TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 04/13 pp500-509
Volume 21, Number 5, October 2016

PCA-Based Network Traffic Anomaly Detection

Meimei Ding and Hui Tian*

Abstract: The use of a Traffic Matrix (TM) to describe the characteristics of a global network has attracted significant interest in network performance research. Due to the high dimensionality and sparsity of network traffic, Principal Component Analysis (PCA) has been successfully applied to TM analysis. PCA is one of the most common methods used in analysis of high-dimensional objects. This paper shows how to apply PCA to TM analysis and anomaly detection. The experiment results demonstrate that the PCA-based method can detect anomalies for both single and multiple nodes with high accuracy and efficiency.

Key words: traffic matrix; network performance; principal component analysis; anomaly detection

1 Introduction

With the continuous expansion of network scale and the rapid development of network applications, network security is becoming more and more important. Therefore, network anomaly detection has become an important research topic. Network anomalies include malicious attacks, node or link disconnection, and more. Existing work has shown that any anomaly will cause an abnormal change in traffic volume. Thus, it is feasible to monitor traffic volume changes in a network to detect network anomalies.

However, most existing detection methods, such as wavelet-based and exponential-smoothing-technique-based methods, aim at anomaly detection on a single link[1] or a network terminal. They regard traffic as a one-dimensional signal in a temporal domain. But in practice, many traffic volume anomalies at the link traffic level may occur at one or more links. They are often overwhelmed within normal traffic patterns, caused by the high level of traffic volume aggregation on backbone links. Therefore, it is quite hard to discover anomalies at the link level. If an anomaly is detected in the network terminal, instantaneity cannot be guaranteed. Thus, this paper proposes a PCA-based method to detect anomalies within overall network traffic. The method deals with a two-dimensional Traffic Matrix (TM) of the whole network and detects anomalies effectively and efficiently.

A TM consists of all Origin-Destination (OD) flows in the network. An OD flow is a collection of traffic volume which flows from a node (called the entry point) into the network, out of the network from another node (called the exit point). The OD flow goes through the path determined by the routing matrix. The superposition of the OD flows constitutes the traffic observed on backbone links. Therefore, it is effective to study OD flows in order to find out the traffic characteristics of the network. That is, the characteristics of a TM imply the characteristics of the link-level traffic.

The TM of a network with n nodes contains n^2 entries. So a medium-scale network with tens of nodes may include hundreds or thousands of OD flows. As the TM changes frequently as time advances, researchers have explored many traffic analysis methods. With the analysis results from these methods, various applications in network anomaly detection, fault diagnosis, network resource optimization, and related areas were developed. A Diffusion Wavelet (DW) approach was proposed for TM analysis by Willinger et al.[2] But the complexity of this method is high. Moreover, it can be applied only to one traffic matrix

• Meimei Ding and Hui Tian are with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China. E-mail: htian@bjtu.edu.cn.

∗ To whom correspondence should be addressed.

Manuscript received: 2016-06-28; revised: 2016-08-04; accepted: 2016-08-30

at a time, which means that the computation required for all the TMs in a network will be excessive. In Ref. [3], a gravity model is applied to the TM. But the model itself can not accurately represent the original TM. Graph-wavelet-based Multi-Resolution Analysis (MRA) was first applied to TM analysis by Crovella and Kolaczyk[4]. The method does not provide a sparse model of the TM but rather an over-complicated decomposition.

In this paper, we apply a PCA-based method to analyze TMs. The least number of principal components is extracted to reduce the dimension of the data space. The extracted principal components can contain the maximum features of the original data and lose as little information as possible. TMs can be analyzed effectively by PCA because the TM is sparse; a lot of OD flows are zero or very small compared with other OD flows. The characteristics of a sparse model for the original data space can be represented perfectly by a small number of principal components. PCA-based analysis results can be applied to traffic inference, flow prediction, anomaly detection, and related areas.

PCA is an algorithm for dimensionality reduction and multivariate analysis. It was first applied in data compression, image processing, neural networks, data mining, and pattern recognition. The widespread use of PCA is mainly due to its three significant characteristics. First, after high-dimensional data is compressed into a set of low-dimensional data, the mean square error of the reconstructed data is inversely proportional to the dimension. Second, the model is stable without adjusting parameters in the process. Third, for given parameters, compression and decompression are easy to conduct.

The remainder of the paper is organized as follows. In Section 2, related work is briefly described. Section 3 introduces the process of PCA analysis on TMs. In Section 4, we propose an approach to diagnosing single-node anomalies by selecting two significant parameters. In Section 5, we extend the approach to detect multi-node anomalies by improving selected parameters. The comparison with existing methods is given in Section 6. The paper is concluded in Section 7.

2 Related Work

The application of the PCA technique to TMs is similar to that of Compressed Sensing (CS) applied in many fields. The sparsification technique indicates that any sufficiently compressible data can be accurately reconstructed from a small quantity of non-adaptive, randomized linear projection samples[5]. In Ref. [6], the CS method is proposed for inferring network characteristic parameters such as end-to-end delays and bit-error rates, but not for other important tasks such as anomaly detection and localization.

