Information Sciences 278 (2014) 900–913

Contents lists available at ScienceDirect

Information Sciences

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

A seven-dimensional flow analysis to help autonomous network management

Marcos V.O. de Assis [a,*], Joel J.P.C. Rodrigues [b], Mario Lemes Proença Jr. [a,1]

[a] Computer Science Department, State University of Londrina (UEL), Rodovia Celso Garcia Cid, Pr 445 Km 380, Campus Universitário, Londrina, Brazil
[b] Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal

ARTICLE INFO

Article history:
Received 19 July 2013
Received in revised form 24 December 2013
Accepted 8 March 2014
Available online 12 April 2014

Keywords:
DSNSF
Holt–Winters
HWDS
Seven-dimensional flow analysis
Traffic characterization
Network management

ABSTRACT

Due to the increasing need of more agility in information exchange, computer networks are continuously expanding both in magnitude and complexity of the management processes. An essential component of these processes is the anomaly detection and identification. Although there are several studies in this area, simple and efficient anomaly detection mechanisms are still required due to the lack of suitable approaches for large-scale network environments. In this paper, we present an anomaly detection system using a seven-dimensional flow analysis. The core of this system is composed by the Holt–Winters for Digital Signature (HWDS) method, an improvement of the traditional Holt–Winters, which characterizes the traffic of each one of the analyzed dimensions in order to generate profiles able to describe the network's normal behavior, here called Digital Signature of Network Segment using Flow analysis (DSNSF). The low complexity of the presented approach enables fast anomaly detection, mitigating the impact on final users. The system not only warns the network administrator about the problem, but also provides the necessary information to identify and solve it. To measure the efficiency and accuracy of the system, we use real data collected from a large-scale network environment.

1. Introduction

The information era, characterized by the agility, speed and ease of data transmission, is entirely based on computer networks. Due to the convergence of the communication processes to the Internet, these networks are under a continuous growth to support new users and applications. Although the evolution of network physical technologies is able to guarantee this scalability, the management complexity of large-scale networks is still a complex task.

This management is commonly performed manually by network administrators, who analyze graphics and statistical information to seek traces and behavioral patterns in order to detect and identify attacks or failures. This approach is ineffective, laborious and error-prone, which hampers the quality of the provided services. Thus, new management tools able to automatically analyze the network and proactively identify anomalous behavioral patterns are needed. This new management paradigm is known as Autonomic Management [10,16,20], which promotes an increase in network reliability and availability, allowing administrators to quickly identify and solve problems.

* Corresponding author. Tel.: +55 43 99443836.
E-mail addresses: mvoassis@gmail.com (M.V.O. de Assis), joeljr@ieee.org (J.J.P.C. Rodrigues), proenca@uel.br (M.L. Proença Jr.).
[1] Tel.: +55 43 33714534.

http://dx.doi.org/10.1016/j.ins.2014.03.102
0020-0255/

Another network management approach that has been increasingly used in many different tools and models [27,28,30,31] is the analysis of IP flows data. Contrary to the Simple Network Management Protocol (SNMP) approach, this analysis is able to provide a wide range of information through Netflow, IPFix and SFlow collectors, such as the number of bytes and packets, protocols, IP addresses, port numbers, among others [26].

Among the studies and tools that already use flows on network management and anomaly detection, most are based on the analysis of a single flow feature [1,11]. The use of multiple correlated flow features, according to [17], enriches the network management processes, such as anomaly detection and identification, in computer networks. Furthermore, several studies [7,13,32] are based on the network analysis in five minutes time windows. This classical approach is becoming impracticable due to the increase of the transmission rate in large-scale networks. On a 10, 40 or 100 Gigabit Ethernet network, for example, in 5 min, up to 3, 12 and 30 Terabits of data can be compromised, respectively. Thus, more agile approaches, able to detect and identify network problems faster, are required to support the management of large-scale networks.

This paper presents a novel network anomaly detection system based on a seven-dimensional flow analysis, using the following features: bits/s, packets/s, flows/s, IP Addresses of origin and destination and Ports of origin and destination. The main contributions of this system are the simultaneous analysis of seven flow dimensions in one minute time windows, autonomously detecting different anomalies and generating specific alarms. Furthermore, it provides the network administrator with important information, allowing the recognition of the origin, destination and application (port) of the detected anomaly in order to help its identification and the countermeasure process. The presented system is divided into two main modules: the Detection and the Information module.

The Detection module is responsible for the autonomous analysis of the seven mentioned flow dimensions. This analysis includes: data treatment processes, like the transformation of qualitative dimensions into quantitative ones through the use of Shannon Entropy and the data granularity reduction provided by the Exponential Smoothing technique; and anomaly detection through a hybrid approach capable of detecting specific anomaly signatures and unknown abnormal behaviors. To make it possible, the system is based on one of the most important steps towards anomaly detection: traffic characterization [19]. The introduced system characterizes the normal network behavior of each one of the seven analyzed dimensions, generating seven different profiles that represent a normal day. These profiles are called Digital Signature of Network Segment using Flow analysis (DSNSF), and are the core of the presented system. To perform this characterization process (DSNSFs generation), we used an improved version of the statistical forecasting Holt–Winters method [29], named Holt–Winters for Digital Signature (HWDS) [2,8]. The HWDS method was specifically designed to generate good network profiles, providing a low computational cost approach with higher efficiency than the traditional method.

The Information module is responsible for providing relevant information about the detected anomaly to the network administrator through the use of two approaches: the global view, a seven-dimensional view of the network behavior in a specific time interval; and the ranking of TOP users, which provides the IP addresses of origin and destination, ports of origin and destination and protocols with higher occurrence frequency on the analyzed time interval. This information helps the decision-making process, reducing the network administrator reaction time and, consequently, mitigating the impact of the anomaly on final users.

To measure the efficiency and accuracy of the presented system, three test scenarios are proposed. The first one is the analysis of the system's efficiency in a real large-scale network environment, collecting data from the State University of Londrina, a large network composed by more than 7000 different hosts. The second scenario tests its efficiency with specific types of anomalies through the use of simulated attacks over the collected real data. Finally, the third one is the complexity analysis of the system.

The remainder of this paper is composed of the following sections: Section 2 presents the related works which present important concepts of this manuscript. Section 3 explains the HWDS method and the processes of traffic characterization (DSNSF) and confidence bands generation. Section 4 introduces the new anomaly detection system, its structure and its functionalities. Section 5 presents the system's performance evaluation through three different scenarios. Finally, Section 6 presents the conclusions of the paper and future work projects.

2. Related works

In recent years, the use of flow analysis on network management approaches and tools has increased significantly due to its large number of different available features, providing detailed information about all the network's communication processes. In [6], the authors use the flow analysis on concurrent systems to detect unreachable states and actions. In [12], the authors present a bandwidth management system for the QoS-assured network architecture through the analysis of IP flows, analyzing its capability to estimate the bandwidth usage, the performance of its utilization and the scalability to analyze hundreds of IP flows. In [17], the authors use three flow dimensions or features on the traffic characterization process of the analyzed network simultaneously, proving the existence of a correlation between them.

Among the different approaches used on the anomaly detection process of modern network management, traffic characterization has been widely applied due to its high efficiency and ability to identify unknown anomalies. In [15], the authors propose and present a hybrid traffic forecasting method through the use of Covariation Orthogonal and Artificial Neural Networks. In [18], the author proposes an optimized BiLinear Recurrent Neural Network (BLRNN) to predict the traffic of