Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2015, Article ID 923792, 26 pages
http://dx.doi.org/10.1155/2015/923792

Research Article

Network-Wide Traffic Anomaly Detection and Localization Based on Robust Multivariate Probabilistic Calibration Model

Yuchong Li,[1,2,3] Xingguo Luo,[1] Yekui Qian,[2] and Xin Zhao[2]

[1] *National Digital Switching System Engineering & Technological Research Center, Jianxue Street No. 7, Jinshui District, Zhengzhou 450002, China*
[2] *Air Defence Forces Academy of PLA, Zhengzhou, China*
[3] *Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory, Shijiazhuang, China*

Correspondence should be addressed to Yuchong Li; yuchonglee@163.com

Received 29 April 2015; Revised 5 August 2015; Accepted 6 August 2015

Academic Editor: Mark Leeson

Network anomaly detection and localization are of great significance to network security. Compared with the traditional methods of host computer, single link and single path, the network-wide anomaly detection approaches have distinctive advantages with respect to detection precision and range. However, when facing the actual problems of noise interference or data loss, the network-wide anomaly detection approaches also suffer significant performance reduction or may even become unavailable. Besides, researches on anomaly localization are rare. In order to solve the mentioned problems, this paper presents a robust multivariate probabilistic calibration model for network-wide anomaly detection and localization. It applies the latent variable probability theory with multivariate t-distribution to establish the normal traffic model. Not only does the algorithm implement network anomaly detection by judging whether the sample's Mahalanobis distance exceeds the threshold, but also it locates anomalies by contribution analysis. Both theoretical analysis and experimental results demonstrate its robustness and wider use. The algorithm is applicable when dealing with both data integrity and loss. It also has a stronger resistance over noise interference and lower sensitivity to the change of parameters, all of which indicate its performance stability.

1. Introduction

Network traffic anomalies are unusual and significant changes at network's traffic level. Intrusions such as DDos attacks and zombie networks significantly jeopardize the Internet security, and network jams and malfunctions have unpleasant impact on service quality; therefore it is critical to detect and locate network anomalies for both network operators and end users. It is a challenging task to detect and locate them because one must extract and interpret anomalous patterns from large amounts of high-dimensional, intricate, and noisy background traffic data.

There are a great number of researches on anomaly detection. Host-based anomaly detection system monitors and analyzes the internals of a computing system by applying data mining of the system logs and audit records [1, 2]; another detection method based on performance measurement data such as end to end round-trip time and packet loss probability in a single path can be implemented by single variable time series analysis [3, 4]; network anomaly detection based on traffic measurements from single link can be implemented by applying machine learning and signal analysis [5, 6]. All of these methods have their limits, because they only concentrate on a part of the information and their detection area is limited. When the scale of network enlarges and data transfer rate speeds up, many network anomalies often exhibit strong network-wide characteristics [7, 8], and their impact may always spread to multiple links, while their local characteristics may not be that obvious. It is difficult to conduct network-wide analysis with above methods and their accuracy cannot be guaranteed.

In order to solve the problems mentioned above, Lakhina et al. come up with network-wide anomaly detection based on subspace construction via PCA [9]. This method employs network traffic of many Origin-Destination (OD) flows to establish a model of normal behavior and detects anomalies by measuring deviations from that model. With the concept of network-wide detection, researches are conducted in space-time expansibility [10–13], robustness [14–16], real-time processing [17, 18], and anomaly measure [19, 20], which enrich network-wide anomaly detection. This kind of methods uses whole-network traffic data, which has huge performance advantages over single point, single path, and single link. At the same time, compared with other methods, it sets up normal behavior model which enables us to avoid building anomalous feature library. Thus it can be implemented to detect known anomalies as well as unknown anomalies, and it can be used widely. Network-wide anomaly detection improves detection performance by introducing wider and multidimensional network information, but it also faces some real problems when implemented in large scale and high speed backbone network. Firstly, because of its wider collection area, more collection equipment, and faster network speed, it might not be applicable if collected data were lost during collection or transfer process [21]; secondly, traffic flows in backbone network continue to grow in volume and complexity, and hidden noise like anomalous traffic could degrade performance of the anomaly detection algorithms [22], while some of the attacks might even pollute the detectors [14–16]; thirdly, the above anomaly detection methods can only find when anomalies happen, but they still have some defects on locating those anomalies [22].

Therefore, we propose an approach named RMPCM based on robust multivariate probabilistic calibration model to overcome these problems discussed above. This anomaly detection and locating algorithm introduces a latent variable probabilistic model based on t-distribution instead of a Gaussian distribution to establish a normal traffic model. By judging if the sample's Mahalanobis distance from the normal model exceeds the threshold, traffic anomaly detection is achieved. Locating anomalies is attained with contribution analysis. RMPCM approach is more robust; not only can it be widely used for processing complete data as well as missing data, but also it acquires stronger robustness under the noise interference and lower sensibility of model parameters. The contributions of this paper consist of following 4 aspects; the RMPCM approach can

(1) solve anomaly detection problem when data loss occurs by establishing a latent variable probabilistic model,

(2) increase detection accuracy by introducing multivariate t-distribution to relieve noise interference in modeling the normal traffic behavior,

(3) correctly locate the underlying Origin-Destination (OD) flows being the source of the anomaly,

(4) reduce the amount of work involved in implementing complicated parameter testing, because RMPCM has a better stability and lower sensibility for model parameters.

This paper is organized as follows. We begin in Section 2 with a discussion of the related work. We describe data source model and problems that need to be solved in Section 3. In Section 4, we describe the RMPCM approach in detail and solve three problems raised in Section 3. We validate our approach in three different ways of experiments and contrast our RMPCM with existing approach on traffic anomaly detection in Section 5. A discussion of several details is presented in Section 6. Concluding remarks and our ongoing work are presented in Section 7.

2. Related Work

Back in 1987 Denning had demonstrated statistic model for detecting network anomalies [23]. And it is becoming more and more important with the development of the Internet. There are many traditional anomaly detection approaches based on host computer, single path, and single link. Researches in [7] indicated the generation and development of network anomalies have exhibited a tendency of network-wide characteristic. They found that the performance of the anomaly detectors increases with enlarging the range of detection beyond linear growth, which sets up the foundation for network-wide anomaly detection. The authors of [9] proposed network-wide anomaly detection algorithm based on network traffic for the first time, which illustrated low dimensionality of OD flows. They also integrated traffic statistics of multiple OD flows to build up a model of normal behavior and detected anomalies by measuring deviations from that model. The authors of [11] came along to improve the anomaly detection approach based on PCA by applying stochastic matrix perturbation theory and proposed a PCA-based distributed approach for network-wide anomaly detection. Reference [13] expanded the classical PCA and proposed the Karhunen-Loeve expansion for network-wide anomaly detection. Reference [17] proposed an online anomaly detection approach using kernel recursive least squares algorithm to solve the problem of online detection. All of them did not address the problems of how to detect anomalies in the condition of noise interference and data loss and how to locate the anomalies in the actual network environment.

The authors of [22] took in-depth study on the influence of anomalous traffic on the performance of detector and indicated that large anomalies may cause the offset of normal model based on PCA, which increased the false positive rate (FPR) of anomaly detection. References [14, 15] took further steps to study poisoning attacks on anomaly detectors and evaluated poisoning techniques and developed defense. The authors of [16] listed 3 mechanisms of poisoning attacks and proposed defense based on robust PCA with projection pursuit. All of the above only focus on the poisoning and defense techniques based on PCA detector, but there is lack of researches on common modeling approaches in the intricate and noisy environment.

Data loss is very common in many fields, and the question of how to get enough information from missing data needs to be answered. The authors of [24, 25] proposed an algorithm to solve the problem of goodness-of-fit test for varying coefficient models with missing data. The authors of [21] gave their opinion on data loss problem when network traffic flows were measured in large scale and high speed backbone network; they proposed an approach of sparsity regularized matrix factorization (SRMF) to make the data complete. This is applicable in traffic engineering, capacity planning, forecasting, and so forth, but it did not research deeply in network anomaly detection when facing data loss.

Network anomaly detection can determine when anomalies take place, but locating anomaly is an extremely challenging task. The authors of [22] point out the deficiency of network-wide detection algorithm based on PCA in locating anomaly. Reference [26] then proposed an approach of Basis-Detect for network-wide anomaly detection and locating, but it could only locate anomalies to border router; it was unable to pinpoint the position.

In this paper we propose a network-wide anomaly detection algorithm based on RMPCM, which will later be proved to have a better performance in solving problems of noise interference, data loss, and locating anomalies.

3. Data Model and Problems Description

3.1. Data Model. Conventionally the researches of the Internet traffic flow mainly focused on temporal characteristics of data package on a single link, which help in developing concepts of self-similar stochastic processes, long-range dependence, and so forth. One ISP (Internet service provider), however, consists of hundreds of those links which are connected all over, and the Internet contains several thousand ISPs. In such a vast background the spatial characteristics of network traffic come to people's attention inevitably. However, it is difficult to analyze traffic flow data of all links in the network simultaneously, because it amplifies the complexity of modeling traffic on a single link which is itself a complicated task. As compact and elegant descriptions of traffic flows between nodes in a certain network structure, traffic matrix is a constantly employed model to conduct explorations on the spatiotemporal component of network-wide traffic. Traffic matrix is an overview of network-wide traffic. Instead of studying traffic on all links, applying traffic matrix provides more straightforward and fundamental insights into network-wide traffic study [8]. We employ traffic matrix at PoP (point of presence) level as data source in our research.

Traffic matrix at PoP level: assume that an autonomous system (AS) has n PoPs. Continuous measuring of the traffic between each PoP pair at a certain period can obtain the traffic of Origin-Destination (OD) flows. An OD flow denotes the collection of all traffic that enters the network from an ingress node and departs from an egress node. Arranging these point-to-point measured values in $N \times D$ matrix can obtain this AS's traffic matrix $\mathbf{X}$. As shown in Figure 1, N denotes the number of measurement periods, and D denotes the number of measured value of OD flows at each measurement ($D = n \times n$). The element x_{ij} in $\mathbf{X}$ denotes the volume of a certain traffic measure at the ith period and the jth OD. Traffic volume (the number of bytes, packets, or flows) is adopted as a traffic measure in this paper.

3.2. Problem Description

3.2.1. Data Loss. In the process of collecting data, data loss may occur. This is because massive data in high speed backbone network may increase burden of collecting equipment and reduce its stability. Another reason is due to network congestion, equipment or link malfunctions when transferring data.

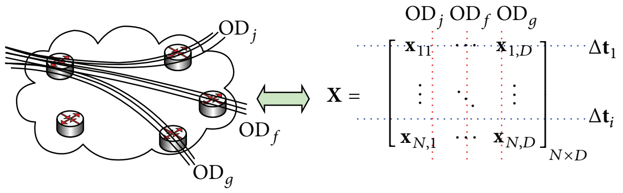

OD_j
OD_f
OD_g
$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_{1,D} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{N,1} & \cdots & \mathbf{x}_{N,D} \end{bmatrix}_{N\times D}$
OD_j OD_f OD_g
$\Delta\mathbf{t}_1$
$\Delta\mathbf{t}_i$

Figure 1: Schematic diagram of traffic matrix.

Traffic data loss is not all completely random, and many of the cases are highly structural. In order to describe the scenario of data loss in the process of data collection and transfer, four kinds of loss mechanisms are adopted.

(1) PureRandLoss. The elements in X are missing independently at random with probability p. This may be due to unexpected congestion by chance in measuring equipment or unreliable transfer mechanism.