Barford et al.[7] proposed an accurate and efficient link-level anomaly detection approach using a wavelet filter in IP flow and SNMP data. Anomalies are detected by observing a sharp increase of the local variance of the filtered traffic in high-frequency and medium-frequency bands, which can detect volume anomalies even with abundant background traffic. But the method's weakness is that the approach has many tunable metrics and can perform poorly if the metrics are set inappropriately. Another scheme for anomaly detection uses probabilistic or analytical-model-based methods. An alarm is sent as soon as a deviation from the model is found[8]. In Ref. [9], a typical Bayes-based approach is applied to forecast disk drive failures. In Ref. [10], a specific analytical model is presented for effectively diagnosing anomalies within I/O systems. These works mostly focus on one-dimensional data in the temporal domain. However, it may be impractical to model and analyze all link data simultaneously, because even single-link models employing this approach were complex.

The approach proposed in this paper targets anomaly detection with a global view of network traffic, represented by its TM. First, TMs are analyzed by PCA, and then several significant metrics are studied. Finally, an effective method to detect both single-node and multi-node anomalies is put forward for two cases: node disconnection and Distributed Denial of Service (DDoS) attacks.

Similar to our approach, Diffusion Wavelet (DW) applied to anomaly detection in Refs. [11, 12] also deals with TMs of the whole network. DW-based analysis is combined with matrix energy to detect and locate anomalies in their work. But they did not consider complicated situations and mainly developed the applications for physical locating of the concerned anomaly. In contrast, our paper considers both single-node and multi-node situations. We compare detection results of the PCA-based approach with those of DW-based and other approaches, and show that it achieves a higher detection accuracy.

3 Principal Component Analysis

A general backbone network is composed of many nodes, which are connected by links, called Points of Presence (PoPs). The backbone network of America, Abilene, is presented in Fig. 1. In this context, the term anomaly is specifically used to refer to a sharp fluctuation in OD flow, representing traffic experiencing a positive or negative change[13], where both legal and illegal behaviors[14] are involved.

Since the backbone network has 12 nodes, its traffic matrix includes 12 by 12 entries. It's difficult to directly supervise each OD flow of the traffic matrix to judge whether the network is experiencing an anomaly. Thus, OD flows in a traffic matrix during a time interval are reorganized as column vectors. Combining these column vectors defined as the OD flow vectors of succeeding sampling instants together constitutes a high-dimensional OD flow matrix. We apply PCA to this matrix.

PCA is a coordinate transformation scheme[15] that maps a given high-dimensional set of samples onto new axes, called principal axes or principal components. Before applying PCA, a normalization procedure is required to process the OD flow matrix. Notice that different OD flows may have different scales. Since we consider all the OD flows to be equally important, we need to transform the traffic volume to a uniform scale. Then the data must be adjusted to a zero mean.

The principal components have the following features. The first principal component lies in the direction of maximum variance of the samples. The second principal component corresponds to the direction of maximum variance in the remaining data, except for the variance represented by the first component. The other principal components obtain the maximum variance within the remaining data. All these principal components are orthogonal. Thus, the principal axes are sorted by the amount of data variance that they capture, in descending order.

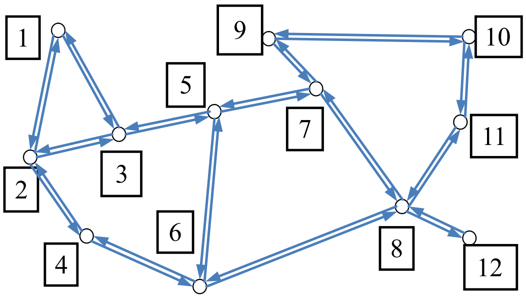


Fig. 1 The Abilene network.

Applying PCA to the normalized OD flow matrix $\boldsymbol{Y}$ in IR^n generates a gathering of n principal components, $\{\boldsymbol{v}_i\}_{i=1}^n$, represented also by $\boldsymbol{v}$ as follows. According to Ref. [15], the first principal component $\boldsymbol{v}_1$ is the vector that corresponds to the direction of maximum variance of data, and is denoted by

$$\boldsymbol{v}_1 = \arg_{\|\boldsymbol{v}\|=1} \max \|\boldsymbol{Yv}\| \tag{1}$$

where $\|\boldsymbol{v}\|$ is the 2-norm of $\boldsymbol{v}$, and $\|\boldsymbol{Yv}\|$ is proportional to the variance of the data distributed along $\boldsymbol{v}$. Proceeding iteratively, if the previous $m-1$ principal axes have been selected, the residual is the difference between the original samples and the samples corresponding to the previous $m-1$ principal axes. Therefore, according to Ref. [15], the m-th principal component is defined as

$$\|\boldsymbol{v}_m\| = \arg_{\|\boldsymbol{v}\|=1} \max \left\| \left(\boldsymbol{Y} - \sum_{i=1}^{m-1} \boldsymbol{Y}\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}} \right) \boldsymbol{v} \right\| \tag{2}$$

A significant use of PCA is to investigate the intrinsic dimensionality[16] of OD flows. By computing the variance obtained by each principal component, it is discovered that the variance along the previous t dimensions is non-negligible. Thus the data denoted by $\boldsymbol{Y}$ can be effectively represented by a t-dimensional $(t < n)$ subspace. In fact, the low effective dimensionality of OD flows forms the basis of our proposed detection method.

