Future Generation Computer Systems 33 (2014) 42–52

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

An IP Traceback Protocol using a Compressed Hash Table, a Sinkhole Router and Data Mining based on Network Forensics against Network Attacks

EunHee Jeong [a], ByungKwan Lee [b,*]

[a] Department of Regional Economics, College of Humanity and Social Science, Kangwon National University, Samcheok, Gangwon-do 245-711, Republic of Korea

[b] Department of Computer, College of Engineering, Kwandong University, Gangneung-si, Gangwon-do 210-701, Republic of Korea

HIGHLIGHTS

- A hash table by using CHTM is compressed and its result is stored in DB.
- The traceback against attacks can be done in real time and even after some time.
- Its attack patterns are analyzed by AAM with the attack packets transferred from a Sinkhole Router.
- New attack patterns are extracted by analyzing the collected attack information.
- Its results make the attack filtering of routers strengthened.

ARTICLE INFO

Article history:
Received 14 October 2012
Received in revised form
18 October 2013
Accepted 23 October 2013
Available online 31 October 2013

Keywords:
IP Traceback Protocol
Compressed Hash Table
Sinkhole router
Hash table
Association rule
Attack pattern
Attack packet rule

ABSTRACT

The Source Path Isolation Engine (SPIE) is based on a bloom filter. The SPIE is designed to improve the memory efficiency by storing in a bloom filter the information on packets that are passing through routers, but the bloom filter must be initialized periodically because of its limited memory. Thus, there is a problem that the SPIE cannot trace back the attack packets that passed through the routers earlier. To address this problem, this paper proposes an IP Traceback Protocol (ITP) that uses a Compressed Hash Table, a Sinkhole Router and Data Mining based on network forensics against network attacks. The ITP embeds in routers the Compressed Hash Table Module (CHTM), which compresses the contents of a Hash Table and also stores the result in a database. This protocol can trace an attack back not only in real time using a hash table but also periodically using a Compressed Hash Table (CHT). Moreover, the ITP detects a replay attack by attaching time-stamps to the messages and verifies its integrity by hashing it. This protocol also strengthens the attack packet filtering function of routers for the System Manager to update the attack list in the routers periodically and improves the Attack Detection Rate using the association rule among the attack packets with an Apriori algorithm.

1. Introduction

We live in a society that has a highly-developed internet attack technology. The target of the attacks is expanded to PCs as well as networks. The purpose of these attacks appears extensive, and the extent of the damage caused by them cannot be foreseen. An IP communication in an All-IP environment is done between nodes on a network, and the IP addresses are also allocated not only to the existing computing devices and mobile nodes but also to home appliances in a ubiquitous environment. Therefore, because a number of devices connected to a network in this environment are vulnerable to security threats, an IP traceback technique against network attacks is necessary.

The Source Path Isolation Engine (SPIE) is based on a bloom filter. It is designed to improve the memory efficiency by storing in a bloom filter the information on the packets that are passing through the routers, but the bloom filter must be initialized periodically because of its limited memory. Thus, there is a problem in that the SPIE cannot trace back the attack packets that passed through the routers earlier.

The Probabilistic Packet Marking (PPM) method has the shortcoming that it must receive a minimum number of packets for a victim to reconstruct the attack routes. The ICMP Traceback (iTrace) has the shortcoming that more traffic occurs, which requires additional traceback.

* Corresponding author. Tel.: +82 10 4441 3373.
E-mail addresses: jeongeh@kangwon.ac.kr (E.H. Jeong), bklee@kd.ac.kr, bklee@kwandong.ac.kr (B.K. Lee).

0167-739X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.future.2013.10.023

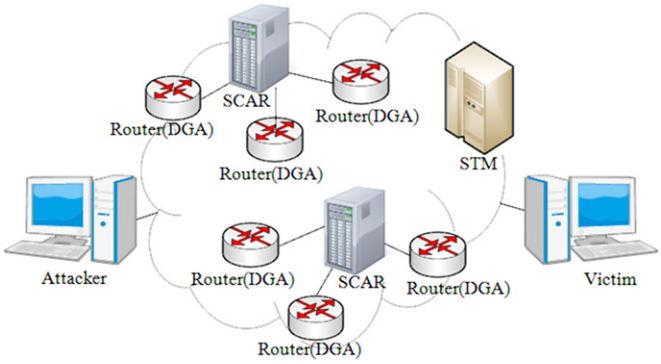

Fig. 1. The SPIE architecture.

To tackle all of these problems, this paper proposes an IP Traceback Protocol (ITP) that uses a Compressed Hash Table, a Sinkhole Router, and Data Mining based on network forensics against network attacks. The ITP embeds in routers the Compressed Hash Table Module (CHTM), which compresses the contents of a hash table and also stores the result in a database. It can trace an attack back not only in real time with a hash table but also by periods with a Compressed Hash Table (CHT). The routing information on traceback against an attack can be provided as forensic evidence for cyber investigation.