(2) PeriodRandLoss. The rows of traffic matrix $\mathbf{X}$ are corresponding to measurement periods, and PeriodRandLoss means the rows in X are missing at random with probability p. The reason of this type of structured loss may be that storage devices are overloaded or program breaks down when centralized processing large number of measured data during this period.

(3) ODRandLoss. The columns of traffic matrix $\mathbf{X}$ are corresponding to OD flows, and ODRandLoss means the columns in $\mathbf{X}$ are missing at random with probability p. This type of structured loss could stem from OD identification error caused by either flow filtration or links/routers' malfunction.

(4) PieceRandLoss. PieceRandLoss means the submatrixes in X are missing at random with probability p. This type of structured loss may be caused when storage devices are full but they are still keeping collecting data for several periods or devices' breakdown for some time, which is corresponding to data loss of traffic matrix multiple adjacent columns. Note that the case single row PieceRandLoss corresponds to PeriodRandLoss, and single column PieceRandLoss corresponds to ODRandLoss.

3.2.2. Noise Interference. Model parameter estimation often uses maximum likelihood estimation (MLE) when samples are known, and the probability distribution of the samples is needed. However, it is very troublesome to accurately describe the distribution of data; therefore it is always assumed that the data generally follows the Gaussian distribution because its nice analytical property always yields tractable algorithms. The MLE is equivalent to the least square estimation in the linear Gaussian regression model which is noted for its unduly sensitive to atypical samples such as outliers and it would affect the accuracy of the model [27]. In real network measurement collected traffic

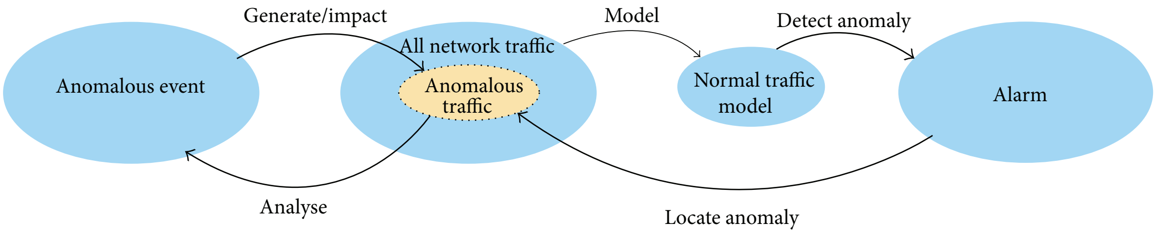


Figure 2: Relation schema of RMPCM.

data contains anomalous traffic which is outliers and will have a huge interference in the establishment of model if a Gaussian noise mode is selected. The question of how to set up a more accurate model of normal traffic under this circumstance needs to be addressed promptly.

3.2.3. Anomaly Localization. The time at which anomalies take place can be determined by implementing network anomaly detection, but locating anomalies is crucial if we want to pinpoint and solve security problems more precisely. Locating anomalies in this paper is corresponding to pointing out the intersections of the rows and columns of traffic matrix $\mathbf{X}$ when anomalies occur. The row of $\mathbf{X}$ corresponds to the time of anomaly occurrence, and the column of $\mathbf{X}$ corresponds to the position where anomaly occurs (i.e., on one OD or a few OD).

4. RMPCM

The relationship between anomalous event generation and its detection and localization is shown in Figure 2. Anomaly events will affect part of traffic flows in a network and thus cause changes of corresponding statistics of overall network traffic. The network anomaly detectors can analyze these changes of statistics and raise an alarm; then network administrators can analyze anomalous traffic by locating anomalies to determine anomaly event. RMPCM which we propose consists of the process of anomaly detection and localization: firstly, model normal traffic with collected traffic data and then determine whether or not this sample is an anomaly by judging if the sample's Mahalanobis distance exceeds the threshold. Secondly, contribution analysis is applied to anomalous samples in order to locate anomalies. The RMPCM method can be divided into 3 steps in detail which include normal traffic modeling, anomaly detection, and anomaly localization.

4.1. Introducing the Model. Applying network anomaly detection algorithm in the real world may encounter difficulties such as data loss in the process of transferring and collecting and modeling deviation caused by noise interference.

Traditional network anomaly detection algorithms will not be applicable any longer in the condition of incomplete data. It is considered to adopt Bayes method, but because of the complexity of network traffic data posterior mean estimation and asymptotic variance cannot be directly derived from this method. Therefore a latent variable probabilistic model is to be introduced, meaning some "latent data" are to be added in the known data in order to simplify the parameter estimation. In this process missing data along with unknown parameters treated as "latent data" will be solved by applying expectation-maximization (EM) algorithm to achieve the maximum likelihood estimation (MLE) of model parameters.

When computing MLE, probabilistic distribution of known data is required. Normally it is assumed that they are normally distributed, but because of some anomalous traffic, this assumption will cause parameter estimation to have a large deviation; therefore, multivariate Gaussian distribution is replaced by t-distribution in our paper. Compared with Gaussian distribution, t-distribution has heavier tails, which is a desirable property to handle data sets in the presence of anomalies. The explanation is as follows.

Specifically, suppose sample data $\mathbf{x}_i$ ($i = 1, \dots, N$) are recorded (N represents the number of the samples), and one assumes that they are independent identically distributed Gaussian random vectors:

$$\mathbf{x}_i \sim N_D \{\boldsymbol{\mu}(\theta), \boldsymbol{\Sigma}(\varphi)\} \quad (1)$$

in which D is the dimension of the sample, $\boldsymbol{\mu}$ is the mean vector function with parameter θ, and $\boldsymbol{\Sigma}$ is the covariance matrix with parameter φ.

t-distribution model replaces the above model, which is

$$\mathbf{x}_i \sim t_D \{\boldsymbol{\mu}(\theta), \boldsymbol{\Lambda}(\varphi), \nu\}. \quad (2)$$

D is also the dimension of the sample, $\boldsymbol{\mu}$ is the location vector, $\boldsymbol{\Lambda}$ is the scale matrix, and ν is degree of freedom for t-distribution; probabilistic density function (p.d.f.) is calculated as

$$S(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu) = \frac{|\boldsymbol{\Lambda}|^{-1/2}\,\Gamma\{(\nu + D)/2\}}{\{\Gamma(1/2)\}^{D}\,\Gamma(\nu/2)\,\nu^{D/2}} \times \left(1 + \frac{\kappa^2}{\nu}\right)^{-(\nu+D)/2}, \quad (3)$$

where $\kappa^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ is the squared Mahalanobis distance and $\Gamma(\cdot)$ is Gamma function. When $\nu < \infty$, t-distribution model has a better robust against outlier's interference over normal distribution by applying MLE. This is due to outliers that have relatively large Mahalanobis distance but comparably small contribution for model parameter estimation. The estimation of θ is taken as example to demonstrate the following.

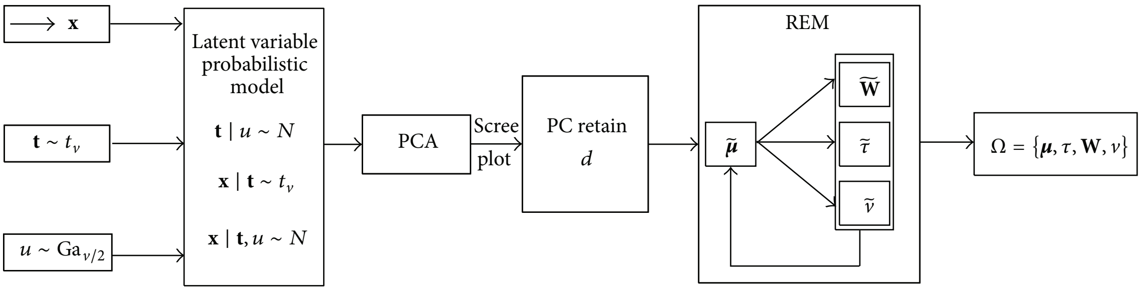

x
$\mathbf{t} \sim t_\nu$
$u \sim \mathrm{Ga}_{\nu/2}$
Latent variable probabilistic model
$\mathbf{t} \mid u \sim N$
$\mathbf{x} \mid \mathbf{t} \sim t_\nu$
$\mathbf{x} \mid \mathbf{t}, u \sim N$
PCA
Scree plot
PC retain d
REM
$\tilde{\boldsymbol{\mu}}$
$\widetilde{\mathbf{W}}$
$\tilde{\tau}$
$\tilde{\nu}$
$\Omega = \{\boldsymbol{\mu}, \tau, \mathbf{W}, \nu\}$

Figure 3: Steps of modeling normal traffic in noisy traffic.

For Gaussian distribution model (1), the equation for its MLE is $\partial l/\partial\theta = \sum_{i=1}^{N} \mathbf{A}_i \sum_i^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_i) = 0$, where l is the corresponding log-likelihood and $\mathbf{A}_i$ is the matrix of partial derivatives of $\boldsymbol{\mu}_i$ with respect to θ; the MLE of θ under t-distribution model (2) satisfies $\partial l/\partial\theta = \sum_{i=1}^{N} w_i \mathbf{A}_i \boldsymbol{\Lambda}_i^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_i) = 0$, where

$$w_i = \frac{(\nu + D)}{(\nu + \kappa_i^2)} \tag{4}$$

is the weight assigned to sample i. The weight w_i is a function of Mahalanobis distance κ_i which indicates that w_i decreases while κ_i increases. The anomaly sample has a relatively larger κ_i; therefore w_i is smaller, meaning adopting t-distribution in model parameter estimation lowers the sensitivity to anomaly samples. Compared with Gaussian distribution, t-distribution model has a stronger robustness.

4.2. Normal Traffic Modeling

4.2.1. Normal Traffic Modeling in Noisy Traffic. RMPCM models normal traffic in noisy traffic by establishing a latent variable probabilistic model based on multivariate t-distribution, and the procedure is illustrated in Figure 3.

In order to solve the problem that Gaussian noise models are too sensitive to atypical observations such as anomalous traffic observations, we suppose the noise is drawn from t-distribution model instead of a Gaussian noise model. Assume each d-dimensional latent vector $\mathbf{t}_i$ comes from a linear probabilistic projection of D-dimensional D $(D \geq d)$ eigenvector $\mathbf{x}_i$, and we build a latent variable probabilistic model where we select a unit variance t-distribution as the prior distribution on the latent vectors. The probabilistic model is

$$\begin{gathered} p(\mathbf{t}_i) = S(\mathbf{t}_i \mid 0, \mathbf{I}_d, \nu), \\ p(\mathbf{x}_i \mid \mathbf{t}_i) = S(\mathbf{x}_i \mid \mathbf{W}\mathbf{t}_i + \boldsymbol{\mu}, \tau\mathbf{I}_D, \nu), \end{gathered} \tag{5}$$

where $\mathbf{W}$ is the projection matrix, $\boldsymbol{\mu}$ is the location vector, and $\mathbf{I}$ is the unit matrix.