4 Single-Node Anomaly Detection

A traffic anomaly behaves as an extraordinary traffic volume load level caused by a variety of reasons, such as worms, network equipment failure, DDoS attacks[17, 18], flash crowd, routing table changes, etc. It is vital for network management to maintain normal network operation by detecting anomalies effectively and taking action rapidly. Network anomalies fall into two major categories: disconnected nodes and hostile anomalous behaviors, such as DDoS. We target these two classes and study how to detect them.

In this section, two significant metrics are described in Section 4.1. In Section 4.2, an anomaly detection method that involves estimating the scale of the normal patterns is proposed. In Section 4.3, performance evaluation of our approach is described.

The datasets used in our experiments are from 2003–2004's open data of the Abilene network, as there is no accessible data source for more recent years. Though the data source used is old, the PCA-based method shall show similar performance on today's data, due to the common sparsity of these datasets. The interval of the obtained TM in Abilene is 5 minutes[2]. In our experiments, we apply PCA to 288 traffic matrices samples in a whole day.

4.1 Selection of parameters

In this subsection, two parameters are selected: dissimilarity, $\bar{d}$, and anomaly score, R. $\bar{d}$ represents the mean dissimilarity between the sample to be detected and other samples. R represents the degree to which the projection of the sample to the first principal axis deviates from the mean state. An approach is presented to detect anomalies based on these two parameters.

4.1.1 Dissimilarity

In our experiments, 12 continous samples are selected per group during a given time window. When the window is too long, large fluctuations caused by the dynamics of network traffic is regarded as abnormal. If the window is too short, PCA is applied more frequently, resulting in massive time overhead. So an hour is selected as the time window, i.e., 12 samples are included per experimental dataset. Each traffic matrix at every snapshot i is recognized and into a long column vector $\boldsymbol{x}'_i = [x_{1,i}, x_{2,i}, \ldots, x_{N,i}]^{\mathrm{T}}$, where $i = 1, 2, \ldots, 12$, $N = 144$. N is the number of OD flows in the Abilene network. Next, 12 column vectors $\boldsymbol{x}'_1 - \boldsymbol{x}'_{12}$ are assembled into a single large matrix $\boldsymbol{X} = [\boldsymbol{x}'_1, \boldsymbol{x}'_2, \ldots, \boldsymbol{x}'_{12}]$, including all the OD flows within the same hour. Each row j in $\boldsymbol{X}$ presents the OD flows between certain node pairs at 12 different sampling times, and each column i presents an instance of all 144 OD flows at the i-th snapshot.

Before applying PCA, we need to process the data by normalization to form the matrix $\boldsymbol{Y}$. All 144 principal components can be obtained after applying PCA to $\boldsymbol{Y}$. Figure 2a represents the plot of variance obtained by each principal component. The variance is calculated using the eigenvalues of the covariance matrix of $\boldsymbol{Y}$.

$$\boldsymbol{C} = \frac{1}{n}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \tag{3}$$

This figure indicates that even though the Abilene network has 144 OD flows, most of the variance in each OD flow can be obtained by the first several principal axes, which conforms to the low effective dimensionality of OD flows.

The projection of the data onto each principal component is denoted by

$$\boldsymbol{u}_i = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{Y}, \quad i = 1, 2, \ldots, 144 \tag{4}$$

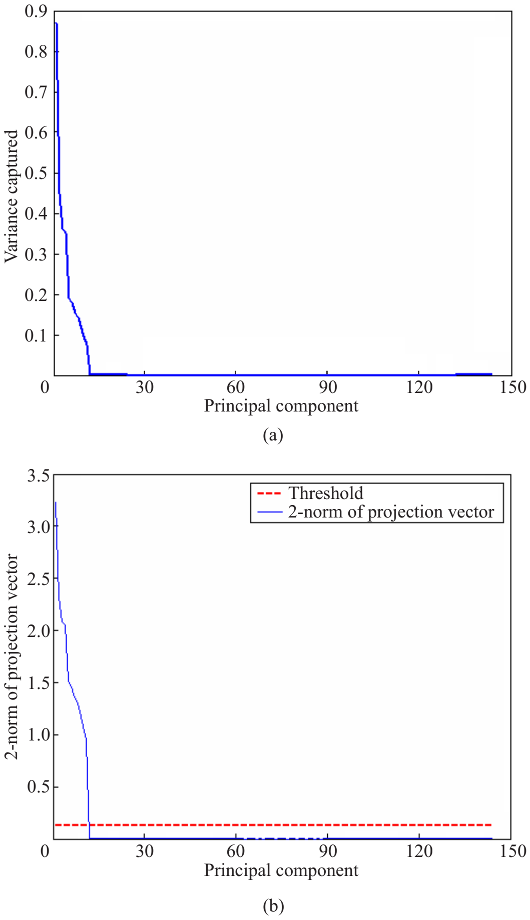


Fig. 2 (a) Variance captured by each principal component; (b) 2-norm of projection vector onto each principal component.

