A Multiple Keyword Fusion Scheme for P2P IDS Alert

Ming Xu, Chaochi Lin, Qin Chen
Institute of Computer Application Technology,
HangZhou Dianzi University, P.R.China
E-Mail: mxu@hdu.edu.cn, lincz048@stu.hdu.edu.cn

Abstract

Alert fusion is a key problem in distributed intrusion detection system (DIDS). The paper proposes a distributed intrusion alert fusion scheme based on multiple keywords and routing infrastructure: distributed hash table (DHT). All the related alerts produced by local sensor can be routed and fused to their corresponding peers by multiple keywords, while evenly distributing unrelated alerts to different peer. We evaluation our scheme with a real-world intrusion detection dataset (DShield Dataset), which has been collected firewall and NIDS logs from over 1600 administrators across the world. Experimental results show that our scheme has well scalable, and can achieve significant improvement in load balancing.

1. Introduction

During the recent years, large-scale coordinated attack, such as port scans, worms and distributed denial-of-service (DDoS) attacks, pose a major threat to network infrastructure security. For instance, over 359,000 computers were estimated to be affected by the "Code-Red V2" worm in less than 14 hours in 2001, with the cost of more than $2 billion [1].Since these types of attacks often occur in multiple networks simultaneously, the attack activities may initially appear innocent at each local network, as the evidence for an attack may be distributed across a large numbers of networks. The current state of the art intrusion detection research uses combination of network-based intrusion detection systems (NIDS) and host-based intrusion detection systems (HIDS) to protect systems from compromise. However, these systems generally are deployed in a limited area and suffer from a high false alarm due to the lack of a global view of the intrusion activity. To address this problem, DIDS have been proposed in the literature. Distributed alert fusion is a major issue in the DIDS. Whereas, current DIDS mainly deal with distributed audit collecting, and have a central coordinator or static hierarchical architecture. Those distributed alert fusion systems generally have unscalable communication mechanisms. Most previous works presume that the local alert classification and identification are precise, but we argue that the local classification to some alert may be imprecise because local sensor has limited view and detecting methods to determine the causes, patterns of such events. To that end, we propose distributed, scalable, robustness overlay security networks based on distributed hash table to facilitate high-speed intrusion detection and alert fusion, and we use multiple potential usefulness keywords to alleviate disadvantage of the local imprecise classification, and implement multiple point of view fusion. We build the DHT-based P2P alert fusion system on the top of Pastry [2], and evaluate the feasibility of system with DShield dataset [3].

The remainder of this paper is structured as follows. In section 2 we introduce the network architecture in detail. In Section 3 details the multiple keywords routing generating mechanism. Section 4 presents the evaluation test bed along with the evaluation results. Section 5 discusses related work, and we finally conclude the paper and outline our future work in Section6.

2. The DHT-based P2P IDS Network Architecture

The network architecture of the distributed intrusion alert fusion system is a DHT-based P2P overlay. The system consists of multiple heterogeneous IDS sensors nodes. Previous researches have shown that information shared between these networks is an effective way to detect intrusion. We use DHT route mechanism based on multiple keywords to share and fuse alert information (see Figure 1). When an attack is detected by a local sensor, which reports an attack alert, the multiple keywords about the intrusion symptoms are embedded into the DHT dimension so that alarms related to the similitude intrusion (with the same keyword) will be routed to the same peer to be

978-0-7695-3391-9/08 $25.00 © 2008 IEEE
DOI 10.1109/ICINIS.2008.43

fuse, evenly distributing unrelated alert to the different peers of the distributed intrusion alert system.

