Journal of Network and Computer Applications 40 (2014) 292–306

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

A transform domain-based anomaly detection approach to network-wide traffic

Dingde Jiang [a,c,*], Zhengzheng Xu [a,b], Peng Zhang [a], Ting Zhu [d]

[a] College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
[b] School of Business Administration, Northeastern University, Shenyang 110819, China
[c] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
[d] Department of Computer Science, State University of New York, Binghamton, NY 13905, USA

ARTICLE INFO

Article history:
Received 4 December 2012
Received in revised form
17 June 2013
Accepted 24 September 2013
Available online 31 October 2013

Keywords:
Network-wide traffic
Anomaly detection
Transform-domain analysis
Feature extraction
Origin–destination flows

ABSTRACT

Traffic anomalies contain existing abnormal changes in network traffic, which are derived from malicious and anomalous behaviors of users or network devices, such as network faults, abuses, network attacks, etc. These anomalies often damage our operation networks and even lead to network disruptions. In the present paper, we propose a novel method for detecting traffic anomalies in a network by exacting and capturing their features in the transform domain. Here, we take in consideration network topology information and network-wide traffic jointly. We find that anomalous network-wide traffic usually exhibits distinct high-frequency nature. This motivates us to utilize transform domain analysis theory to characterize network-wide traffic to identify its abnormal components. Besides, we group all origin–destination flows in the network in accordance with common destination nodes. By combining network topology information and transform-domain analysis in the given time window, the specious traffic components can be found and identified. Simulation results show that our detection algorithm exhibits a fairly robust detection ability and provides the better detection performance than previous algorithms.

1. Introduction

Traffic anomalies exhibit the abnormal network behaviors, which there are many causes that result (Lakhina et al., 2005; Jiang et al., 2011a). The abnormal behaviors can often give rise to the dramatic change of network traffic, including link traffic and origin–destination flow traffic (Nie et al., 2013; Guo et al., 2006; Raghunath et al., 2007). Network traffic yields anomaly changes when networks abnormally operate (Xiang et al., 2011). All these behaviors have the impact on network activities. Hence, detecting traffic anomalies in the network is an important task for network management and network operation (Lakhina et al., 2004; Federico et al., 2011; Shanbhag and Wolf, 2009; Venkatarama et al., 2010; Kanda et al., 2010).

However, performing accurate detection of traffic anomaly is a challenge. Network traffic is difficult to handle because it contains many inherent properties (Akgül et al., 2011; Vishwanath and Vahdat, 2009; Kodialam et al., 2009; Lazarou et al., 2009). Its larger changes often result in the faults and congestions of networks. Traffic estimation is helpful to capture network traffic nature (Anand et al., 2009; Guan et al., 2010; Qin et al., 2011; Lu and Ghorbani, 2009; Jiang and Hu, 2009; Zhu and Yu, 2006). Network traffic anomalies reflect the anomalous or malicious behaviors that appear in the network. Discovering exact/actual network traffic anomalies is to effectively contain these anomalous or malicious behaviors that damage the network (Yu et al., 2010; Thatte et al., 2011; Nawata et al., 2010; Vishwanath et al., 2011; Celenk et al., 2010; Zhu et al., 2011). Anomalous traffic is often much smaller in volume in contrast to the normal ground traffic, and is buried in the huge ground traffic (Jiang et al., 2011b; Guo, 2007). Therefore, this makes it hidden and indiscoverable. On the other hand, some of the anomalous traffic also has burst characteristics and distributed properties. All of the above characteristics increase the difficulties in detecting anomalous network traffic (Casas et al., 2010; Rubinstein et al., 2009; Silveira and Diot, 2010).

To overcome these problems, many methods are proposed for detecting anomalous traffic in the network. Wiese and Hero (2009) presented a decomposable principle component analysis (PCA) method to discover anomalous behaviors in the network. Paschalidis and Smaragdakis (2009) exploited the spatio-temporal correlations to detect network-wide anomalies based on the empirical measures. Xu et al. (2005, 2008) analyzed network traffic behaviors and characterized network traffic. However, owing to the diversity of traffic anomalies, different anomalies have different behavior characteristics, and thus need different methods to analyze them. Lakhina et al. (2005) exploited traffic feature distributions to extract network anomalies. Additionally, they also adopted the PCA method to detect and diagnose network-wide traffic anomalies (Lakhina et al., 2004). Huang

* Corresponding author at: College of Information Science and Engineering, Northeastern University, Shenyang 110819, China. Tel.: +86 24 83684219.
E-mail address: jiangdingde@ise.neu.edu.cn (D. Jiang).

1084-8045/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jnca.2013.09.014

et al. (2007) used a network-wide analysis method to diagnose network disruptions. Singhal and Michailidis (2008) investigated optimal sampling technologies in network detection and monitoring. Kim and Reddy (2008) studied statistical methods that used packet header data to detect traffic anomalies. Nychis et al. (2008) discussed the problem of the empirical evaluation that used entropy techniques to detect traffic anomalies.