The solid curve in Fig. 2b shows the 2-norm of the projection vector $\boldsymbol{u}_i$ of the data onto each principal axis, $\|\boldsymbol{u}_i\|$. All the principal components are separated into a normal set and an abnormal set by a threshold-based mehtod[19]. $\|\boldsymbol{u}_i\|$ is examined successively for i from 1 to 144. If $\|\boldsymbol{u}_{K+1}\|$ passes the threshold, the first K principal components are regarded as normal principal components, with the others belonging to the abnormal principal components. The empirical threshold is employed (i.e., mean value of all the 2-norm projection vectors) as our threshold as shown by the dashed line in Fig. 2b. According to the threshold, the first 11 principal components are classified in the normal set denoted by $\boldsymbol{V}$, and the others in the abnormal set denoted by $\boldsymbol{V}'$. The normal subspace and abnormal subspace can be obtained by mapping the data onto $\boldsymbol{V}$ and $\boldsymbol{V}'$, denoted by $\boldsymbol{S}$ and $\boldsymbol{S}'$, respectively:

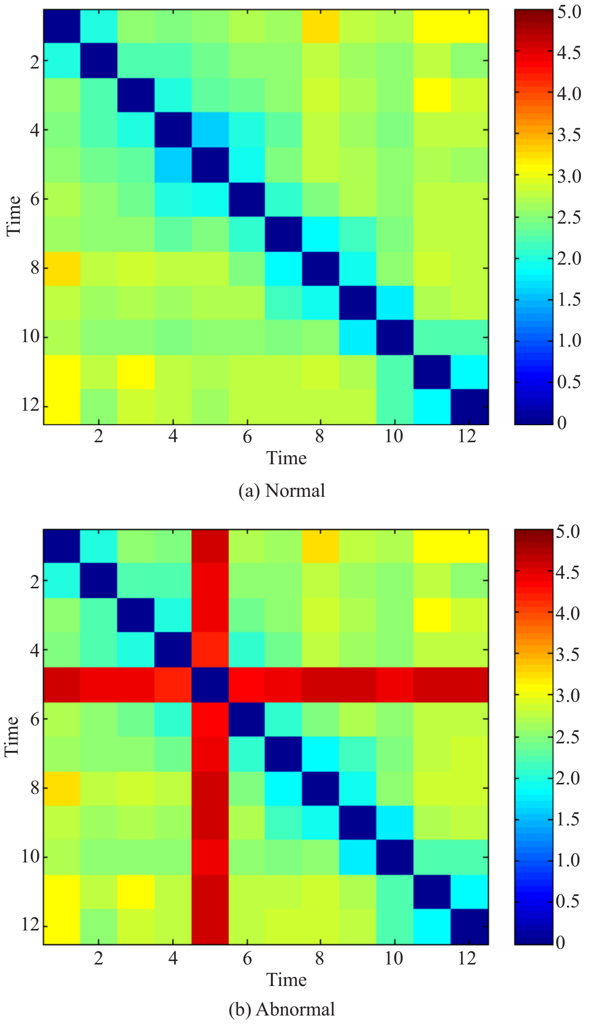
$$\boldsymbol{S} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{Y}, \quad \boldsymbol{S}' = \boldsymbol{V}'^{\mathrm{T}}\boldsymbol{Y} \tag{5}$$

$\boldsymbol{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_{12}]$, where $\boldsymbol{s}_i$ is an 11-dimensional vector, representing the projection of OD flows onto $\boldsymbol{V}$ at time i. Therefore, the underlying normal traffic remains in $\boldsymbol{S}$. Euclidean distance is employed to express the dissimilarity[20] between sample n_x at time x and sample n_y at time y:

$$d(n_x, n_y) = \sqrt{\sum_{i=1}^{K} \left(n_{x,i} - n_{y,i}\right)^2} \tag{6}$$

where $K = 11$ is the dimension of data space $\boldsymbol{S}$. The computational cost can be reduced through dimensionality reduction.

When the network remains normal, d fluctuates slightly. However, when an anomaly occurs in the network, d will change apparently, as seen in Fig. 3. Figure 3 presents the color image of dissimilarity at normal and abnormal situations, where larger color values represent larger dissimilarity. It is a symmetric image. Both the horizontal axes and the vertical axes denote moments. The intersection of the i-th row and the j-th column represents dissimilarity d between data $\boldsymbol{s}_i$ at instant i and data $\boldsymbol{s}_j$ at instant j. The color value in the diagonal is always zero because the dissimilarity between one sample and itself is equal to zero. Figure 3a shows the normal situation, and Fig. 3b demonstrates the abnormal situation, where node 1 is disconnected at the fifth moment. As Fig. 3b shows, the color value in the fifth row and column is apparently larger, indicating the dissimilarity between sample $\boldsymbol{s}_5$ at the fifth instant and the other samples is much larger. So d can reflect network traffic anomalies.



(a) Normal

(b) Abnormal

Fig. 3 Dissimilarity between any two samples.

The parameter $\bar{d}$ is defined as the mean value of the dissimilarity between the sample at time i and any other sample, given by