It is inapplicable to be analyzed and resolved using MLE directly. As noted in [28] t-distribution model can be extended to an infinite Gaussian mixture model with the same mean where the prior distribution on u is a Gamma distribution with parameters depending only on t-distribution's degrees of freedom ν:

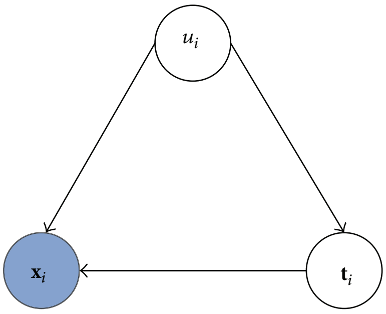


Figure 4: Graphical model of RMPCM. The shaded node is the observed vector, and arrows denote conditional dependencies between these random variables.

$$\begin{aligned} p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu) & \\ &= \int_0^{+\infty} N\left(\mathbf{x} \mid \boldsymbol{\mu}, u\boldsymbol{\Lambda}^{-1}\right) \mathrm{Ga}\left(u \mid \frac{\nu}{2}, \frac{\nu}{2}\right) du, \end{aligned} \tag{6}$$

where $\mathrm{Ga}(u \mid \alpha, \beta) = \beta^{\alpha} u^{\alpha-1} e^{-\beta u}/\Gamma(\alpha)$ is the probability density function of Gamma distribution. Based on (6), latent variable u is introduced, and u is a scalar, $u \sim \mathrm{Ga}(\nu/2, \nu/2)$. As can be known in [29], if $\mathbf{t} \sim t(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$, then the conditional distribution of $\mathbf{t} \mid u$ is Gaussian: $\mathbf{t} \mid u \sim N(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Lambda})$. The latent variable model can be derived as follows. Figure 4 indicates the relationship of variables in the model:

$$p(u_i) = \mathrm{Ga}\left(u_i \mid \frac{\nu}{2}, \frac{\nu}{2}\right), \tag{7}$$

$$p(\mathbf{t}_i \mid u_i) = N\left(\mathbf{t}_i \mid 0, u_i^{-1}\mathbf{I}_d\right), \tag{8}$$

$$p(\mathbf{x}_i \mid \mathbf{t}_i, u_i) = N\left(\mathbf{W}\mathbf{t}_i + \boldsymbol{\mu}, u_i^{-1}\tau\mathbf{I}_D\right). \tag{9}$$

In order to calculate model parameters, $u \mid \mathbf{x}$ and $\mathbf{t} \mid \mathbf{x}, u$ are also needed to be calculated, and from (8) and (9), it can be solved that

$$\begin{aligned} p(\mathbf{x}_i \mid u_i) &= \int p(\mathbf{x}_i \mid \mathbf{t}_i, u_i)\, p(\mathbf{t}_i \mid u_i)\, d\mathbf{t} \\ &= N\left(\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T + \frac{\tau\mathbf{I}_D}{u_i}\right). \end{aligned} \tag{10}$$

Set $\mathbf{\Psi} = \mathbf{W}\mathbf{W}^T + \tau\mathbf{I}_D$, which is $D \times D$ matrix, and the marginal distribution of $\mathbf{x}$ follows t-distribution, which is $p(\mathbf{x}_i) = S(\boldsymbol{\mu}, \mathbf{\Psi}, \nu)$.

From (7) and (10) it can be calculated that

$$\begin{aligned} p(u_i \mid \mathbf{x}_i) &\propto p(\mathbf{x}_i \mid u_i)\, p(u_i) \\ &= \mathrm{Ga}\left(u_i \mid \frac{D+\nu}{2}, \frac{\delta^2+\nu}{2}\right), \end{aligned} \tag{11}$$

where $\delta^2 = (\mathbf{x}_i - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})$.

From (8) and (9) it can be computed that

$$\begin{aligned} p(\mathbf{t}_i \mid \mathbf{x}_i, u_i) &\propto p(\mathbf{x}_i \mid \mathbf{t}_i, u_i)\, p(\mathbf{t}_i \mid u_i) \\ &= N\left(\mathbf{M}^{-1}\mathbf{W}^T(\mathbf{x}_i - \boldsymbol{\mu}), \frac{\tau\mathbf{M}^{-1}}{u_i}\right), \end{aligned} \tag{12}$$

where $\mathbf{M} = \mathbf{W}^T\mathbf{W} + \tau\mathbf{I}_d$ is $d \times d$ matrix and $\mathbf{t} \mid \mathbf{x}$ follows t-distribution, which is $p(\mathbf{t}_i \mid \mathbf{x}_i) = S(\mathbf{M}^{-1}\mathbf{W}^T(\mathbf{x}_i - \boldsymbol{\mu}), \tau\mathbf{M}^{-1}, \nu)$.

The above procedure establishes the latent variable probabilistic model by replacing Gaussian distribution model with t-distribution model.

The model parameters $\{\boldsymbol{\mu}, \tau, \mathbf{W}, \nu, d\}$ need to be estimated. Parameter d is the intrinsic dimension of traffic matrix $\mathbf{X}$. As shown in Figure 3 the value of d is determined by scree plot based on PCA, and the details are in Section 6.1. Parameters $\{\boldsymbol{\mu}, \tau, \mathbf{W}, \nu\}$ are estimated by MLE, where corresponding log-likelihood is

$$L = \sum_{i=1}^{N} \ln p(\mathbf{x}_i, \mathbf{t}_i, u_i). \tag{13}$$

MLE can be calculated by applying EM algorithm. In order to simplify the calculation, REM (rapid expectation-maximization) is employed in this paper. REM remarkably increases the algorithm's rate of convergence. This algorithm consists of 2 stages, estimating different parameters by applying EM algorithm in every stage, and does 2 stages' iteration until satisfying the condition of convergence.

The First Stage. It is only estimating $\boldsymbol{\mu}$, ignoring $\mathbf{t}_i$ in the first stage [30].

The log-likelihood is

$$L_1 = \sum_{i=1}^{N} \ln p(\mathbf{x}_i, u_i) = \sum_{i=1}^{N} \ln\{p(\mathbf{x}_i \mid u_i)\, p(u_i)\}. \tag{14}$$

E-Step. As can be known from Jensen inequality, when $p(u_i) = p(u_i \mid \mathbf{x}_i)$, the inequality becomes equality, and the lower bound of $L_1$is set. Its expectation can be calculated from (10), (11), and (14) and get (15), where $\langle\cdot\rangle$ is the expectation operator:

$$\langle L_1 \rangle = -\sum_{i=1}^{N} \langle u_i \rangle (\mathbf{x}_i - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}), \tag{15}$$

where

$$\langle u_i \rangle = \frac{\nu + D}{\nu + (\mathbf{x}_i - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})}. \tag{16}$$

M-Step. Maximization of $\langle L_1 \rangle$ with respect to $\boldsymbol{\mu}$ can give the estimated value of $\boldsymbol{\mu}$. This process can be implemented by setting the partial derivative of $\langle L_1 \rangle$ with respect to $\boldsymbol{\mu}$ at 0:

$$\widetilde{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{N} \langle u_i \rangle \mathbf{x}_i}{\sum_{i=1}^{T} \langle u_i \rangle}. \tag{17}$$

The Second Stage. In the second stage, REM algorithm introduces the latent variable $\mathbf{t}_i$ and estimates parameters $\{\tau, \mathbf{W}, \nu\}$. The estimated value $\widetilde{\boldsymbol{\mu}}$ from the first stage is used in the second stage.

The log-likelihood is

$$\begin{aligned} L_2 &= \sum_{i=1}^{N} \ln p(\mathbf{x}_i, \mathbf{t}_i, u_i) \\ &= \sum_{i=1}^{N} \ln\{p(\mathbf{x}_i \mid \mathbf{t}_i, u_i)\, p(\mathbf{t}_i \mid u_i)\, p(u_i)\}. \end{aligned} \tag{18}$$

E-Step. When $p(\mathbf{t}_i, u_i) = p(\mathbf{t}_i, u_i \mid \mathbf{x}_i)$, the expectation of L_2 can be calculated from (7), (8), and (9):

$$\begin{aligned} \langle L_2 \rangle = -\sum_{i=1}^{N} \Bigg\{ & \frac{D}{2}\ln\tau + \frac{\langle u_i \rangle}{2\tau} (\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})^T (\mathbf{x}_i - \widetilde{\boldsymbol{\mu}}) \\ & - \frac{1}{\tau} \langle u_i \mathbf{t}_i \rangle^T \mathbf{W}^T (\mathbf{x}_i - \widetilde{\boldsymbol{\mu}}) + \frac{1}{2\tau} \operatorname{tr}\left(\mathbf{W}^T\mathbf{W} \left\langle u_i \mathbf{t}_i \mathbf{t}_i^T \right\rangle\right) \\ & + \frac{\nu}{2}\log\frac{\nu}{2} + \frac{\nu - 2}{2} \langle \log u_i \rangle - \log\Gamma\left(\frac{\nu}{2}\right) \\ & - \frac{\nu}{2} \langle u_i \rangle \Bigg\}, \end{aligned} \tag{19}$$

where $\langle u_i \rangle$ is in (16). Consider

$$\langle \mathbf{t}_i \rangle = \mathbf{M}^{-1}\mathbf{W}^T (\mathbf{x}_i - \boldsymbol{\mu}), \tag{20}$$

$$\langle u_i \mathbf{t}_i \rangle = \langle u_i \rangle \langle \mathbf{t}_i \rangle, \tag{21}$$

$$\left\langle u_i \mathbf{t}_i \mathbf{t}_i^T \right\rangle = \tau\mathbf{M}^{-1} + \langle u_i \rangle \langle \mathbf{t}_i \rangle \langle \mathbf{t}_i \rangle^T, \tag{22}$$

$$\begin{aligned} \langle \log u_i \rangle = {} & \psi\left(\frac{\nu + D}{2}\right) \\ & - \log\left(\frac{\nu + (\mathbf{x}_i - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})}{2}\right), \end{aligned} \tag{23}$$

where $\psi(\cdot)$ denotes the digamma function in (23).

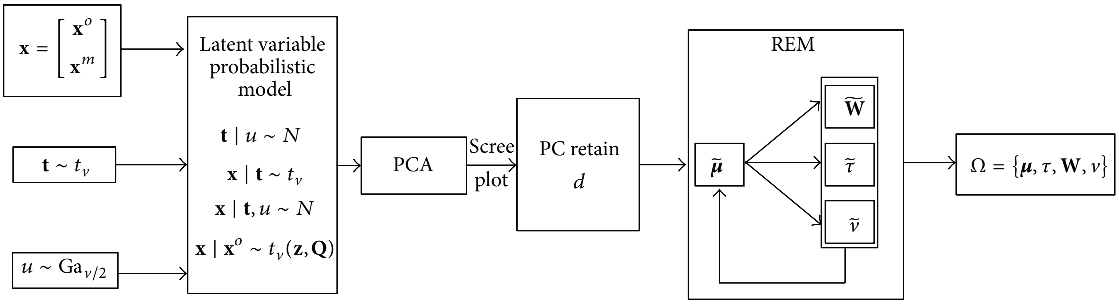

$\mathbf{x} = \begin{bmatrix}\mathbf{x}^o\\ \mathbf{x}^m\end{bmatrix}$
$\mathbf{t} \sim t_\nu$
$u \sim \mathrm{Ga}_{\nu/2}$
Latent variable probabilistic model
$\mathbf{t} \mid u \sim N$
$\mathbf{x} \mid \mathbf{t} \sim t_\nu$
$\mathbf{x} \mid \mathbf{t}, u \sim N$
$\mathbf{x} \mid \mathbf{x}^o \sim t_\nu(\mathbf{z}, \mathbf{Q})$
PCA
Scree plot
PC retain d
REM
$\widetilde{\boldsymbol{\mu}}$
$\widetilde{\mathbf{W}}$
$\widetilde{\tau}$
$\widetilde{\nu}$
$\Omega = \{\boldsymbol{\mu}, \tau, \mathbf{W}, \nu\}$

Figure 5: Steps of modeling normal traffic with data loss.

