Contents lists available at ScienceDirect

Information Sciences

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

An improved ant colony system algorithm for solving the IP traceback problem

Ping Wang[a,*], Hui-Tang Lin[b], Tzy-Shiah Wang[b]

[a] *Department of Information Management, Kun Shan University, Taiwan*
[b] *Institute of Computer and Communication Engineering, National Cheng Kung University, Taiwan*

ARTICLE INFO

Article history:
Received 15 August 2013
Revised 10 June 2015
Accepted 4 July 2015
Available online 29 July 2015

Keywords:
IP traceback (IPTBK)
Attack path
Ant colony optimization (ACO)
Waxman model
Ns2

ABSTRACT

The difficulty in identifying the origin of an attack over the Internet is termed the IP traceback (IPTBK) problem. The probable origin of an attack is commonly investigated using some form of ant colony system (ACS) algorithms. However, such algorithms tend to converge to a local suboptimal solution, meaning that the perpetrator of the attack cannot be found. Therefore, the present study proposes a modified ACS scheme (denoted as ACS–IPTBK) that can identify the true attack path even without the entire network routing information. The ability of the ants to search all feasible attack paths was enhanced using a global heuristic mechanism in which the ant colony was partitioned into multiple subgroups, with each subgroup having its own pheromone updating rule. The performance of the ACS–IPTBK algorithm in reconstructing the attack path was investigated through a series of ns2 simulations by using network topologies generated by the Waxman model. The simulations focused specifically on the effects of the ACS model parameters and network characteristics on the performance of the ACS–IPTBK scheme in converging towards the true attack path. Finally, the robustness of the proposed scheme against spoofed IP attacks was investigated. The results showed that the proposed scheme has a slightly slower convergence speed than does the conventional ACS algorithm, but yields a more globally optimal solution for the attack path, particularly in large-scale network topologies.

1. Introduction

According to survey results released by the US Computer Science Institute/Federal Bureau of Investigation [11], the three malware-based threats most experienced by respondents in 2010 were malware infections, bots, and phishing messages. Furthermore, the average financial losses incurred per respondent because of security incidents amounted to US$100 000. Most respondents reported that approximately 20% of these losses were due to the theft of personal information by a malicious third party. Consequently, developing efficient and robust methods for establishing the identity of malicious attackers is increasingly attracting interest.

The difficulty in identifying the origin of an attack over the Internet is termed the IP traceback (IPTBK) problem. Typically, the IPTBK problem involves collecting sufficient routing information to determine all possible paths between the attacker and victim under the constraints of the required number of routing packets and computational time. Solving the IPTBK problem is crucial in information security management for detecting the origin of a malicious attack and bringing the perpetrator to

[*] Corresponding author. Tel.: +886 6 2052139, fax: +886 6 2050545.
E-mail address: pingwang@mail.ksu.edu.tw (P. Wang).

http://dx.doi.org/10.1016/j.ins.2015.07.006
0020-0255/© 2015 Elsevier Inc. All rights reserved.

court. Many methods for reconstructing the attack path have been proposed [1,2,5,6,36,37]. These methods invariably assume the full cooperation of all servers between the victim and the command and control (C&C) server in providing the routing information required to reconstruct the attack path. However, in practice, some service providers may be unwilling (or unable) to provide this information, thus necessitating the reconstruction of the attack path using only partial knowledge of the routing information. Furthermore, the convergence time and the amount of routing information required to reconstruct the attack path must be minimized to ensure a prompt and effective response to perceived or actual attacks. Thus, the optimization scheme used for solving the IPTBK problem must not only be operable with limited routing information but also have low time complexity.