Specifically, the Attack Analysis Manager (AAM) in the ITP improves an Attack Detection Rate by generating the attack pattern and the rules of attack packets with an Apriori algorithm.

The remainder of this paper is organized as follows. Section 2 explains the related work and the problems of the existing research. Section 3 explains the ITP, and in Section 4, the ITP is estimated. Section 5 concludes this paper.

2. Traceback approaches and bloom filters

2.1. Traceback approach

2.1.1. Source Path Isolation Engine (SPIE)

The Source Path Isolation Engine (SPIE), which is called the hash-based approach [1,2], is a log-based traceback system that uses efficient auditing techniques at network routers to support the traceback of individual IP packets [3]. In SPIE, routers compute a 32-bit digest of the packets based on the 20-byte IP header and the first 8 bytes of the payload. The routers store these 32-bit packet digests instead of the packets themselves in a space-efficient data structure called a bloom filter [4,5].

Fig. 1 shows the three major architectural components of the SPIE system. Each SPIE-enhanced router has a Data Generation Agent (DGA) associated with it.

The DGA produces packet digests of each packet as it is forwarded through the router and stores them in time-stamped digest tables. The digest tables are stored locally at the DGA for some period of time, depending on the resource constraints of the router. SPIE Collection and Reduction Agents (SCARs) are responsible for a specific region of the network, serving as data concentration points for several routers and facilitating traceback of any packets that traverse the region. When a trace is requested, each SCAR produces an attack graph for its specific region. The attack graphs from each SCAR are grafted together to form a complete attack graph by the SPIE Traceback Manager (STM). The STM controls the whole SPIE system. The STM is the interface to the intrusion detection system or other entity requesting a packet trace. When a request is presented to the STM, it verifies the authenticity of the request, dispatches the request to the appropriate SCARs, gathers the resulting attack graphs, and assembles them into a complete attack graph. Upon completing the traceback process, the STM replies to the intrusion detection system with the final attack graph [5–7].

The SPIE can trace back attack packets only if the STM, SCAR, and DGA are installed in it. In particular, if the DGA is not installed in a router, then the router cannot trace back the attack packets because they have no digest information on the packets. Thus, there is a problem that the SPIE can trace back attack packets only if the DGA is installed in all of the routers. In addition, the DGA stores the information on the packets passing through the routers in its bloom filter. Because the bloom filter is stored in the memory of the routers and the router has limited memory, it must be initialized after a regular time. Thus, the SPIE cannot trace back the attack packets that passed through the routers before that.

This paper solves this problem by compressing and storing the bloom filter and cuts off the DoS attack beforehand by filtering packets in routers and using a Sinkhole router.

2.1.2. Probabilistic Packet Marking (PPM)

The main idea of Probabilistic Packet Marking (PPM) is to allow routers to mark the packets with path information probabilistically and allow the victim to reconstruct the attack path using the marked packets.

The PPM was first suggested by Burch and Cheswick in [8]. The first actual schemes for the PPM were introduced by Savage, Wetherall, Karlin and Anderson in [9], which propose the following clever approach to the IP traceback problems: some fixed number of bits in the packet header are allocated to IP traceback and are used to store an IP address and a hop count. Every router that forwards a packet, independently with some probability p, writes its (unique) IP address to those bits, and sets the hop count to 0. With probability $1 - p$, the IP address is left unchanged, and the hop count is incremented [10,11].

The PPM approach requires the victim to have a prohibitively high number of attacking packets to be collected and a high computational overhead for the traceback procedure, mostly due to the high complexity of coding/decoding of the path information using limited marking space.

2.1.3. iTrace

In the ICMP Traceback mechanism, a new ICMP message type, ICMP Traceback (iTrace), is defined to carry information on routes that an IP packet has taken. The principle idea in this scheme is for every router to sample, with low probability (e.g., 1/20,000), one of the packets it is forwarding and copy the contents into a special ICMP traceback message that includes information about the adjacent routers along the path to the destination. During a flooding-style attack, the victim host can then use these messages to reconstruct a path back to the attacker.

This scheme has many benefits compared to the previous work and is in many ways similar to the packet marking approach that we have taken. However, there are several disadvantages in the current design that complicate its use. These disadvantages include the following: ICMP traffic is increasingly differentiated and could be filtered or rate limited differently from normal traffic, the ICMP Traceback message relies on an input debugging capability (i.e., the ability to associate a packet with the input port and/or MAC address on which it arrived) that is not available in some router architectures; if only some of the routers participate, it appears to be difficult to positively "connect" traceback messages from participating routers separated by a non-participating router; and finally, it requires a key distribution infrastructure to address the problem of attackers sending false ICMP Traceback messages [9].

2.2. Bloom filters

A bloom filter, conceived by Burton Howard Bloom in 1970 [12], is a space-efficient probabilistic data structure that is used to test whether an element is a member of a set.