M-Step. Maximizing $\langle L_2\rangle$ with respect to $\mathbf{W}$ and τ gives the following updating formulas of $\{\mathbf{W}, \tau\}$:

$$\widetilde{\mathbf{W}} = \left(\sum_{i=1}^{N}(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})\langle u_i\mathbf{t}_i\rangle^T\right)\left(\sum_{i=1}^{N}\left\langle u_i\mathbf{t}_i\mathbf{t}_i^T\right\rangle\right)^{-1},$$
$$\widetilde{\tau} = \frac{1}{ND}\sum_{i=1}^{N}\left\{\langle u_i\rangle\left\|\mathbf{x}_i - \widetilde{\boldsymbol{\mu}}\right\|^2 - 2\langle u_i\mathbf{t}_i\rangle^T\widetilde{\mathbf{W}}^T(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}}) + \mathrm{tr}\left(\widetilde{\mathbf{W}}^T\widetilde{\mathbf{W}}\left\langle u_i\mathbf{t}_i\mathbf{t}_i^T\right\rangle\right)\right\}. \quad (24)$$

The MLE of $\widetilde{\nu}$ can be computed by solving the following equation by line search:

$$1 + \log\frac{\nu}{2} - \psi\left(\frac{\nu}{2}\right) + \frac{1}{N}\sum_{i=1}^{N}\left(\langle\log u_i\rangle - \langle u_i\rangle\right) = 0. \quad (25)$$

4.2.2. Normal Traffic Modeling with Data Loss. If data loss occurs on some dimensions, D-dimensional data $\mathbf{x}_i$ can still be used to conduct estimation for model parameters. The sample data $\mathbf{x}_i$ can be divided into observed data and missing data; that is,

$$\mathbf{x}_i = \begin{bmatrix}\mathbf{x}_i^o\\ \mathbf{x}_i^m\end{bmatrix}. \quad (26)$$

Its mean and covariance can be also divided into blocks:

$$\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}^o\\ \boldsymbol{\mu}^m\end{bmatrix};$$
$$\boldsymbol{\Psi} = \begin{bmatrix}\boldsymbol{\Psi}_{oo}\boldsymbol{\Psi}_{om}\\ \boldsymbol{\Psi}_{mo}\boldsymbol{\Psi}_{mm}\end{bmatrix}. \quad (27)$$

As noted in [31], $\mathbf{x}_i^m \mid \mathbf{x}_i^o \sim t_D(\boldsymbol{\mu}^m + \boldsymbol{\Psi}_{mo}\boldsymbol{\Psi}_{oo}^{-1}(\mathbf{x}_i^o - \boldsymbol{\mu}^o), \boldsymbol{\Psi}_{mm} - \boldsymbol{\Psi}_{mo}\boldsymbol{\Psi}_{oo}^{-1}\boldsymbol{\Psi}_{om}, \nu)$, and then $\mathbf{x}_i \mid \mathbf{x}_i^o \sim t_D(\mathbf{z}_i, \mathbf{Q}_i, \nu)$.

Thus

$$\mathbf{z}_i = \begin{bmatrix}\mathbf{x}_i^o\\ \boldsymbol{\mu}^m + \boldsymbol{\Psi}_{mo}\boldsymbol{\Psi}_{oo}^{-1}(\mathbf{x}_i^o - \boldsymbol{\mu}^o)\end{bmatrix};$$
$$\mathbf{Q}_i = \begin{bmatrix}0 & 0\\ 0 & \left(\boldsymbol{\Psi}_{mm} - \boldsymbol{\Psi}_{mo}\boldsymbol{\Psi}_{oo}^{-1}\boldsymbol{\Psi}_{om}\right)\end{bmatrix}. \quad (28)$$

The procedures for normal traffic model establishment with data loss are shown in Figure 5. When calculating model parameters, the method of determining d is the same as Section 4.2.1, and REM is also used to implement estimations of model parameters $\{\boldsymbol{\mu}, \tau, \mathbf{W}, \nu\}$ under the condition of data loss.

The First Stage

E-Step. Consider

$$\left\langle L_1'\right\rangle = -\sum_{i=1}^{N}\mathrm{tr}\left(\left\langle u_i(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T\right\rangle\boldsymbol{\Psi}^{-1}\right), \quad (29)$$

where

$$\langle u_i\rangle = \frac{\nu + D_i^o}{\nu + (\mathbf{x}_i^o - \boldsymbol{\mu}^o)^T\boldsymbol{\Psi}_{oo}^{-1}(\mathbf{x}_i^o - \boldsymbol{\mu}^o)}, \quad (30)$$

$$\left\langle u_i(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T\right\rangle = \mathbf{Q}_i + \langle u_i\rangle(\mathbf{z}_i - \boldsymbol{\mu})(\mathbf{z}_i - \boldsymbol{\mu})^T. \quad (31)$$

M-Step. Maximization of $\langle L_1'\rangle$ with respect to $\boldsymbol{\mu}$ results in the updating formula of $\boldsymbol{\mu}$:

$$\widetilde{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{N}\langle u_i\rangle\mathbf{z}_i}{\sum_{i=1}^{N}\langle u_i\rangle}. \quad (32)$$

The Second Stage. In the second stage, the latent variable $\mathbf{t}_i$ is introduced, and then the algorithm estimates parameters $\{\tau, \mathbf{W}, \nu\}$. The estimated value $\widetilde{\boldsymbol{\mu}}$ from the first stage is used in the second stage.

E-Step. Consider

$$\left\langle L_2'\right\rangle = -\sum_{i=1}^{N}\left\{\frac{D}{2}\ln\tau + \frac{1}{2\tau}\mathrm{tr}\left[\left\langle u_i(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})^T\right\rangle\right] - \frac{1}{\tau}\mathrm{tr}\left[\left\langle u_i(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})\mathbf{t}_i^T\right\rangle\mathbf{W}^T\right] + \frac{1}{2\tau}\mathrm{tr}\left(\mathbf{W}^T\mathbf{W}\left\langle u_i\mathbf{t}_i\mathbf{t}_i^T\right\rangle\right)\right\}, \quad (33)$$

where $\langle u_i(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})^T\rangle$ is in (31). Consider

$$\begin{aligned}\left\langle u_i\left(\mathbf{x}_i-\widetilde{\boldsymbol{\mu}}\right)\mathbf{t}_i^T\right\rangle &= \left\langle u_i\left(\mathbf{x}_i-\widetilde{\boldsymbol{\mu}}\right)\left(\mathbf{x}_i-\widetilde{\boldsymbol{\mu}}\right)^T\right\rangle \mathbf{W}\mathbf{M}^{-1},\\ \left\langle u_i\mathbf{t}_i\mathbf{t}_i^T\right\rangle & \\ &= \tau\mathbf{M}^{-1}+\mathbf{M}^{-1}\mathbf{W}^T\left\langle u_i\left(\mathbf{x}_i-\widetilde{\boldsymbol{\mu}}\right)\left(\mathbf{x}_i-\widetilde{\boldsymbol{\mu}}\right)^T\right\rangle \mathbf{W}\mathbf{M}^{-1}.\end{aligned} \tag{34}$$

M-Step. Maximizing $\langle L_2\rangle$ with respect to $\mathbf{W}$ and τ gives the following updating formulas of $\{\mathbf{W},\tau\}$:

$$\begin{aligned}\widetilde{\mathbf{W}} &= \left(\sum_{i=1}^{N}\left\langle u_i\left(\mathbf{x}_i-\widetilde{\boldsymbol{\mu}}\right)\mathbf{t}_i^T\right\rangle\right)\left(\sum_{i=1}^{N}\left\langle u_i\mathbf{t}_i\mathbf{t}_i^T\right\rangle\right)^{-1},\\ \widetilde{\tau} &= \frac{1}{ND}\sum_{i=1}^{N}\left\{\operatorname{tr}\left[\left\langle u_i\left(\mathbf{x}_i-\widetilde{\boldsymbol{\mu}}\right)\left(\mathbf{x}_i-\widetilde{\boldsymbol{\mu}}\right)^T\right\rangle\right]\right.\\ &\quad -2\operatorname{tr}\left[\left\langle u_i\left(\mathbf{x}_i-\widetilde{\boldsymbol{\mu}}\right)\mathbf{t}_i^T\right\rangle\mathbf{W}^T\right]\\ &\quad \left.+\operatorname{tr}\left(\widetilde{\mathbf{W}}^T\widetilde{\mathbf{W}}\left\langle u_i\mathbf{t}_i\mathbf{t}_i^T\right\rangle\right)\right\}.\end{aligned} \tag{35}$$

The updating formula of v is the same as (25).

The algorithm in Section 4.2.2 is applicable when data loss occurs as well as noise interference.

4.3. Anomaly Detection. Anomaly traffic flow samples of complex traffic flow data need to be determined by choosing measurement standards. There are mainly 2 strategies for determining anomaly samples: one is to determine whether samples are leverage outliers by judging if Hotelling's T^2 exceeds the threshold; the other is to determine whether samples are orthogonal outliers by judging squared prediction error (SPE) which exceeds the threshold [32]. Because probabilistic model is established in this paper and Mahalanobis distance is employed to conduct anomaly measurement, there is no need for the two strategies [33].

For intact data samples, the squared Mahalanobis distance δ^2 is

$$\delta^2 = \left(\mathbf{x}_i-\boldsymbol{\mu}\right)^T\boldsymbol{\Psi}^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}\right). \tag{36}$$

For samples containing some of dimension loss, the squared Mahalanobis distance δ_m^2 is to be computed using its expectation:

$$\delta_m^2 = \operatorname{tr}\left\{\boldsymbol{\Psi}^{-1}\left[\left(\mathbf{z}_i-\boldsymbol{\mu}\right)\left(\mathbf{z}_i-\boldsymbol{\mu}\right)^T+\mathbf{Q}_i\right]\right\}. \tag{37}$$

Normal distribution "3σ" control chart is adopted to determine anomaly, and the choosing reason is discussed in Section 6.2. When anomaly occurs in time series, corresponding distribution goes beyond the control boundary.

Establishing time series of the squared Mahalanobis distance as $\delta^2(t)$, its mean is μ_M, and the variance is σ_M^2, and then the configuration of "3σ" control chart is as follows:

Centre line: CL $= \mu_M$.

Upper control limit: UCL $= \mu_M + 3\sigma_M$.

Lower control limit: LCL $= \mu_M - 3\sigma_M$.

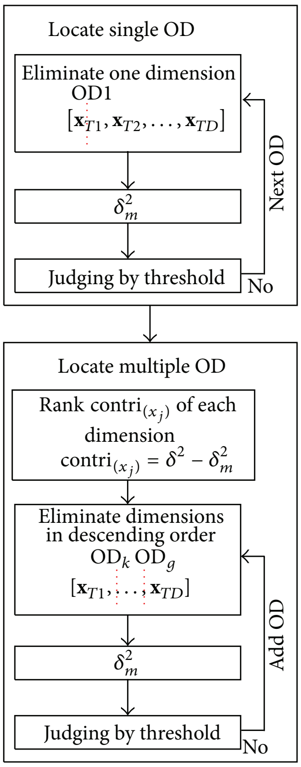

Locate single OD
Eliminate one dimension
OD1
$[\mathbf{x}_{T1}, \mathbf{x}_{T2}, \ldots, \mathbf{x}_{TD}]$
δ_m^2
Judging by threshold
No
Next OD
Locate multiple OD
Rank contri$_{(x_j)}$ of each dimension
contri$_{(x_j)} = \delta^2 - \delta_m^2$
Eliminate dimensions in descending order
OD$_k$ OD$_g$
$[\mathbf{x}_{T1}, \ldots, \mathbf{x}_{TD}]$
δ_m^2
Judging by threshold
No
Add OD

Figure 6: Steps of locating OD of anomaly occurrence.