Most existing methods for IPTBK focus on distributed denial of service (DDoS) attacks [1,2,5,6,16,27,35,39]. In such attacks, a malicious user, referred to as a herder, uses a C&C server to initiate an attack on a single target by using a multitude of compromised systems. The resulting flood of incoming messages overwhelms the target and causes shut down, thereby denying access to legitimate users. The compromised systems are referred to as zombies or bots, which collectively form a botnet. Although most herders establish botnets for financial gain or peer recognition, botnets may also be rented to third parties for conducting a range of nefarious activities, including sending spam messages, distributing viruses, and installing spyware.

Given the vast size of most botnets (typically thousands to even millions of nodes), locating the true origin of the attacker is a substantial challenge. The IPTBK problem is practically a combinatorial optimization problem, in which the number of feasible attack paths increases exponentially with the number of C&C servers. Furthermore, it is a nondeterministic polynomial time-complete problem: a set of feasible solutions must be attained within polynomial time under certain constraints (e.g., a limited number of nodes within the topology). Such problems are most commonly solved using heuristic artificial intelligence algorithms, such as ant colony optimization (ACO), genetic algorithms, particle swarm optimization, simulated annealing, tabu search, and evolutionary algorithms. Among these, ACO schemes are particularly effective in solving various combinatorial optimization problems, such as the traveling salesman problem [9,22], vehicle routing problem (VRP) [7,19,21], and network routing problem [17,33].

This study proposes a modified ant colony system (ACS) algorithm, ACS–IPTBK, for solving the IPTBK problem by minimizing the number of routing packets required to reconstruct the path and the time required to converge to the most probable attack path. To implement the algorithm, the ability of the ants to search all feasible attack paths was enhanced using a global heuristic mechanism in which the ant colony was partitioned into multiple subgroups, each with its own pheromone updating rule. The effectiveness of the proposed approach was demonstrated using a series of ns2 simulations conducted using network topologies of various sizes.

The remainder of this paper is organized as follows: Section 2 reviews studies related to the ant system (AS) and describes the use of the AS methodology in solving the IPTBK problem. Section 3 discusses the application of the conventional AS and ACS methods to the IPTBK problem, and describes the proposed enhanced ACS scheme. Section 4 compares the performance of IPTBK in identifying the true attack path with that of the conventional AS and ACS methods. Section 5 examines the effects of the main ACS model parameters on the convergence performance of the proposed scheme. In addition, the effects of the network size on the number of packets required for constructing the attack path and the convergence time of the proposed scheme are investigated and compared with those of the conventional AS and ACS methods. Finally, Section 6 provides brief concluding remarks and indicates the direction of future research.

2. Related work

This section reviews several existing methods for solving the IPTBK problem in DDOS attacks and introduces ACO schemes.

2.1. Existing IP traceback systems for DDOS attacks

A DDoS attack is an attempt to prevent legitimate users of a service from using that service by overwhelming the target server with external communication requests generated by a multitude of coordinated hosts (i.e., zombies). The zombies are compromised by installing malware distributed through their network connexions; once compromised, they are manipulated by a malicious third party (remote controller) through a network of C&C servers. In practice, tracing the remote controller is extremely difficult because their true identity is generally disguised by their constantly changing spoofed IP addresses. Consequently, DDoS attacks severely threaten today's highly connected Internet environment.

In general, the network of compromised hosts (referred to collectively as a botnet) has a centralized architecture, with multiple zombies linked to each C&C. In the event of a DDoS attack (real or perceived), the web defender must identify the C&C servers responsible for controlling the attack to (a) identify the compromised hosts and purge them of malware and (b) trace the attack path to the remote controller. Thus, reconstructing the attack path with the collected routing information is critical in information security management.

Fig. 1 presents a typical DDoS scenario. In theory, the path reconstruction in response to such an attack is a type of graph optimization and can therefore be processed using an ACO scheme. For applying such a scheme, each ant lays a pheromone trail along the route they select between the victim (food source) and attacker (nest) (e.g., path V–R_{43}–R_{41}–R_{21}–R_{13}–R_{11}–S_{11}), and the relative probability of each path being the actual attack path is given by the intensity of the pheromone along the corresponding trail.