$$\bar{d}_i = \frac{1}{M-1} \sum_{j=1}^{M} d(n_i, n_j) \tag{7}$$

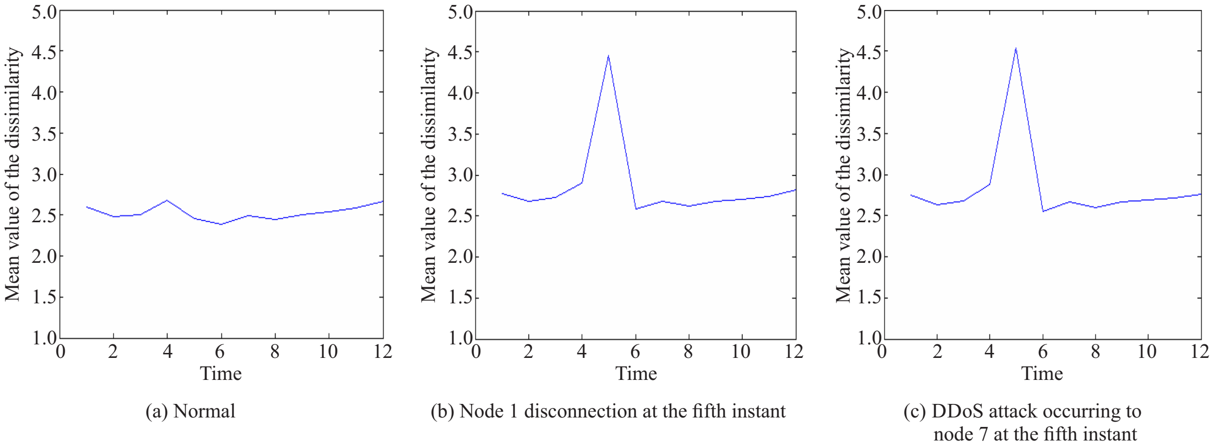
where $i = 1, 2, \ldots, 12$, $j = 1, 2, \ldots, 12$, $j \neq i$, $M = 12$, and M is the number of samples in the set of experiments. The metric $\bar{d}$ is regarded as a significant parameter for detecting anomalies.

Figure 4a demonstrates the normal situation. It is discovered that the average of $\bar{d}$ is 2.5258 and the variance of $\bar{d}$ is 0.0081. Therefore, $\bar{d}$ fluctuates smoothly when the network remains normal. Figure 4b shows the situation of a node 1 disconnection at the fifth instant, where $\bar{d}$ exhibits a sudden "spike" indicative of a network anomaly at the abnormal instant. Figure 4c presents the situation of a DDoS attack occurring to node 7 at the fifth instant. The same phenomenon is observed when the DDoS attack occurs to node 7. Thus, the anomaly can be easily observed.

Through many groups of experiments, similar scenarios can be obtained. In situations where a single node is abnormal, a sudden "spike" in the plot of parameter $\bar{d}$ can be captured at the abnormal instant. Therefore, single-node anomalies including node disconnection and DDoS attacks can be detected effectively by means of the parameter $\bar{d}$.

4.1.2 Anomaly scores

In Section 4.1.1, a threshold-based method is presented to separate all the principal components into a normal set and an abnormal set. As we know, the variance captured by one normal principal component is larger than that captured by one abnormal component, which means the projections of data onto normal principal components will contain more information regarding the original data. The first principal component contains the largest amount of information. The second principal



(a) Normal

(b) Node 1 disconnection at the fifth instant

(c) DDoS attack occurring to node 7 at the fifth instant

Fig. 4 Mean dissimilarity $\bar{d}$.

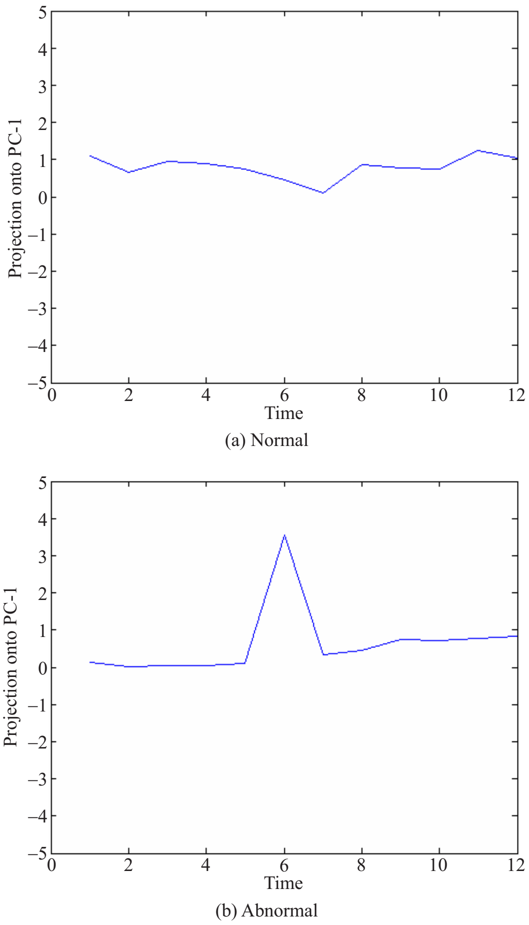
component contains the largest amount information of the remainder, and so on. Thus, the characteristic of the projections onto the first principal axis, $|\boldsymbol{u}_1|$, is the closest to the nature of the original OD flows, and carries the most significant patterns of the OD flows. In Fig. 5, the projections of samples onto the first principal axis are plotted. Figure 5a reflects the normal case, and Fig. 5b corresponds to the abnormal case. As shown, the time series of projections are fairly smooth in the normal case. In the abnormal case, a remarkable increase of OD flows can be observed. The projection $|\boldsymbol{u}_1|$ increases to 4 times the mean value of the normal situation. Therefore, it is considered that projections onto the first principal component can reflect network traffic changes.