According to the Gaussian distribution,

$$\begin{aligned}&P\left\{-3\sigma_M < \delta_M^2(t) < 3\sigma_M\right\}\\ &\quad = \Phi\left(\frac{3\sigma_M-\mu_M}{\sigma_M}\right)-\Phi\left(\frac{-3\sigma_M-\mu_M}{\sigma_M}\right) = 99.74\%,\end{aligned} \tag{38}$$

where $\Phi(\cdot)$ is the probability distribution function of standard normal distribution.

Adopting normal distribution "3σ" control chart, it is certain that anomaly occurs when the values deviate from μ_M by more than 3 times standard deviation σ_M, and its confidence coefficient is 99.74%.

4.4. Anomaly Localization. After confirming anomalous samples, it is needed to be determined which dimension (corresponding to OD) of the selected anomalous sample should be responsible for the anomaly, which is anomaly localization. The anomalous sample $\mathbf{x}_a$ is a D-dimensional vector, and x_j $(j = 1, \ldots, D)$ is the jth-dimensional variable of $\mathbf{x}_a$, adopting following contribution analysis (as shown in Figure 6) to locate OD of anomaly occurrence, which is divided into 2 stages.

Stage One. (1) Eliminate any x_j in $\mathbf{x}_a$, and get $\mathbf{x}_a^o$ and $\mathbf{x}_a^m$.

(2) Compute the conditional distribution of the selected anomalous sample with the jth-dimensional variable missing

Table 1: Execution time of RMPCM anomaly detection.

Simulation experiment	Testbed experiment		Real network data analysis	
Complete data	Complete data	Missing data (mean)	Complete data	Missing data (mean)
2.15 s	1.94 s	3.56 s	2.53 s	4.86 s

Table 2: Evaluation content and method.

<table>
<tr><th rowspan="3">Method</th><th colspan="6">Content</th></tr>
<tr><th colspan="2">Robustness under noise interference</th><th rowspan="2">Robustness under data loss</th><th rowspan="2">Anomaly localization</th><th colspan="2">Sensitivity</th></tr>
<tr><th>Noisy traffic</th><th>Poisoning</th><th>Intrinsic dimension</th><th>Traffic measure</th></tr>
<tr><td>Simulation experiment</td><td></td><td>√</td><td></td><td>√</td><td></td><td></td></tr>
<tr><td>Testbed experiment</td><td>√</td><td></td><td>√</td><td>√</td><td></td><td></td></tr>
<tr><td>Real network data analysis</td><td></td><td></td><td>√</td><td></td><td colspan="2">√</td></tr>
</table>

from formula (28) to derive corresponding $\mathbf{z}$ and $\mathbf{Q}$, and compute the squared Mahalanobis distance δ_m^2 with formula (37).

(3) Calculate the contribution of x_j using $\text{contri}_{(x_j)} = \delta^2 - \delta_m^2$. It represents the degree of change in anomaly measurement after a certain dimension of anomaly sample missing.

(4) Take each different value of j $(j = 1, \ldots, D)$, and repeat (1) ~ (3). If δ_m^2 is smaller than the threshold of anomaly discriminant, the corresponding dimension $j(\text{OD}_j)$ is mainly responsible for the anomaly, since its elimination would bring the sample back to normal.

Stage Two. If there is not any δ_m^2 in Step (4) of Stage 1 smaller than the threshold of anomaly discriminant, it is indicated that the anomaly is not caused by anomalous traffic of only one OD flow but multiple OD, and then Stage 2 is to be carried out.

(1) Arrange $\text{contri}_{(x_j)}$ of every dimension of $\mathbf{x}_a$ in descending order.

(2) Remove the first two dimensions of $\mathbf{x}_a$ after the arrangement in (1), and if δ_m^2 is smaller than the threshold, the corresponding two dimensions are responsible for the anomaly.

(3) Otherwise, sequentially increase the number of missing dimensions in the order arranged in (1) until corresponding δ_m^2 is smaller than the threshold, and then several missing dimensions are jointly responsible for the anomaly.

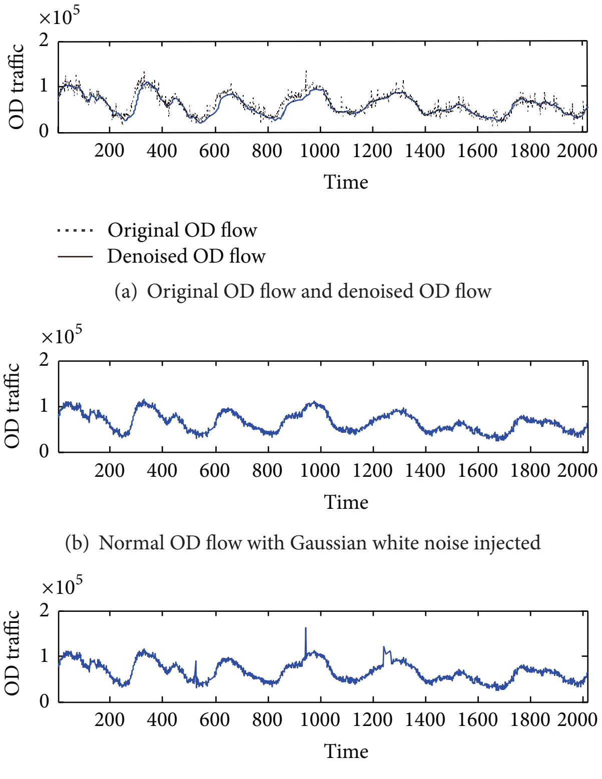
4.5. Algorithm Complexity Analysis. In RMPCM the major overheads are the inverse of $\mathbf{\Psi}$ and iterations of REM. $\mathbf{\Psi}$ is $D \times D$ matrix; D is dimensions of $\mathbf{X}$ corresponding to the number of columns of traffic matrix, which is the number of OD $(n \times n)$. In the calculation process, directly computing $\mathbf{\Psi}^{-1}$ affects the algorithm complexity significantly, and Woodbury matrix identity is adopted in the paper: $\mathbf{\Psi}^{-1} = (\mathbf{W}\mathbf{W}^T + \tau\mathbf{I}_D)^{-1} = \tau^{-1}\mathbf{I}_D - \tau^{-1}\mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T$, where $\mathbf{M} = \mathbf{W}^T\mathbf{W} + \tau\mathbf{I}_d$ is $d \times d$ matrix. d is determined by applying PCA dimensionality reduction, and $d \ll D$ (refer to Section 6.1). Calculating the inverse of $D \times D$ matrix $\mathbf{\Psi}$ is changed into inverting $d \times d$ matrix $\mathbf{M}$, which simplifies the algorithm complexity noticeably, and the time complexity is $O(Nd^2)$. The time complexity is also related with the iterations of EM algorithm. Iterative times are smaller than 15 in the paper. Applying Matlab to run RMPCM anomaly detection algorithm on selected data, execution time is shown in Table 1, and computer specifications are Windows 7, i7 3.5 GHz CPU, and 4 GB RAM. In the process of implementation the calculation will be faster when converting Matlab code into binary executable program.

5. Evaluation

Normally there are 3 methods that can be applied to assess the performance of network anomaly detection algorithm: network traffic simulation experiment [10, 14–16], testbed experiment [19], and real network data analysis [9, 10, 20, 32, 34]. There are pros and cons regarding each method. For example, network traffic simulation experiment produces synthetic data which can be entirely controlled by researchers, but it is not so close to reality; real network data analysis is closest to reality but it is difficult to establish a benchmark; testbed experiment gives attention to both reality and controllability, but the real output data might also deviate from the experiment settings. In order to assess the performance of RMPCM more objectively, all the three of them were combined to conduct analysis. Experimental content and method are shown in Table 2.

RMPCM will be compared with the anomaly detection method based on subspace construction via PCA and its improved method ANTIDOTE [15] to assess its real performance. As it is known, the anomaly detection method based on PCA is accepted and applied commercially (a commercial anomaly detection system (NetReflex by Guavus) is based on the well-known anomaly detector using PCA [34]). Therefore, the results should be persuasive.

5.1. Network Traffic Simulation Experiment. Network anomalous traffic, especially some poisoning attack traffic, may cause the skewing of detection model, which would significantly decrease the performance of anomaly detectors



(a) Original OD flow and denoised OD flow

(b) Normal OD flow with Gaussian white noise injected

(c) OD flow with anomalies injected

Figure 7: Steps for synthetic generation of anomalies.

[14, 22]. In this section the performance of RMPCM under poisoning will be assessed.

5.1.1. Anomaly Generation. The structure of network-wide traffic was revealed for the first time by Lakhina et al., which is that OD flows consist of large number of periodic and deterministic trends, some of the noises, and few of spikes [8]. In order to be closer to the reality, the real network data as shown in Table 1 is chosen as the blueprint of simulation, and Discrete Wavelet Transform is adopted to obtain periodic and deterministic trends of the real data which are the normal daily mode by eliminating noises and spikes [10]. After that, artificial noises and anomalies are synthetically injected into the periodic daily mode components. The process is as follows.

(1) Daubechies 5 orthogonal wavelet is adopted for multiresolution analysis, and high-frequency components containing noise and anomalies are removed by filtration from 121 OD flows' data in the traffic matrix, which computes the periodic smoothing components as the basis, as shown in Figure 7(a).

(2) Zero-mean Gaussian white noise is injected after (1), as shown in Figure 7(b).

(3) Typical anomalies are injected after (1) and (2), as shown in Figure 7(c).

121 OD flows are produced in this way. Data collection occurs every 5 minutes, which is marked as one collection cycle. Traffic matrix with 2016 rows and 121 columns is created with one week's collections in 121 OD flows.

Table 3: Typical anomalies in the Internet.

Type	Description
DoS	Single source node sending large amount of data to single destination node
DDoS	Multiple source nodes sending large amount of data to single destination node
ALPHA	Abnormal high speed rate transferring between two nodes
Ingress/egress shift	Change of routing causing traffic ingress/egress shift
Flash crowd	Abnormal large data request for a certain service

Table 4: Anomaly injection.

Type	Injection method
DoS/DDoS	Increasing the volume of single/multiple OD flows gradually
ALPHA	Promptly increasing the volume of a single OD flow
Ingress/egress shift	Reducing a portion of volume of a certain OD flow which then added to another OD flow
Flash crowd	Increasing the volume of multiple OD flows rapidly and then tuning them back to normal gradually

As we mainly focused on traffic volume anomalies, five kinds of typical anomalies were simulated: DoS, DDoS, ALPHA, ingress/egress shift, and flash crowd. Their brief description is shown in Table 3. Table 4 indicates how we have them injected.