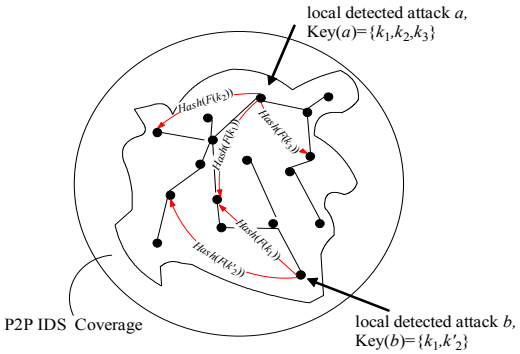

Figure 1.DHT network of P2P IDS

The DHT system provide two basic interfaces: put( ) and get( ). The interface for inserting, put(keyword, object), causes the application-specific objects to be inserted by routing a Pastry message, using the keyword as the key. The Pastry node will effectively route the given object to the node with a nodeId that is numerically closest to the keyword. The interface for retrieval, object=get(keyword), cause the Pastry to obtain the object from a node identifier closest to the keyword. Among all currently live nodes in the alert fusion system, Pastry provides a mechanism for mapping keys to nodes with consistent hashing. The expected number of the forwarding steps in the Pastry overlay network is O(log N), which is the size of the routing table maintained in each Pastry is only O(log N) in size (where N is the number of live Pastry nodes in the overlay network). This implies that the system built on the top of the Pastry can be scalable to a very large network. When a local sensor detects an attack, it generates an alert report that will be routed to the appropriate node with the put(k_i,alert) operation. The keywords set of the alert, $\{k_1,k_2,\ldots,k_n\}$, will route the alert to the appropriate peer in the system. Based on the deterministic characteristic of the keywords routing, the attack which has the same keyword will be routed and reported to the corresponding peer. For an attack alert with a keywords set, every local sensor can query its prevalence with the get(keyword) operation.

3. Multiple Keywords Generating

When cyber intrusions are detected by the local sensor, the alert produced by the local IDS must be reported and routed to appropriate peer perform data fusion and inferences about such an attack. There are two challenges. On the one hand, the local classifying and identifying to some alert may be imprecise because local sensor has limited view and detecting methods, For instance, alert generated by a new known attack or statistics anomaly detection method. So the really meaning of alert may not be comprehensive before global fusing. Moreover, an attack alert could denote different meaning from different point of view. So the multiple keywords, which could potentially 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 may heterogeneous IDSs.

To ensure that related event alert information will be routed to the appropriate peer, and 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 the basic reference to contrast keywords set. I.e. the source and the destination IP address IP_s, IP_d, the source and destination port number P_s, P_d are select. When the attack can be identified or classified accurately, for instance, 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 Sid in sid-msg file can be used as a keyword, because the Snort message ID identifies a known attack, and often gives a corresponding ID of Bugtraq, CVE, and Nessus.

4. Preliminary Implementation and Evaluation

4.1. Prototype system

We implemented a prototype in Java on a Redhat Linux sever, which consists of a Dual Xeon^{TM} 2.4GHz, with 1024 Mbytes memory, 120GB integrated storage. The prototype is used Pastry as its peer-to-peer protocol.

4.2. Experiments Methodology

Firstly, two experiments we conducted to evaluate the system scalability by varying the number of the alerts and peers. Secondly, the effectiveness of the proposed load balancing scheme was evaluated also.

1) Effect of the varying sending alert number: In our first experiment we varied the sending alert number of each peer for a fixed 100 peers. And we use multiple keywords as the route key. The results are shown in Figure 2. The time for the distributed P2P alert fusion system increases from 28s when each peer alert is 100 records to 4min50s on 1000 records. It indicates that the growth of the alert number increases the processing time on each node which received the alert.

2) Effect of varying peer number: In this experiment, we varied the number of peer in the system

from 32 peers to 1024 peers and fixed each peer alert numbers to 100. The results are shown in Figure 3. The time DHT-based P2P distributed alert fusion system to send the alert increases from 28.1s in case of 32 nodes to 1min24s for 1024 nodes. The first reason comes from the route time as the nodes increase. From the figure we can see that as the time increase slightly when the system varied from 32 to 256 nodes, which is decided by the determinate characteristic of the pastry route methods. In the Pastry-based system, nodeIds and keys are thought of as a sequence of digits with based 2^b. A node's routing tables is organized into [$\log_{2^b}$N] rows with 2^b-1 entries each. The default b is set to be 4. So when the peers increase 32 to 256, the route hop number is nearly 1. However, even there are 1 million peers in the system, the average peer route table includes 75 entries and expect route hop is only 5. Secondly, as we simulated all of the peers on the one server, all of the peers run parallel. And as the nodes increase the time increases sharply for the short of process capacity and memory. From above we can clearly see that it is suitable for the full-scale deployment of the P2P distributed alert fusion system.