The metric *Anomaly Score* R is defined as the ratio of the projection of data to be detected onto the first principal component to the average of all the projections onto the first principal axis:

$$R_i = \frac{|u_{1,i}|}{\dfrac{1}{M}\displaystyle\sum_{j=1}^{M} u_{1,j}} \tag{8}$$

where $i, j = 1, 2, \ldots, 12$. $u_{1,i}$ denotes the projection of a sample at the i-th instant onto the first principal axis. R represents the degree to which the projection at the current instant deviates from the mean state.

As with $\bar{d}$, an anomaly can be detected by R. R undulates slightly with mean value 1 and variance 0.143 in the normal state, and increases suddenly when disconnection or a DDoS anomaly occurs to the backbone network. This constitutes one of foundations for detecting the respective anomaly. It is confirmed in practice that the metric R works well for detecting anomalies including DDoS and disconnected nodes.



(a) Normal

(b) Abnormal

Fig. 5 Projections onto the first principal component.

4.2 Anomaly detection

As illustrated above, anomalies can be diagnosed visually by monitoring the two parameters $\bar{d}$ and R. But it is unknown how large a peak must be to be recognized as an anomaly. In the following, the problem can be addressed by studying the distributions of the two parameters.

The experimental dataset contains 288 samples during the whole day from the Abilene network on March 2nd, 2004. Using the function *normplot*, it is confirmed that the parameters $\bar{d}$ and R follow a normal distribution, which constitutes the basis for estimating the scale of normal patterns.

Since $\bar{d}$ and R follow a normal distribution, most of the data is distributed within a limited scale in normal cases. This property is used for anomaly detection. It can be observed that the confidence intervals of $\bar{d}$ and R at the 98% confidence level in normal situations, as Table 1 shows. Relying on the scales for $\bar{d}$ and R_1 in Table 1, it can be determined when a single-node anomaly is happening to the network with high accuracy. When two parameters of one sample are obtained, we first determine whether both values are within normal scale. If not, it indicates that traffic may be abnormal. Node disconnections and DDoS attacks can be detected by comparing their two parameters to normal scales, as Table 1 shows. By combining both parameters, high accuracy for detecting single-node anomalies (including disconnected nodes and DDoS attacks) and locating anomalous instants can be guaranteed. The approach works well in practice.

4.3 Performance evaluation

4.3.1 Result

We first implement our approach, and then develop a way to evaluate the detection performance of our algorithm. 540 samples from the Abilene network are used for the evaluation. For single-node anomalies, three groups of experiments are conducted. In each group, we employ 180 samples with 72 injected anomalies, so that we have enough data to compute accuracy. We find that the average accuracy is 93.33%, with a false-alarm rate controlled to below 8.33% for single-node anomalies. Therefore, our approach is effective for the two types of anomalies in the single-node case. The accuracy rate is defined as the fraction of the number of correctly detected samples (including true positive and true negative) over the total count of samples, and the false alarm rate is represented by the ratio of false alarms to the total count of normal samples.

Table 1 Scale of normal cases.

Parameter	Scale
$\bar{d}$	[2.3127, 3.1313]
R_1	$[-0.0662, 2.0662]$

4.3.2 Dimension discussion

As mentioned above, K, the number of normal principal components, is determined to be 11 by the empirical-threshold method. For different values of K, the accuracy rates may be different. To study the problem, experiments for various values of K are conducted. The results are shown in Fig. 6. Figure 6a demonstrates the change in the false alarm rate as K ranges from 1 to 11. The fluctuation of the accuracy is plotted in Fig. 6b as the dimension of the data changes. As can be seen, the accuracy increases and the false alarm rate decreases when the dimension increases. However, when $K \geqslant 7$, the accuracy and false alarm rate change little. Thus, for computational cost effectiveness, we take $K = 7$ as the optimal scheme.

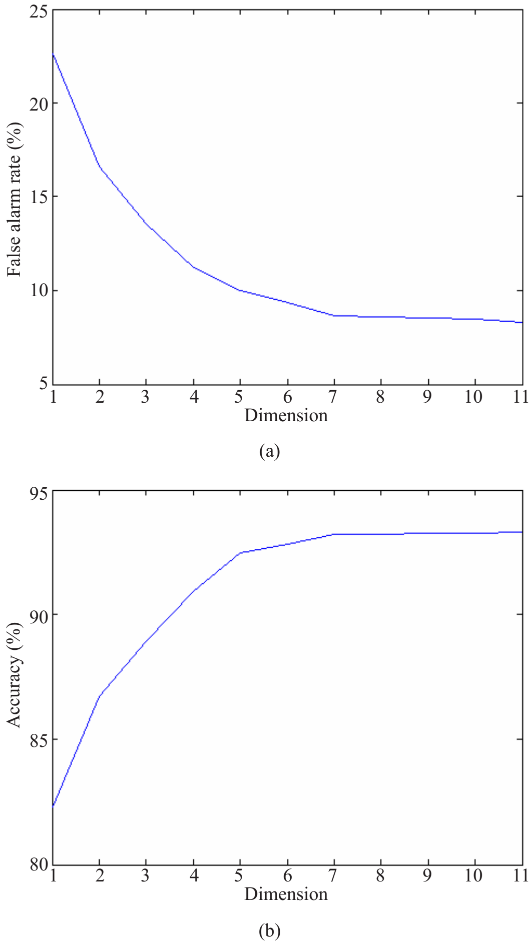


Fig. 6 False-alarm rate and accuracy.