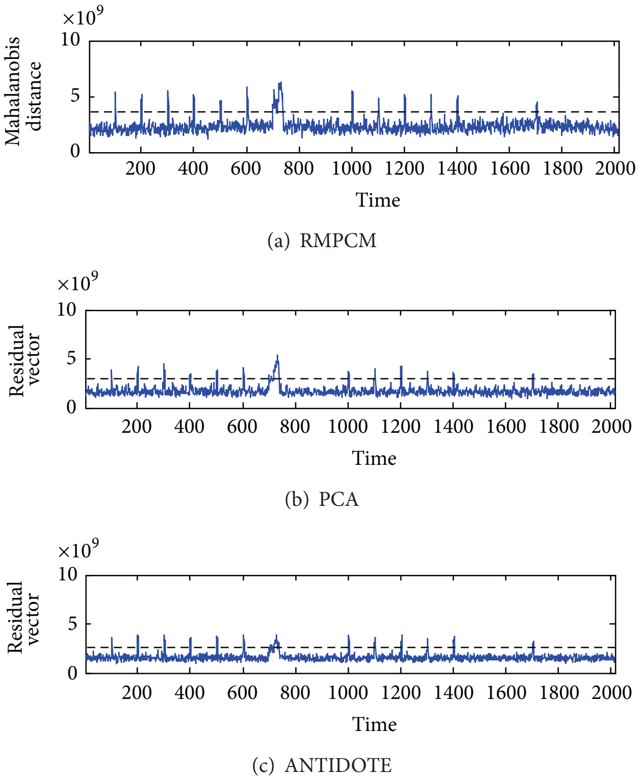
5.1.2. Poisoning Generation. The poisoning method Add-More-If-Bigger discussed in [15] was employed in this section. It suggests injecting poisoning traffic flow into targeted OD flow to increase its variance, which skews the detector's normal traffic model and thereby increases the escape probability of anomalies. This poisoning method requires attackers to grasp the real-time traffic volume of background flows. Add-More-If-Bigger means attackers inject poisoning traffic when the background traffic volume exceeds its mean value. The volume of poisoning traffic is $c_t = (\max\{0, x_s(t) - \text{mean}\})^{\beta}$, where $\text{mean} = (1/T)\sum_{t=1}^{T} x_s(t)$, $x_s(t)$ is the volume of targeted OD traffic at the moment of t. c_t is a function with parameter β which determines the degree of poisoning. In this experiment $\beta = 0.8$ means medium poisoning, and $\beta = 1.2$ means high poisoning.

5.1.3. Anomaly Detection under Poisoning. Anomalies were injected according to the method in Section 5.1.1. Different anomalies usually have different characteristics; thus injection should also be used in a different way. Usually DoS or DDoS attacks are the gradually increasing traffic volume of single or multiple OD flows, which normally lasts 5 to 30 minutes; ALPHA is always the sharp rise of single OD flow with uncertain duration; flash crowd occurs when traffic volume of multiple OD flows uprises fast and goes back to normal gradually; ingress/egress shift is likely to show a step change of 2 OD flows' volume, which normally lasts a long time until the routing policy change. According to their own properties, 5 ALPHA attacks were injected in 100~500 with its volume accounting for 50% of the mean volume of OD flows; each of them lasted 20 minutes; at 600, an hour's flash crowd anomaly was injected into OD flows from 3 origin nodes to one destination node, the volume of which was 10% ~40% of the mean volume of OD flows; at the time of 700, 200 minutes' ingress/egress shift were injected, and the volume of transferred traffic was 50% of the mean of OD flows; 5 Dos attacks were injected in 1000~1400 with its volume accounting for 30% ~50% of the mean volume of OD flows, and every one of them lasted 30 minutes; at the time of 1700, 30 minutes' DDoS attacks were injected into OD flows from 4 origin nodes to 1 destination node, and the volume of the anomalies is from 5% to 30% of the mean volume of OD flows.

Subspace construction via PCA and ANTIDOTE was selected in order to be compared with RMPCM under exactly the same circumstance to evaluate its detection performance under poisoning, and we plot the squared norm of the residual vector as a function of time to show PCA and ANTIDOTE method results. Receiver Operations Characteristic (ROC) curve is also applied to the overall performance estimation of three above methods. The x axis means false positive rate (FPR), and y axis represents true positive rate (TPR). ROC curve can display the relationship between TPR and FPR with the change of thresholds. The area under the curve is the performance indicator. The larger the area is, the better the performance is.

The experiment was conducted with the data generated in one week to compare those three methods. Three ROC curves are drawn corresponding to no poisoning, medium poisoning, and high poisoning (Figure 10), and part of the experiment results is shown in Figures 8 and 9. Before poisoning (Figure 8) these three methods have a similar performance. After medium poisoning (Figure 9) the volume of the residual vector of background traffic increases significantly in PCA method, and the performance of PCA reduces dramatically because lots of anomalies submerge in the background traffic (Figure 9(b)); ANTIDOTE performs well against poisoning, but low volume anomalies could easily escape its detection (Figure 9(c)); RMPCM has a stable performance, which does not decline too much after injection (Figure 9(a)). Figure 10 reveals that poisoning has a large impact on PCA anomaly detector, and FPR increases quickly after TPR exceeded 60%; RMPCM and ANTODOTE have a better antipoisoning ability; the detection performance of RMPCM is better than ANTIDOTE because its ROC curve rises up more quickly.

5.1.4. Anomaly Localization Test. Table 5 illustrates the setting of anomaly locating test, and a single OD or multiple OD flows were injected with anomalies generated in Section 5.1.3. The result of anomaly localization test in RMPCM method is shown in Figure 11, the histograms represent anomaly contribution degree of each OD at the time of anomaly

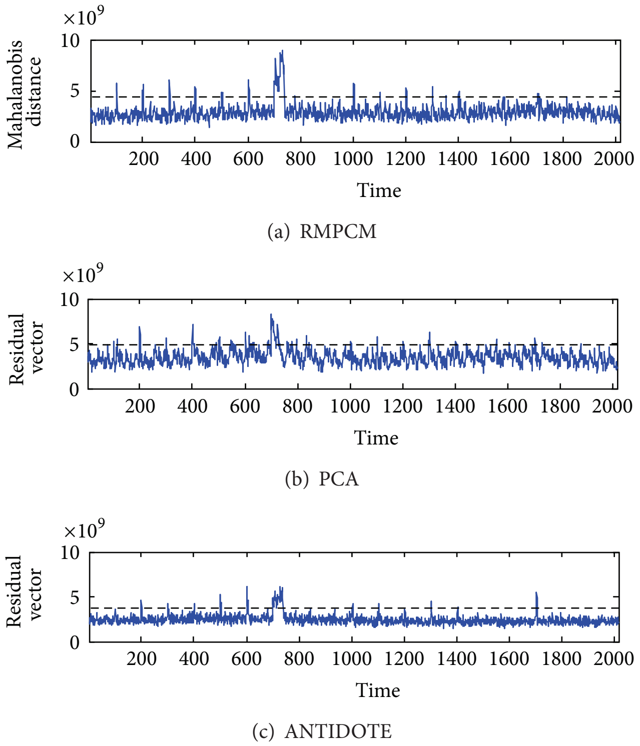


(a) RMPCM

(b) PCA

(c) ANTIDOTE

Figure 8: Comparison of three methods' test results with no poisoning.



(a) RMPCM

(b) PCA

(c) ANTIDOTE

Figure 9: Comparison of three methods' test results with medium poisoning.

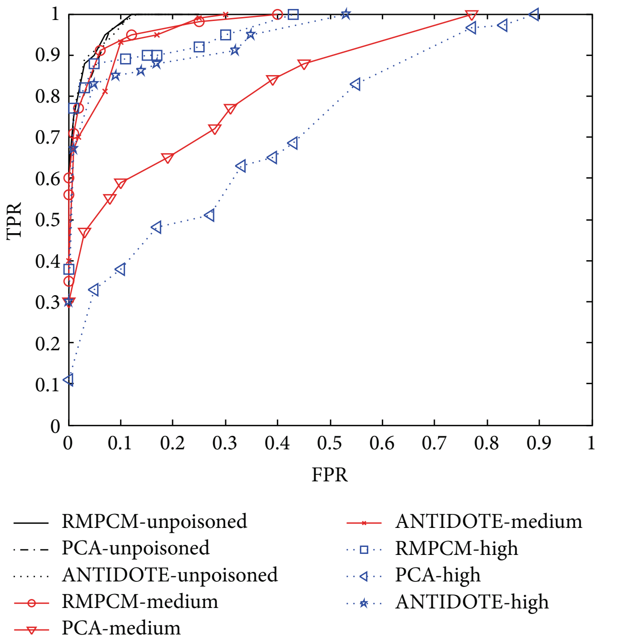


Figure 10: ROC curve of three anomaly detectors under poisoning.

Table 5: Setting of anomaly localization test.

Time of anomaly occurrence	Injection position	Type
300	OD50	ALPHA
703	OD50, OD100	Ingress/egress shift
602	OD7, OD40, OD60	Flash crowd
1704	OD10, OD20, OD60, OD80	DDoS

occurrence, and the result agrees with the setting from Table 5. As shown in Figure 11(a), anomaly contribution degree of OD50 noticeably exceeds the contribution degree threshold; therefore it is certain that anomaly at the time of 300 occurs on OD50; from Figures 11(b), 11(c), and 11(d) it is found that there is no single OD contribution degree exceeding the threshold, and this requires locating multiple OD when anomalies occur.

5.2. Testbed Experiment. Cyber-defense technology experimental research laboratory testbed (DETERLab) [35] proposed by the University of Southern California is chosen to conduct the testbed experiment. Not only can DETERLab interconnect nodes in the prototype system with any topological structure, but also its configuration for experiment condition is flexible. It provides researchers with needed background traffic and attack traffic injection for conducting network attack-defense experiments and then deploys and evaluates feasible solutions. It can fully integrate local hardware resources and has a better simulation advantage over software such as NS2.

In our experiment attacking tools based on Metasploit frame was integrated into DETERLab's security experimentation environment (SEER) toolset to generate multiple anomalies on the DETERLab platform. Our experiment sets up 10 PoP nodes and chooses the adjacent node of each PoP node as collection device, and the topological configuration is shown in Figure 12. The duration of the experiment is one week, data are collected every 5 minutes which is a collection cycle, and there are total 2016 cycles for a week. The collecting data is byte count of traffic flows.

5.2.1. Anomaly Detection in Noisy Traffic. In order to verify RMPCM's performance in noisy traffic, the experiment set up three situations to compare with the detection method based on subspace construction via PCA, which test the detection accuracy, factors impacting performance, and poisoning of large anomalies, respectively, in the two methods.

At the time of 500 and 1000 DoS attacks using TCP SYN Flood were initiated from PoP1 to PoP2, at the time of 1800 DoS attacks were carried out from PoP3 to PoP4, and the duration of DoS attacks was all 4 cycles; at the time of 800 port scan was initiated from PoP1 to PoP2, PoP5, and PoP6 applying by Nmat, and the duration was 5 cycles; at the time of 1200 ingress/egress shift was initiated by transferring 50% of the traffic volume of the OD path (PoP1 to PoP2) to another (PoP7 to PoP8), and the duration was 40 cycles; at the time of 1500 DDoS attacks using UDP Flood on PoP10 were initiated from PoP2, PoP4, PoP5, and PoP8 simultaneously, and it lasted 6 cycles. Figure 13(a) and Table 6(a) illustrate the test results applying RMPCM and subspace construction via PCA. Both methods succeed in detecting 6 anomalies, but the latter does not detect anomalies every time in the corresponding cycles of anomaly occurrence, especially for the 1200–1239 ingress/egress shifts. The anomaly cycles detected by PCA are shorter than the settings, while RMPCM made a notable improvement over PCA.