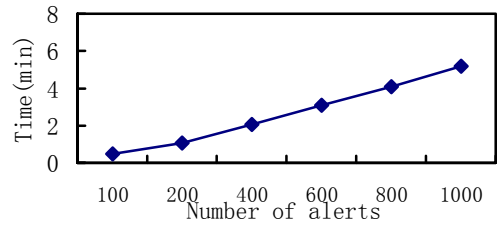

Figure 2.Effect of varying alert number

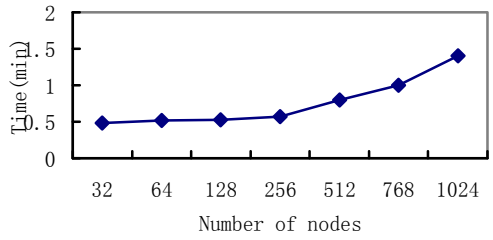

Figure 3.Effect of varying peer number

3) Multiple keywords VS Single keyword: Firstly, we conduct an empirical analysis on dataset from Dshield.org during a seven day from 10-16 December 2004. There is no specific worm outbreak in this time period. There are 294,730,000 records in these logs. We select small partition of the dataset in order to investigate the source behavior during a normal scan scenario where there is no specific worm at that time. We plot the total number of accesses generated by the top source address in every two minutes period across 1 hours from 00:00am to 1:00am on December 15, 2004, which was selected at random from DShield dataset, there are 48,609,000 records in these in that day. As show in Figure 4, during particular time periods, the top 10 sources are responsible for most scans on a given day, around 15% of all accesses with the time period. During the worm outbreaks, there will be more proportion of accesses by the top source. If we only based on the single keyword, such as source address as route keyword in this scenario, it will generate a highly skewed load distribution among the peer that is hosting the alert for the top source address. According our analysis it is strongly uneven popularity distribution of the port [4]. It will result a single peer being severely overload. To the end, we use multiple keywords as route key, we currently use IP_s, P_d and protocol as multiple keywords, we consider that because most coordinated attacks such as worms or coordinated scans share the same IP_s, while the IP_d is specific to a monitored network, and will not appear across multiple subnetworks. We select P_d and protocol for the reason that destination port and protocol often can determinate the application service, but the source port often randomize.

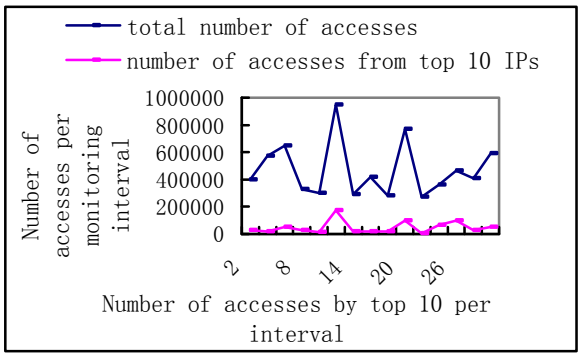

Figure 4.Suspicious IPs in a normal scan scenario

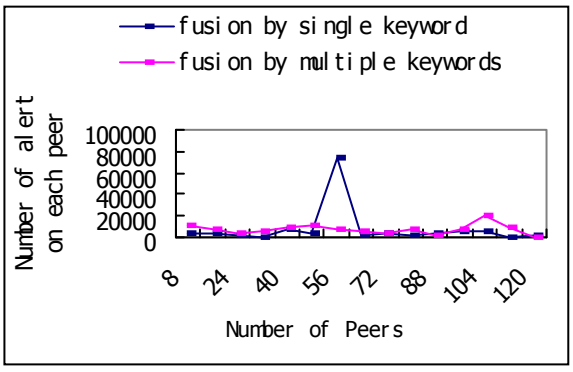

Figure 5.Evaluation of multiple keywords scheme

In the simulation, we applied our proposed scheme to the dataset, and set participating nodes to be 128. Figure 5 plots the distribution of the alert among the peers. From the picture we can see that the there is an

uneven distribution with the single keyword, where a small number of peers are responsible for most of the alerts. Furthermore, with the number of participant peers increase the load of some peers may become even worse. In contrast, our multiple keywords scheme evenly distributed alerts among the peers, i.e., over 70% reduction in peak load. Therefore, our scheme can achieve significant improvement in loading balancing.

