Computer Networks 104 (2016) 27–42

Contents lists available at ScienceDirect

Computer Networks

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

A transparent and scalable anomaly-based DoS detection method

Ognjen Joldzic[a,*], Zoran Djuric[a], Pavle Vuletic[b,c]

[a] Faculty of Electrical Engineering, University of Banja Luka, Patre 5, 78000 Banja Luka, Bosnia and Herzegovina
[b] University of Belgrade, School of Electrical Engineering, Bulevar Kralja Aleksandra 73, 11000 Belgrade, Serbia
[c] AMRES - Serbian National Research and Education Network, Bulevar Kralja Aleksandra 90, 11000 Belgrade, Serbia

ARTICLE INFO

Article history:
Received 2 August 2015
Revised 13 January 2016
Accepted 3 May 2016
Available online 4 May 2016

MSC:
68M14
68W05

Keywords:
Intrusion detection
Intrusion prevention
Distributed processing
Load balancing
Security

ABSTRACT

Intrusions and intrusive behaviour can be aimed at different parts of the system, ranging from lower-level network attacks intended to disrupt the flow of data in general, to higher-level attacks targeted against specific applications or services. Due to the constant growth of network traffic and the need to inspect the traffic thoroughly, intrusion detection and prevention are becoming increasingly complex and require significant computational resources. This paper presents a distributed, scalable solution for the detection of lower-level Denial-of-Service (DoS) attacks which are executed by transmitting overwhelming amounts of data with the intention of disrupting regular network service. Scalability is achieved by active traffic balancing among multiple traffic processors, exploiting the flexibility and network programmability that Software Defined Networking paradigm brings and packet processing based on device polling. Traffic processors can be elastically added into the pool depending on the traffic volume. The whole system is completely transparent to the external observers. The paper shows that the implemented balancing algorithm further improves the reliability of the intrusion detection.

1. Introduction

With the constant advances in information and network technologies, the entire landscape of computer networks has changed significantly in recent years. Various new types of client devices and available content and offered services are only few of the factors that contributed to the current state of technology in this field. At the same time, these factors have entailed a number of changes in network design and the set of network technologies needed to support the increased complexity of the environment, new needs and a larger user base. Contemporary networks support simultaneous transfer of various content and services multiplexed over a single physical infrastructure. The number of services supported by this converged network infrastructure is constantly growing as is our reliance on them. Therefore, one of the most important aspects of network implementation which is affected by this converged network model is security [1]. Modern security protocols and techniques have to be robust enough to filter anomalous behaviour from a large volume of very diverse traffic. This paper presents a solution for detecting and mitigating Denial-of-Service (DoS) attacks, which are among the most common types of attacks aimed at the disruption of service operations and reliability. DoS attacks are generated by overwhelming amounts of unwanted network traffic or by forcing computing resources to waste processing and/or storage capacities on unwanted tasks. Detecting DoS attacks requires processing large volumes of traffic and its detailed inspection. This can present a bottleneck and disrupt the other services itself if inserted along the traffic path and if the computing power of the detectors is not sufficient. Therefore, creating a scalable solution for attack detection, capable to be deployed in the most challenging environments is required. Recently, a new network management paradigm, Software Defined Networking (SDN), emerged. It allows flexible and adaptable definitions of the packet forwarding and rewriting rules, and network element programmability from a central controller. These properties enable the design of the attack detection systems that can easily split the load to multiple processors and implement feedback actions that can stop the attack. The solution described in this paper leverages the advantages of multicore processing and a proposed load balancing mechanism, which enable it to overcome the limitations of the standard approach to network threat prevention, and to maintain security without disrupting normal network operation.

The rest of the paper is organized as follows. Section 2 discusses the common characteristics and issues of intrusion detection and prevention. Section 3 shows a detailed view of the current state of technology and research in this field. All parts of the proposed solution and the underlying technologies are described

* Corresponding author.
E-mail addresses: ognjen.joldzic@etfbl.net (O. Joldzic), zoran.djuric@etfbl.net (Z. Djuric), pavle.vuletic@etf.bg.ac.rs (P. Vuletic).

http://dx.doi.org/10.1016/j.comnet.2016.05.004
1389-1286/

in Section 4. This section is divided into 6 subsections, each discussing a distinct segment of the proposed solution: System architecture, software components, communication protocol, load balancing mechanisms and, finally, the detection algorithm and attack mitigation. Section 5 contains an overview of the experiments conducted to prove the efficiency of the platform and the results obtained by those experiments. Finally, at the end of the paper some concluding remarks and outlines of the future work are given.