In order to reveal the impacting factors and performance differences of two methods further, anomalies were adjusted. The volume of DoS attacks beginning at 500 and 1800 was reduced by 50%; the range of port scan beginning at 800 decreased to the range only from PoP1 to PoP2, and scanning frequency was deducted by 50% as well; the duration of ingress/egress shift beginning at 1200 was cut down by 20 cycles; DDoS attacks beginning at 1500 were narrowed down to the range from PoP2 and PoP4 to PoP10 and kept the same attack volume; DoS attacks beginning at 1000 remained unchanged. Figure 13(b) and Table 6(b) indicate that the change of volume and range of anomalies would influence the performance of both of the methods, and the impact of duration deduction is hardly noticeable. PCA gives us false positive alarms at total of 10 time points including 1272, 1407, and 1451, which is higher than RMPCM obviously.

The attack volume of DoS increased by 220% to produce a large anomaly, and other factors remained the same as the first test. A large anomaly could raise variance level of the path that it was on, and other anomalies causing small variance would be mistaken for a normal event in this situation. Figure 13(c) and Table 6(c) show that the large anomaly leads to the dramatic decrease in the detection

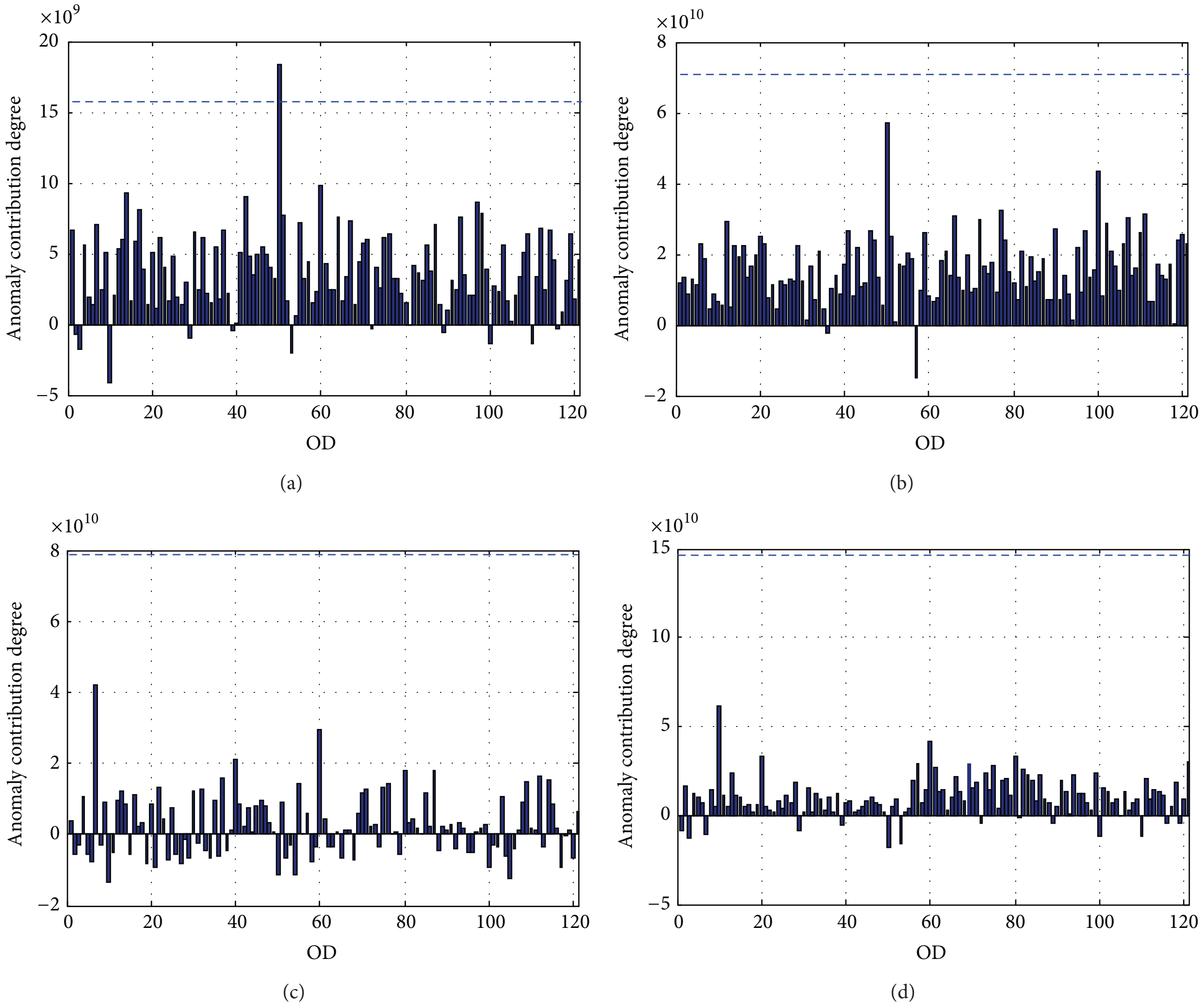


Figure 11: Results of locating OD of anomaly occurrence in simulation experiment.

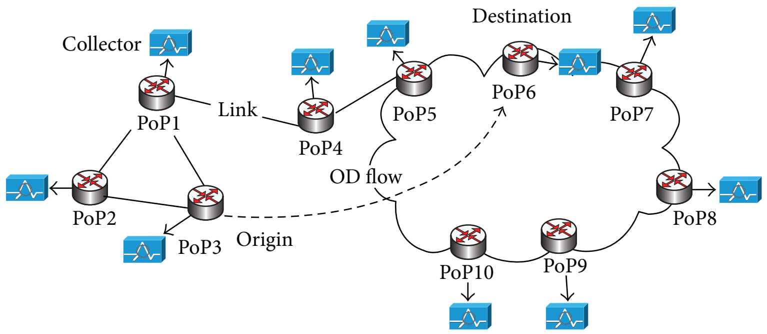
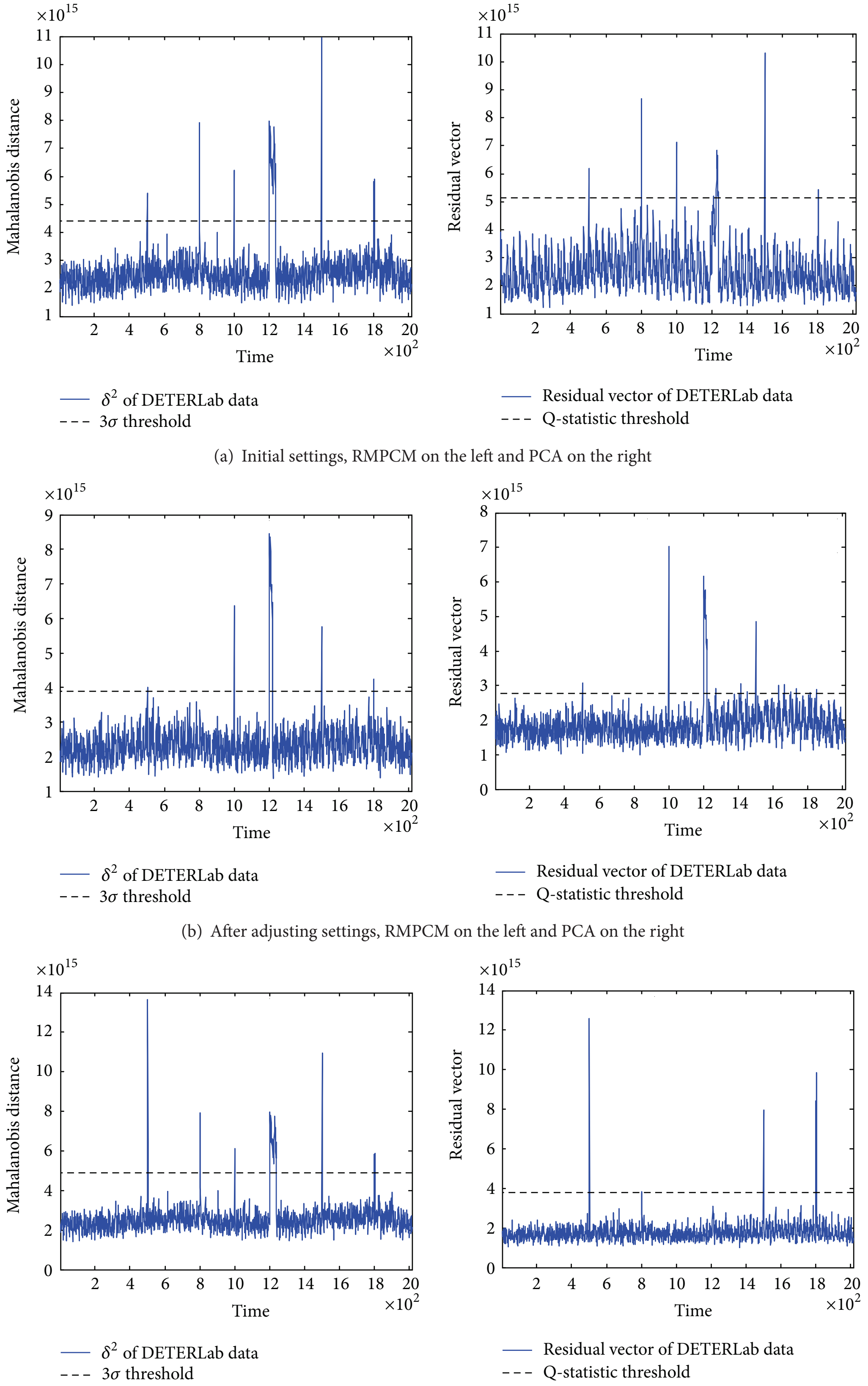


Figure 12: Topological configuration on DETERLab.

accuracy of PCA method. OD paths of anomaly occurrence contain the OD path of the large anomaly (PoP1 to PoP2), and then the corresponding anomaly detection rate is affected: Dos beginning at 1000 and the large anomaly are on the same path, and the Dos attack is completely missing from the figure; ingress/egress shift involves 2 OD paths which contain the OD path of the large anomaly, and the residual vector also seriously decays and does not reach the detection threshold; port scan and DDoS also contain OD path where the large anomaly occurs, and its residual vector is affected in varying degrees; DoS beginning at 1800 does not involve the OD path of the large anomaly, thus the detection to the above DoS attack is not impacted. Comparably RMPCM has a strong robustness, and detection accuracy is not affected.



(a) Initial settings, RMPCM on the left and PCA on the right

(b) After adjusting settings, RMPCM on the left and PCA on the right

(c) Injecting the large anomaly, RMPCM on the left and PCA on the right

FIGURE 13: Comparisons of RMPCM and PCA test results on DETERLab.

Table 6: Comparisons of RMPCM and PCA test results on DETERLab.