5 Multi-node Anomaly Detection

When the network experiences an anomaly at one instant, there may be more than one anomaly source concerning multiple nodes. Multi-node situations can be more complex than single-node. Therefore, an improvement on the prior parameters is required in order to deal with multi-node anomaly detection. The parameters are introduced in Sections 5.1 and 5.2. In Section 5.3, we present the experimental evaluation.

5.1 Relative error of $\bar{d}$

It is discovered that when an anomaly occurs to multiple nodes in the backbone network, the value of $\bar{d}$ both at abnormal instants and normal instants increases as the number of anomalous nodes increases. Therefore, multi-node anomalies cannot be detected simply according to the scale of the normal pattern of the parameter $\bar{d}$ used in single-node situations, because it would result in a higher false-alarm rate.

To address the problem, the following formula is proposed, based on the concept of relative error.

$$\text{RE\_}\bar{d}_i = \frac{\bar{d}_i - \dfrac{1}{M-1}\sum\limits_{j \neq i} \bar{d}_j}{\dfrac{1}{M-1}\sum\limits_{j \neq i} \bar{d}_j} \tag{9}$$

where $i = 1, 2, \ldots, 12$. The parameter $\text{RE\_}\bar{d}_i$ represents the value of the relative error of $\bar{d}_i$ at the i-th instant. $\dfrac{1}{M-1}\sum\limits_{j \neq i} \bar{d}_j$ in the formula is considered as the "standard value", denoted by the average of parameter $\bar{d}$ for $M-1$ samples, leaving out the sample corresponding to the instant to be detected. As we know, when the network is normal, $\text{RE\_}\bar{d}$ should be smaller. When an anomaly occurs, $\text{RE\_}\bar{d}$ should be larger.

The parameter has a smooth trend with variance 0.0011 in a normal situation. It is revealed that when multiple nodes are disconnected from the network, a sharp increase of $\text{RE\_}\bar{d}$ is captured. So $\text{RE\_}\bar{d}$ is regarded as a parameter for multi-node anomaly detection, which can distinguish anomalous and normal states with high accuracy. Similar to the approach used in single-node situations, for estimating the scale of normal pattern of $\text{RE\_}\bar{d}$, a 98% confidence interval is calculated: [−0.1186, 0.1191]. When an anomaly occurs, $\text{RE\_}\bar{d}$ inclines to increase at the anomalous instant. So the upper limit of a 98% confidence interval is considered as the threshold for detecting anomalies and localizing abnormal instants.

5.2 Anomaly score

In single-node anomaly detection, the *Anomaly Score* R was introduced, which is formulated by the ratio of projection of sample points at the detected instant onto the first principal axis to the average of projections of all the samples. It is confirmed that R has good stability, and can therefore also be applied to multi-node anomaly detection. So through a large number of experiments, we come to a conclusion that R can be used to detect multi-node anomalies. The confidence interval is [−0.0662, 2.0662] at a 98% confidence level. Likewise, an upper limit of 2.0662 is selected as the threshold.

The scale of normal patterns for parameters $\text{RE\_}\bar{d}$ and R is shown in Table 2. By combining the two parameters, we can guarantee a high accuracy.

5.3 Performance evaluation

As in our approach to evaluating single-node anomaly detection performance, for the multi-node case we also conduct three groups of experiments. We find that the accuracy is 94.27% with a 6.02% false-alarm rate. Thus, multi-node anomalies can be detected effectively by our approach. What's more, as we can see, through improvement of the parameters, the accuracy rises in multi-node anomaly detection, although the complexity also increases.

6 Comparison

To compare with the existing anomaly detection techniques in Refs. [11, 12, 21], we conduct experiments on our datasets. In Refs. [11, 12], anomaly detecting based on DW was explored. The author diagnosed anomalies by a formulation combined with the coefficient matrix in Ref. [11]. In Ref. [12], the concept *contributed rate of energy* was proposed for detecting anomalies. In Ref. [21], the author proposed a model-based method, in which the anomaly nodes were discovered by an Exponentially Weighted Moving Average (EWMA) filter. The false-judgment rate (including False Positive (FP) and False Negative (FN)) obtained by the method in Ref. [10] was 18.55%,

Table 2 Scale of normal cases.

Parameter	Scale
$\text{RE\_}\bar{d}$	0.1191
R	2.0662

while the rate of our approach is 8.33% on average. Moreover, in our experiments, FN is always 0, which means our approach is able to recognize the entire set of anomaly nodes.

Our scheme is less complex than the method in Refs. [11, 12]. Although the complexity of the scheme proposed in Ref. [21] is much lower, the accuracy is remarkably lower than others. Table 3 demonstrates the detailed complexity and accuracy for DW-, PCA-, and Model-based algorithms. This shows that our approach compares favorably to Model-based and DW-based methods. Moreover, multi-node anomaly detection, for which there are many research works, is also explored by PCA in the paper. Thus, our approach outperforms the techniques in Refs. [11, 12, 21] in anomaly detection for large-scale backbone networks.

7 Conclusion

In this paper, we showed that the PCA-based approaches can carry out an effective analysis of OD flows by separating network traffic into anormal subspace and an abnormal subspace. Through exploring the analysis results, we developed a novel detection method for node disconnection and DDoS attacks in a backbone network by selecting two significant parameters from OD flows. The approach proposed in the paper is able to detect not only single-node anomalies but also multi-node anomalies by parameter improvement, with a high accuracy and a low false-alarm rate. It is clear that anomaly detection in backbone networks will attract more and more attention along with the increasing importance of network security. In the future, we intend to explore more PCA-based applications by combining our approach with other suitable techniques, such as routing-related anomaly detection, anomaly localization, and traffic prediction.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61100218).