Different from the above methods, this paper investigates the network-wide traffic detection based on the transform domain analysis. Firstly, we here take into account the network topology information and network-wide traffic in order to accurately detect the abnormal traffic. The end-to-end network traffic is not apparently definite as link-level traffic. As mentioned above, it holds many hidden inherent features that are difficult to discover. Fortunately, network-wide traffic gives us a network-level point of view to formulate and grasp its inherent and hidden properties. Moreover, this network-wide traffic also describes the network activities and user behaviors from the network-level perspective. Secondly, we classify origin–destination flows in the network into different groups in accordance with common destination addresses. Origin–destination flows arriving at common destination addresses hold some relative characteristics. This classification method is therefore reasonable. Thirdly, we extract the inherent properties in network-wide traffic by introducing the transform domain analysis. As a time sequence, network-wide traffic holds the time and frequency properties. Through the transform domain analysis, we find that the anomalous network traffic exhibits the distinct high-frequency features in the transform domain. Furthermore, we also observe that the transform domain analysis in the different time window has some impact on characterizing the inherent nature of network-wide traffic. Accordingly, this motivates us to utilize/employ the time window to analyze network-wide traffic. This time window analysis is useful to decrease the computation overhead and realize the detection. By grouping all the origin–destination flows, we make the transform domain analysis for network traffic of the origin–destination flows in the different group. Then the hidden properties are extracted from the transform domain. We can thus detect accurately the specious and abnormal traffic. Finally, we exploit network traffic data from the real network as background traffic to validate our approach. Simulation results show that our method can well detect and identify the anomalous network traffic, is robust to different time windows, and exhibits better detection performance than previous approaches.

The rest of this paper is organized as follows: Section 2 presents the related works; Section 3 introduces the network-wide traffic and system model; Section 4 derives our anomaly detection method; Section 5 presents the simulation results and analysis, evaluates detection results, detection ability, impact of window size, and detection performance of our approach; finally, we conclude our work in Section 6.

2. Related work

The network-wide traffic holds diverse inherent properties and is significantly difficult to predict and estimate (Vishwanath and Vahdat, 2009; Jiang et al., 2010). The entropy method has extensively been used for detecting traffic anomalies. Nychis et al. (2008) studied the impact of the entropy analysis of multiple traffic distributions on traffic anomaly detection. Lakhina et al. (2005) used entropy to analyze the traffic feature distributions and capture anomalies. Xiang et al. (2011) introduced the generalized entropy metric and the information distance metric to identify low-rate DDoS attacks. However, these methods are difficult in the capture of network-wide nature.

Statistic analysis is a useful method for anomalous identification in network traffic. Thatte et al. (2011) used aggregate traffic statistics to find network anomalies. Federico et al. (2011) used a α-stable first-order model and a generalized likelihood ratio test to identify network traffic anomalies. PCA can extract main characteristics in network traffic. Lakhna et al. (2004) exploits the PCA to detect and diagnose the abnormal traffic. They analyzed the network-wide anomalous traffic. Rubinstein et al. (2009) analyzed poisoning and defense techniques for the PCA-subspace detector in backbone networks. Wiese and Hero (2009) propose the decomposable PCA method, which can detect the specious traffic only by the lower computation overhead. In the present paper, we will also detect network-wide traffic anomalies as in Lakhina et al. (2004), while our method is completely different from it.

The signal processing is often exploited to identify network traffic anomalies. Celenk et al. (2010) used the adaptive Wiener fil-tering process and auto-regressive moving average model to capture network feature changes. Casas et al. (2010) relied on the coarse-grained measurement information to detect and isolate abnormal traffic. Silveira and Diot (2010) used the unsupervised root cause analysis to diagnose anomalous traffic. Barford et al. (2002) proposed a deviation score anomaly detection (DSAD) approach to find the specious part in network-wide traffic. Maggi et al. (2009) used the fuzzy alert aggregation to improve the detection performance of anomaly detectors. Brauckhoff et al. (2010) analyzed in detail the random packet sampling and the impact of its quantification on anomaly detection. Chhabra et al. (2008) presented a spatial anomaly detection (SPAD) method. Here, we also exploit the signal analysis as in Barford et al. (2002) and Chhabra et al. (2008) to detect anomalous components of network-wide traffic, but our approach adopts a new detection method.

Alternatively, Yu et al. (2010) studied the modeling, analysis, and countermeasures for worm attacks on communication networks. Qin et al. (2011) exploited the bind source separation to find the abnormal traffic components in the network. Guan et al. (2010) studied the detection and measurement method for the dynamic changes in the critical traffic patterns. Lu and Ghorbani (2009) studied using the wavelet analysis to detection network anomalies. Eriksson et al. (2010) proposed a model and frame-work to find the abnormal network events. Our work is mainly related to those in Lakhina et al. (2004), Barford et al. (2002) and Chhabra et al. (2008). We will thoroughly compare our method with them in the following section.

3. Problem statement and system model

Traffic anomalies are the deviation of network traffic from normal traffic, which are often found in our networks. There are many causes that result in traffic anomalies. Network abuse, for example, which usually yields large burst traffic, is one of the main causes. Network flash crowd also produces the large traffic change and leads to anomalous traffic patterns. Network faults will lead to network rerouting and consequently give rise to the dramatic change of network traffic, including link traffic and origin–destination flow traffic. Network attacks, such as distributed denial of service (DDoS), often lead to network-wide traffic changes. Worm propagation can also result in network traffic anomaly changes. More importantly, all these behaviors that encourage anomaly traffic will somehow/eventually affect the normal activity of network operation, more or less. Furthermore, some of them are able to can lead to network disruptions.

3.1. Network-wide traffic and anomaly detection

For large-scale networks, traffic flows from origin nodes to destination nodes will transverse many intermediate nodes. These traffic flows are aggregated into the volumes of the traffic of links