5. Relation work

The problem of distributed intrusion detection and alert fusion has become an active research field. We briefly review the most relevant work in the literature comparison to our own approach.

Chen[5] proposed a Global Peer-to-Peer Intrusion Detection System using DHT routing infrastructure, which embedded the attack symptoms into the DHT dimensions so that related to the same intrusion will be routed to the same sensor fusion centers (SFCs), while evenly distributing unrelated alarms to different SFCs. Their scheme resembles ours, but they use single keyword to route, while our approach uses multiple keywords. Min Cai[6] proposed Collaborative Internet Worm Containment NetShield system architecture and trust over a DHT overlay based on Chord. The system is designed to contain unknown worms by using worm signature self generation and dissemination.

In the Publish/subscribe model, which include Indra [7], LarSID [8], MONINO [9], Indra is a distributed scheme of intrusion detection and prevention based on sharing security information between trusted peers in a network to against intrusion attempts. Its system was built upon the Scribe project, which overlays a topic-based publish/subscribe multicast communication mechanism on top of the Pastry P2P network. LarSID provides a service for defending against attacks by sharing secure data between participants from different organizations using a content based peer-to-peer publish/subscribe mechanism, which uses DHT approach to sharing evidence. Though we all use DHT as route mechanism, there are several different differences between LarSID and our proposed solution. LarSID is designed for sharing potential evidence of intrusions between participants. While our approach is mainly focus on how to route the alert effectively. The DOMINO project is a Collaborative Intrusion Detection System (CIDS) aiming to provide a global view of the intrusion scenarios and monitor large-scale outbreak. It is a hybrid of hierarchical and publish/subscribe module. The DOMINO may work well for a small or moderate scale of participants.

6. Conclusion and Future Work

In this paper, we describe our on-going research on a P2P IDS based on multiple keywords and DHT. Our current experimental results show the good load balancing and scalability of the system. The distributed intrusion alert fusion based on multiple keywords scheme can efficiently alleviate the disadvantage of the local imprecise classification, and implement multiple points of view fusion. Moreover our scheme can detect and find new attack or worm earlier than the single view fusing scheme. For future work, we plan to simulate more detailed intrusion scenarios, and consider security issues and trust model of the system. We will consider evaluate our system in a real world further.

7. Acknowledgements

We thank Natural Science Foundation of Zhejiang Province under Grand No.Y106176, and Science and Technology Research Planned Projects of Zhejiang Provincial No.2007C33058 for funding this work.

8. References

[1] D. Moore, C. Shannon, and K. Claffy, “Code Red: A Case Study on the Spread and Victims of an Internet Worm,” ACM SIGCOMM, 2002.
[2] Pastry, http://freepastry.rice.edu.
[3] “Dshield org,” http://www.dshield.org.
[4] Ming XU, Wei Hua, “Distributed Intrusion Alert Fusion Based on Multi Keyword,” In the First International Symposium on Data, Privacy, and E-Commerce, Chengdu, 2007, pp. 469-471.
[5] Zhichun Li, Yan Chen, and Aaron Beach, “Towards Scalable and Robust Distributed Intrusion Alert Fusion with Good Load Balancing,” in Proc. of ACM SIGCOMM Workshop on Large-Scale Attack Defense, 2006.
[6] CAI, M., ET AL, “Collaborative internet worm containment,” in IEEE Security and Privacy Magazine, 2005.ense, 2006.
[7] JANAKIRAMAN, R., ET AL, “Indra: A peer-to-peer approach to network intrusion detection and prevention,” in 12th IEEE WETICE Workshop on Enterprise Security, 2003.
[8] C. V. Zhou, S. Karunasekera and C. Leckie, “Evaluation of a Decentralized Architecture for Large Scale Collaborative Intrusion Detection,” in the Tenth IFIP/IEEE International Symposium on Integrated Network Management (IM), Germany, 2007, pp. 80-89.
[9] V. Yegneswaran, P. Barford and S. Jha, “Global Intrusion Detection in the DOMINO Overlay System,” in proceedings of Network and Distributed Security Symposium, 2004.