Table 3 Comparison of algorithm.

Approach	Complexity	Accuracy (%)
Model-based	$O(n^2)$	75.18
DW-based	$O(n^3 \log n^3)$	88.70
PCA-based	$O(n^3)$	93.33

References

[1] A. Ward, P. Glynn, and K. Richardson, Internet service performance failure detection, *Performance Evaluation Review*, vol. 26, no. 3, pp. 38–44, 1998.

[2] W. Willinger, D. Rincón, and M. Roughan, Towards a meaningful MRA of traffic matrices, in *IMC Proceedings of ACM Sigcomm Conference on Internet Measurement*, 2008, pp. 331–336.

[3] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg, Fast accurate computation of large-scale IP traffic matrices from link loads, *ACM Sigmetrics Performance Evaluation Review*, vol. 31, no. 1, pp. 206–217, 2003.

[4] M. Crovella and E. Kolaczyk, Graph wavelets for spatial traffic analysis, *Proceedings-IEEE INFOCOM*, vol. 3, pp. 1848–1857, 2002.

[5] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, Compressed sensing for networked data, *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 92–101, 2008.

[6] M. Coates, Y. Pointurier, and M. Rabbat, Compressed network monitoring for IP and all-optical networks, in *ACM SIGCOMM Internet Measurement Conference (IMC)*, 2007, pp. 241–252.

[7] P. Barford, J. Kline, D. Plonka, and A. Ron, A signal analysis of network traffic anomalies, in *Proceedings of Internet Measurement Workshop*, 2002, pp. 71–82.

[8] J. Hellerstein, F. Zhang, and P. Shahabuddin, A statistical approach to predictive detection, *Computer Networks*, vol. 35, no. 1, pp. 77–95, 2001.

[9] G. Hamerly and C. Elkan, Bayesian approaches to failure prediction for disk drives, in *ICML'01 Proceedings of the Eighteenth International Conference on Machine Learning*, 2001, pp. 202–209.

[10] K. Shen, M. Zhong, and C. Li, I/O system performance debugging using model-driven anomaly characterization, in *4th USENIX Conference on File and Storage Technologies*, 2005, pp. 309–322.

[11] H. Tian, B. Zhong, and H. Shen, Diffusion wavelet-based analysis on traffic matrices by different diffusion operators, *Computers & Electrical Engineering*, vol. 40, no. 6, pp. 1874–1882, 2014.

[12] T. Sun, H. Tian, and X. Mei, Anomaly detection and localization by diffusion wavelet-based analysis on traffic matrix, *Computer Science and Information Systems*, vol. 12, no. 4, pp. 1361–1374, 2015.

[13] Y. Qian, M. Chen, and Q. Hao, ODC: A method for online detecting & classifying network-wide traffic anomalies, *Journal on Communications*, pp. 134–141, 2011.

[14] Y. Zhang, S. Singh, S. Sen, N. Duffield, and C. Lund, Online identification of hierarchical heavy hitters: Algorithms, evaluation, and applications, in *Proc. of the 4th ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2004, pp. 101–114.

[15] A. Lakhina, M. Crovella, and C. Diot, Diagnosing network-wide traffic anomalies, *Computer Communication Review*, vol. 34, no. 4, pp. 219–230, 2004.

[16] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft, Structural analysis of network traffic flows, *ACM Sigmetrics Performance Evaluation Review*, vol. 32, no. 1, pp. 61–72, 2004.

[17] C. Wang and S. Mahadevan, Multiscale dimensionality reduction based on diffusion wavelets, Technical Report, Department of Computer Science, University of Massachusetts, USA, 2009.

[18] H. Beitollahi and G. Deconinck, Connection score: A statistical technique to resist application layer DDoS attacks, *Journal of Ambient Intelligence and Humanized Computing*, vol. 5, no. 3, pp. 425–442, 2014.

[19] L, Huang, X. L. Nguyen, M. N. Garofalakis, M. I. Jordon, A. D. Joseph, and N. Taft, In-network PCA and anomaly detection, in *Advances in Neural Information Processing Systems 19 (NIPS 2006)*, 2006, pp. 617–624.

[20] Z. Zheng, Y. Li, and Z. Lan, Anomaly localization in large-scale clusters, in *IEEE International Conference on Cluster Computing*, 2007, pp. 322–330.

[21] B. Eriksson, P. Barford, R. Bowden, N. Duffield, J. Sommers, and M. Roughan, BasisDetect: A model-based network event detection framework, in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, Melbourne, Australia, 2010, pp. 1–30.

Hui Tian is currently an associate professor in School of Electronics and Information Engineering, Beijing Jiaotong University. She received the BEng and MEng degrees from Xidian University, China in 2000 and 2003, respectively, and PhD degree from Japan Advanced Institute of Science and Technology in 2006. Her research interests include network performance evaluation, telecommunications, and wireless sensor networks.

Meimei Ding received the BE degree from Changchun University, China in 2010. She is currently a master student in Beijing Jiaotong University, China. Her research interest is network performance analysis.