, IP_d, P_s, P_d, S_{id}. The multi keywords routing scheme can efficiently alleviate load balancing problem. However, when a global attack occurs, for example a DDos attack, many local IDSs will detect it and report alerts, which may easily overload some SFCs. Furthermore, some characterization keywords have strongly uneven popularity distribution, like port numbers 80 and 135.

When hot spots keywords are stable such as P_s and P_d [5], each stable hot spot keyword would route at least one dedicated SFC. We use 7 bits for the port number with 128 buckets. We map each of the 64 popular ports into one unique bucket, and map the other ports randomly into the remaining buckets. This effectively improves the load balancing especially when the SFC nodes are relatively few and thus sparse in the node ID space. Unfortunately, the distribution of IP_s and IP_d are only stability within hours, but not even a day[6].

Some keywords are unstable such as S_{id}, but the S_{id} distribution changing relate to the time. Thus we hope the fresh S_{id} keywords would route to dispersive SFCs, but the old S_{id} keywords may route a same SFC. We use 7 bits for the S_{id} number with 128 buckets. We map each of the freshest 32 S_{id} into one unique bucket, map the left freshest 128 S_{id} into 32 bucket, i.e. 4 S_{id} into one unique, and other S_{id} randomly into the remaining buckets.

5. Preliminary evaluation

The evaluation data set download from DShield[7]. DShield is the distributed Internet firewall and IDS log repository, and receives over 15 million intrusion alerts reported everyday. In this research, attacks number is number of distinct target addresses reporting hits from this source, and reports number is number of packets reported as originating from this IP address.

We use worm Witty which exploits a vulnerability in BlackIce's ICQ parser as a global attack example, because this worm generates large amounts of UDP traffic with source port 4000, and the wide spread use of BlackIce on Mar 20-22, 2004. In these three days, the DShield site denoted the internet was tracking a significant new threat. As shown in figure 2, figure (a) shows the number of reports with port 4000 and attack destinations port 4000 between MAY 1 to MAY 31, and figure (b) shows the number of attacks with source port 4000. In (a) and (b) figure, we find peak in evidence at Mar 20-22, 2004, but in figure (a) we also find some other peaks easily, so that we can detect anomaly easily using figure (b) but (a). More over, before worm Witty analyzed by expert, we do not know using source port 4000 can detection worm Witty. So using multi keywords IP_s, IP_d, P_s, P_d, S_{id} to route and fusing can easy detect this global attack by SFC of P_s, even before we analyze and understand his mechanism. We always have a period in which we do not know to the new attack before it is analyzed by experts, so that using multi keywords to route and fusing alert in P2P is necessary to detect it earlier.

Table 1. **The maximum events in top 10000**

Max	IP Address	Attacks	Reports	First Seen	Last Seen	value
Attack,Report	202.113.096.015	392976	135888	2007-3-22	2007-5-1	392976, 135888
Attack/Day	220.099.176.050	385175	674	2007-4-2	2007-4-2	385175
Report/Day	082.194.075.145	72693	67991	2007-5-3	2007-5-3	67292
Attack/Report	070.255.091.204	824761	513	2007-3-12	2007-5-7	1607.721

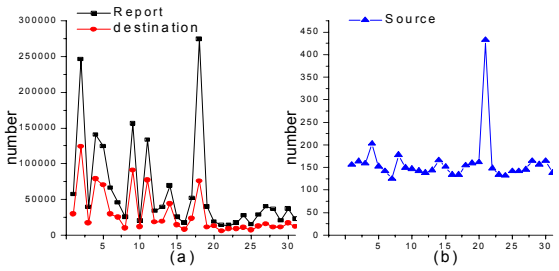

Figure 2. **The cure of Report, destination and source with port 4000**

As to strongly uneven popularity distribution characterization keywords, we check the alert data of port 80 in April, 2007. The detail is showed in figure 3 (a) and (b). Figure (a) shows the number of reports with port 80 and attack destinations port 80, and figure (b) shows the number of attacker source port 80. At April 2007, the reports number is to top peak 434,495, and the attack number is to top peak 87,723 at April 14, but the source port number is to top peak 11631 at April 27. From figure 2 and 3, we can find the cures of report and destination is very similitude, but the cures of report and source or destination and source are not. In fact, we argue that the destination port often can determine the application service, but the source port often randomize. More over, when a global attack occurs, many sensors will detect it and send alerts.

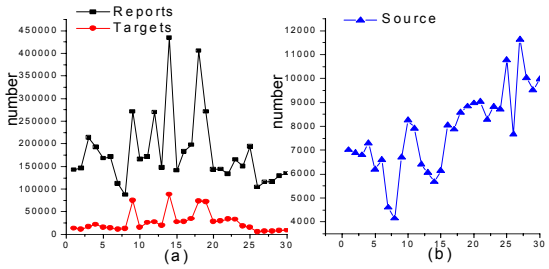

Figure 3. **The cure of report, destination and source**

To evaluation the effect to SFC which fuses alert to the frequent attack, we check the top 10000 source IP attack data by May 10, 2007. The interesting result shows in table 1, the max attack and report event has 392976 and 135888 respectively from March 22, 2007 to May 1, 2007; the max attack/day has 385175 at April 2, 2007; the max report/day has 67292 at May 3, 2007; the max attack/report show average 1607.721 attack in every report. These data show our scheme burden with network size as DShield at top attack event in the rough.

6. Conclusion

In this paper, we describe our on-going research on a P2P IDS alert fusion based on multi keywords routing. All the related alarms produced by local sensor can be evenly routed and fused to its corresponding sensor fusion centers (SFCs) by multi keywords IP_s, IP_d, P_s, P_d, S_{id}. The distributed intrusion alert fusion based on multi keywords scheme can efficiently alleviate disadvantage of the local imprecise classification, and to implement multi points of view fusion. More over our scheme can detect and find new attack earlier than the single view fusing scheme.

7. Acknowledgements

This work is supported by the Natural Science Foundation of Zhejiang Province of China under Grant No. Y106176, and the science and technology search Planned Projects of Zhejiang Provincial No. 2007C33058.

8. References

[1] P. Gross, J. Parekh, and G. Kaiser, "Secure Selecticast for Collaborative Intrusion Detection Systems," presented at the DEBS, Edinburgh, Scotland, UK, 2004.
[2] C. Krugel, T. Toth, and C. Kerer, "Decentralized event correlation for intrusion detection," presented at Proceedings of the ICISC, 2001.
[3] M. E. Locasto, J. J. Parekh, A. D. Keromytis, and S. J. Stolfo, "Towards Collaborative Security and P2P Intrusion Detection," presented at IEEE SMC IAW, West Point, 2005.
[4] Y. Chen, A. Beach, and J. Skicewicz, "Cyber Disease Monitoring with Distributed Hash Tables: A Global Peer-to-Peer Intrusion Detection System," NWU-CS-04-40, 2004.
[5] V. Y. P. B. J. Ullrich, "Internet intrusions: global characteristics and prevalence," presented at the 2003 ACM SIGMETRICS international conference on Measurement and modeling of computer systems, San Diego, CA, 2003.
[6] B. P. Yegneswaran V, Jha S, "Global Intrusion Detection in the DOMINO Overlay System," presented at NDSS, San Diego, CA, 2004.
[7] S. INSTITUTE, "Dshield.org: Distributed intrusion detection system," 2007.