2. Intrusion detection and prevention - characteristics and common issues

Any action performed by one or more hosts on the network aimed at disrupting the normal operation of the network or obtaining otherwise restricted capabilities or information can be considered as an intrusion. Intrusions and intrusive behaviour can be aimed at different parts of the system, from lower-level network attacks intended to disrupt the flow of data in general, to higher-level attacks targeted against specific applications or services. One of the main issues with intrusion detection and prevention is that each type of attack is specific enough to make it almost impossible to create a general-purpose solution that would provide full protection against all types of known and future attacks [2]. One of the main goals of network attacks is to disrupt regular network activity by exploiting protocol vulnerabilities or by sending a large volume of data (i.e. flooding) or specifically created service requests to overwhelm the receiver's network connectivity or processing resources and cause a denial of service to legitimate users. The attack can be initiated by a single sender, or by a number of compromised hosts (bots) coordinated by the attacker. The latter variant is identified as a Distributed Denial of Service (DDoS) attack.

Although similar in execution and their effects, distributed and non-distributed DoS attacks present an important difference from the standpoint of the security device that is supposed to detect and mitigate the attacks. As the number of attack sources is usually very large, the amount of traffic generated by a single source may often be similar in volume and pattern to regular traffic, and thus difficult to detect. It is only when the attack traffic is aggregated at the entry point of the target network that its destructive potential becomes obvious. As a result, there are no widely accepted procedures for preventing DDoS attacks. Some designs have been proposed in literature that would require an Internet-wide deployment of security nodes and proprietary protocols in order to enable efficient detection and mitigation [3]. These protocols would enable an ISP to remotely enforce security policies based on detected attacks, although the compromised hosts may belong to networks which are administered by different ISPs. The mitigation approach involving inter-ISP communication, although currently implemented in a large number of networks [4], may be challenging to implement for all service providers (especially for smaller providers who have no commercial incentive or resources for implementation of such solutions), which arguably leads to reduced efficiency [5].

Intrusion detection systems (IDS) function as passive monitoring devices that alert the network administrator or another network device in case of suspicious network activity. On the other hand, Intrusion Prevention Systems (IPS) actively participates in the flow of traffic, because the process of intrusion prevention requires them to control the transmission of data.

Both types of security devices can be classified by their approach to traffic analysis or by their threat detection methodology. One of the frequently used traffic analysis techniques is the deep packet inspection (DPI), which provides the widest range of possibilities for detection, because the detection process can be based on any information contained in either the header or the payload of the packet. In recent period, another approach to traffic analysis, called the flow-based detection, has seen a significant development and increase in popularity, despite certain inherent performance issues concerning flow export mechanisms and flow sampling in high speed networks [6]. However, several extensions have been proposed that should increase the performance of flow-based attack detection solutions, especially in the field of DDoS attacks [7].

Depending on the threat detection methodology, intrusion detection can be either signature- or anomaly-based [8]. Signature-based detection algorithms assume that every attack can be, with varying precision, described by a set of rules and packet patterns that are compared to every incoming packet by the scanner in real time. Therefore, the attack signature has to be known before the attack happens in order to correctly configure the matching pattern. In reality, this is often not the case and such strategy can be easily exploited by deducing which patterns are recognized by the detection device, and modifying and executing the attack accordingly.

The other major group of detection algorithms, anomaly-based, relies on the fact that, during the attack, one or more network parameters values will significantly differ from the measured baseline. These algorithms require a scanner to undergo a learning period in order to establish the baseline values. Anomaly-based detection requires less configuration by the administrator and is arguably more robust in terms of discovering attacks that present themselves in previously unknown attack patterns [9].

Regardless of the attack detection approach, the IPS must perform its function without introducing any significant latency, since the service levels of the protected network (the network behind the security appliance) are directly affected by the inspection. IDS devices have a slightly higher tolerance for latency, since they usually receive a copy of the traffic which is forwarded along the original path.

The most common point of deployment for an IPS/IDS within the protected network (i.e. the network which is the attack target) is as close as possible to the probable source of the attack. This deployment increases the chance for a positive detection, since the IPS/IDS device has access to the aggregated traffic that enables it to obtain a complete view of the attack in progress [5]. On the other hand, the placement of the device makes it a possible bottleneck and a single point of failure for high speed networks. Therefore, a logical solution to this issue is the introduction of a scalable device that would distribute the processor load required for packet processing to a number of processors, while still maintaining effectiveness.

The solution proposed in this paper is a transparent intrusion detection system (TIDS), developed using the anomaly-based attack detection methodology. The design goals, operation and structure of the proposed system is explained in detail in Section 4.

3. Background and related work

Contemporary security appliances (built either for intrusion detection or prevention) must operate in near real time in order to be able to efficiently respond to threats involving large volumes of traffic commonly found in today's networks. This section highlights the recent work in the domains of parallel traffic processing and scaling to large volumes of traffic during peak usage periods.

Lakhina et al. [10] discuss the possibilities of detecting various attacks based on the entropy values of certain packet header fields of the incoming traffic. The same approach has been taken by Giotis et al. [11], but with the introduction of SDN to increase the programmability and flexibility of the solution. Extending these findings, the solution proposed in this paper is a highly scalable multi-processor detection system based primarily on entropy calculations.