(a) Initial settings

Preset anomaly cycles	Alerts cycles		Type
	RMPCM	PCA	
500~503	501~503	502, 503	DoS
800~804	801	801	Port scan
1000~1003	1000~1003	1000~1003	DoS
1200~1239	1200~1239	1217, 1231–1239	Ingress/egress shift
1500~1505	1501~1505	1501~1505	DDoS
1800~1803	1802, 1803	1803	DoS

(b) After adjusting settings

Preset anomaly cycles	Alerts cycles		Type
	RMPCM	PCA	
500~503	503	502, 503	DoS
800~804			Port scan
1000~1003	1001, 1002	1001, 1002	DoS
1200~1219	1200~1219	1200~1219, 1272	Ingress/egress shift
1500~1505	1503, 1504	1407, 1416, 1451, 1503, 1504, 1637, 1665	DDoS
1800~1803	1802	1701, 1733, 1814, 1849	DoS

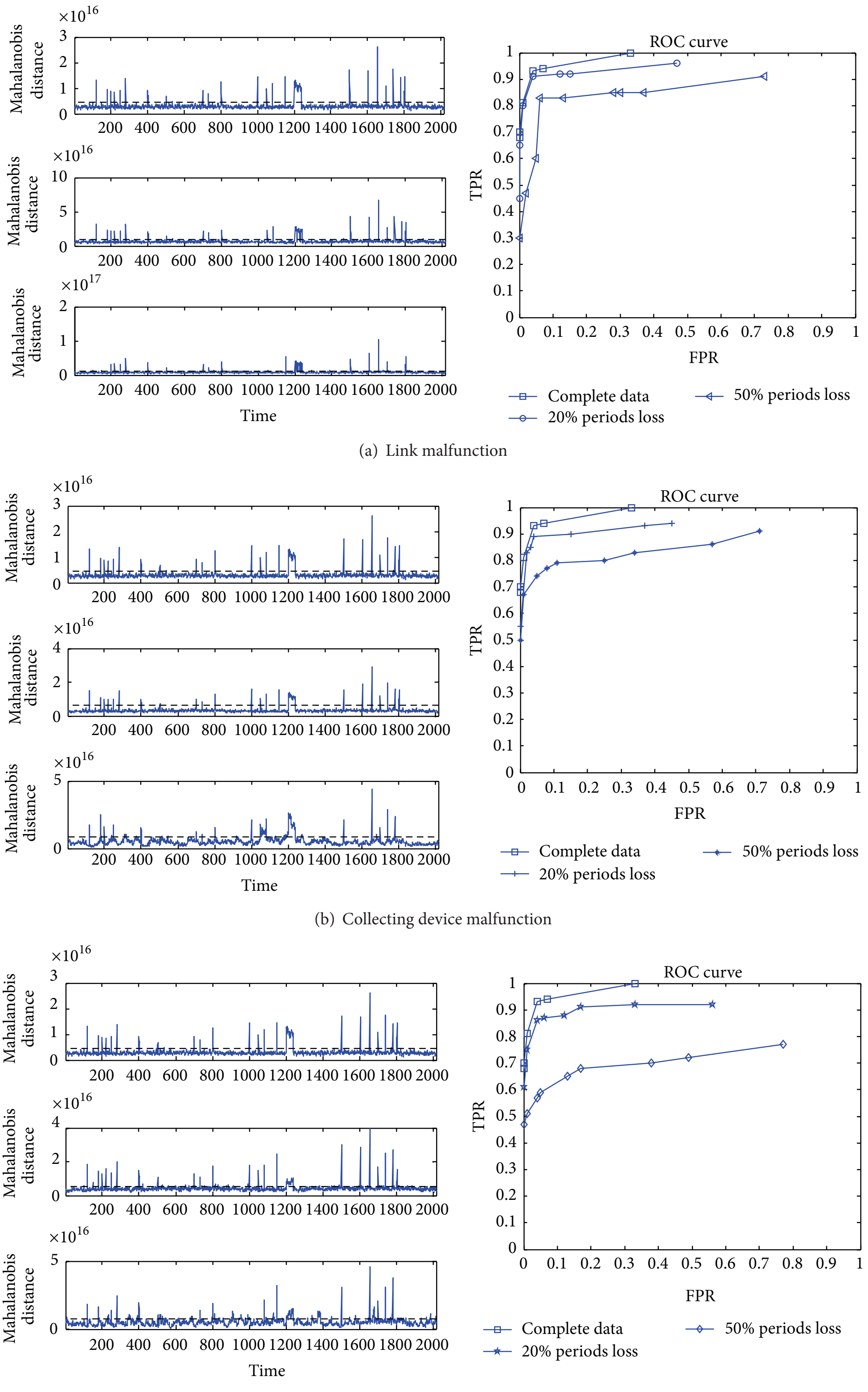
(c) Injecting the large anomaly

Preset anomaly cycles	Alerts cycles		Type
	RMPCM	PCA	
500~503	500~503	500~503	DoS
800~804	801	801	Port scan
1000~1003	1001, 1002		DoS
1200~1239	1200~1239		Ingress/egress shift
1500~1505	1501~1504	1502~1504	DDoS
1800~1803	1801, 1803	1801~1803	DoS

The above experiment demonstrates RMPCM has a better detection performance than the method based on subspace construction via PCA. Add-More-If-Bigger poisoning experiment in Section 5.1.3 and large anomaly poisoning experiment verify that RMPCM has a strong robustness under the poisoning of variance injection.

5.2.2. Anomaly Detection with Data Loss. Data loss may be caused by network malfunction, device breakdown, and so on; thus algorithms based on conventional nonstatistical model become inapplicable because of the incomplete input data. RMPCM based on the latent variable probability model of multivariate t-distribution has its distinct advantage when solving the problem of data loss. The detection performance of RMPCM with data loss is tested on DETERLab. The scenarios are set to the occurrence of data loss when facing link fault, collection device breakdown, and PoP node malfunction. Link fault will incur data loss when OD flows go through the link; collection device breakdown will cause the data loss of OD flows whose source node is the node connected with the broken-down device; data loss caused by PoP node malfunction is related with malfunction types, network topology, and router strategies. Figure 12 displays experimental topology configuration, and 100 anomalies were injected in the way that was used in Section 5.2.1. In order to eliminate occasionality 10 experiments were conducted for each type of malfunctions, 403 cycles and 1008 cycles were randomly selected (which accounted for 20% and 50% of all 2016 cycles in one week resp.), in which a randomly selected link (collection device, PoP node) broke down, and the malfunction duration lasted 20 cycles. The mean of 10 experiment results was used to draw the ROC curves. The experiments results indicate that link fault, collection device breakdown, and PoP node malfunction regularly and increasingly deteriorate the accuracy of anomaly detection, but overall RMPCM has better robustness in this situation. Its TPR is close to 70% when FPR is 20% even in the worst scenario (PoP node malfunction occurred in 50% cycles) (Figure 14(c)).

5.2.3. Anomaly Localization Test. Table 7 illustrates the setting of anomaly locating test, and a single OD or multiple OD flows were injected with anomalies generated in Section 5.2.1. The result of anomaly localization test in RMPCM method

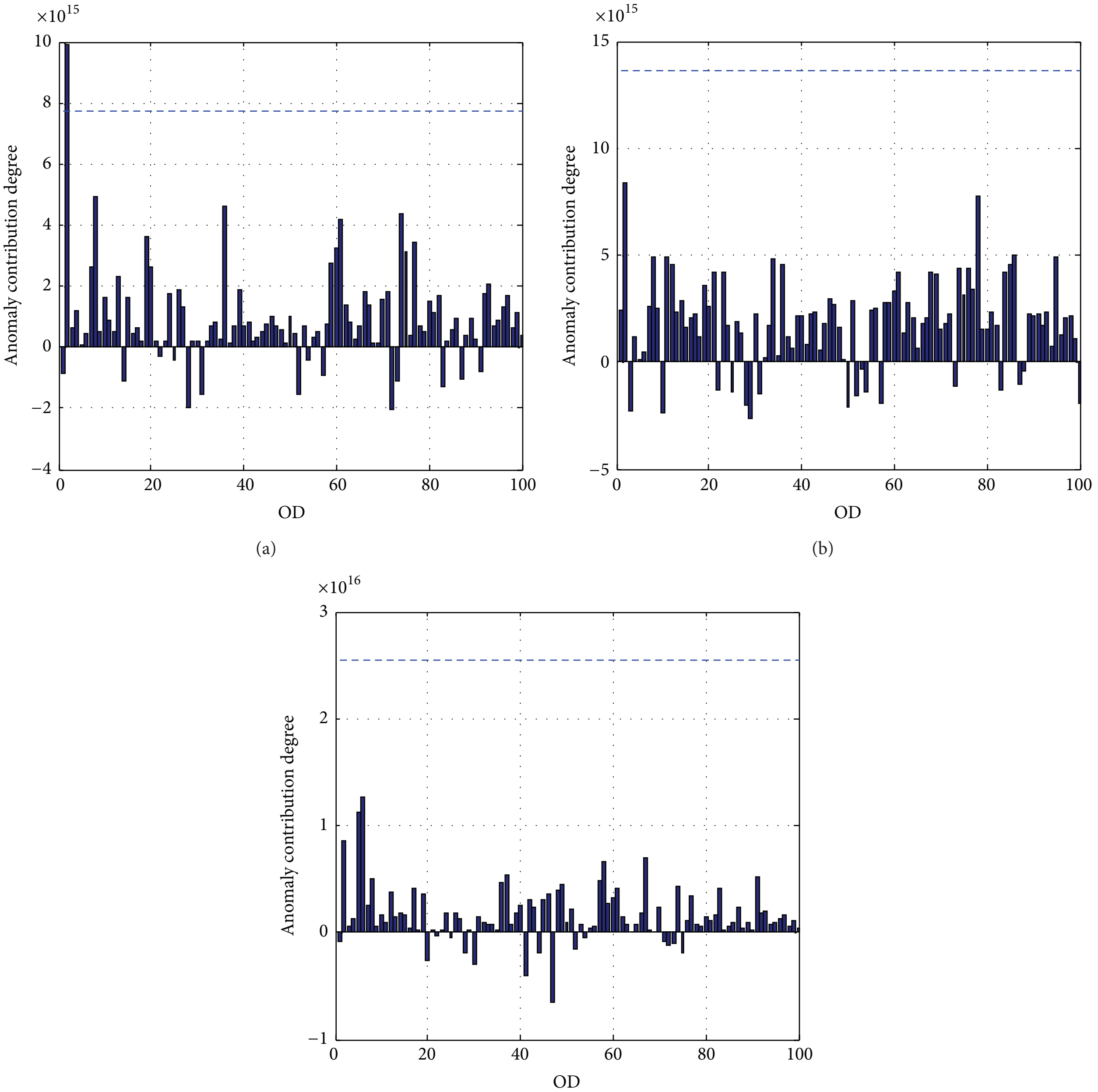


(a) Link malfunction

(b) Collecting device malfunction

(c) PoP node malfunction

Figure 14: RMPCM detection results (left) and ROC curves (right) with data loss on DETERLab.



(a)

(b)

(c)

Figure 15: Results of locating OD of anomaly occurrence on DETERLab.

Table 7: Setting of anomaly localization test.

Time of anomaly occurrence	Injection position	Type
502	OD2	DoS
1203	OD2, OD78	Ingress/egress shift
802	OD2, OD5, OD6	Port scan

is shown in Figure 15, the histograms represent anomaly contribution degree of each OD at the time of anomaly occurrence, and the results agree with the setting from Table 7. As shown in Figure 15(a), anomaly contribution degree of OD2 noticeably exceeds the contribution degree threshold; therefore it is certain that anomaly at the time of 502 occurs on OD2; from Figures 15(b) and 15(c) it is found that there is not a single OD contribution degree exceeding the threshold, and this requires locating multiple OD when anomalies occur. Some problems of anomaly localization will be discussed in Section 6.5.

5.3. Analysis of Real Network Data

5.3.1. Data Sets. Real network data sets are obtained from backbone network Abilene [2, 9, 10, 16, 18, 20], whose main

TABLE 8: Abilene traffic matrix data set.

Duration	Time bin	Measure	Matrix form	Data set
15 December–21 December 2003	5 min	Byte	2010 × 121	B
15 December–21 December 2003	5 min	Packet	2010 × 121	P
15 December–21 December 2003	5 min	Flow	2010 × 121	F

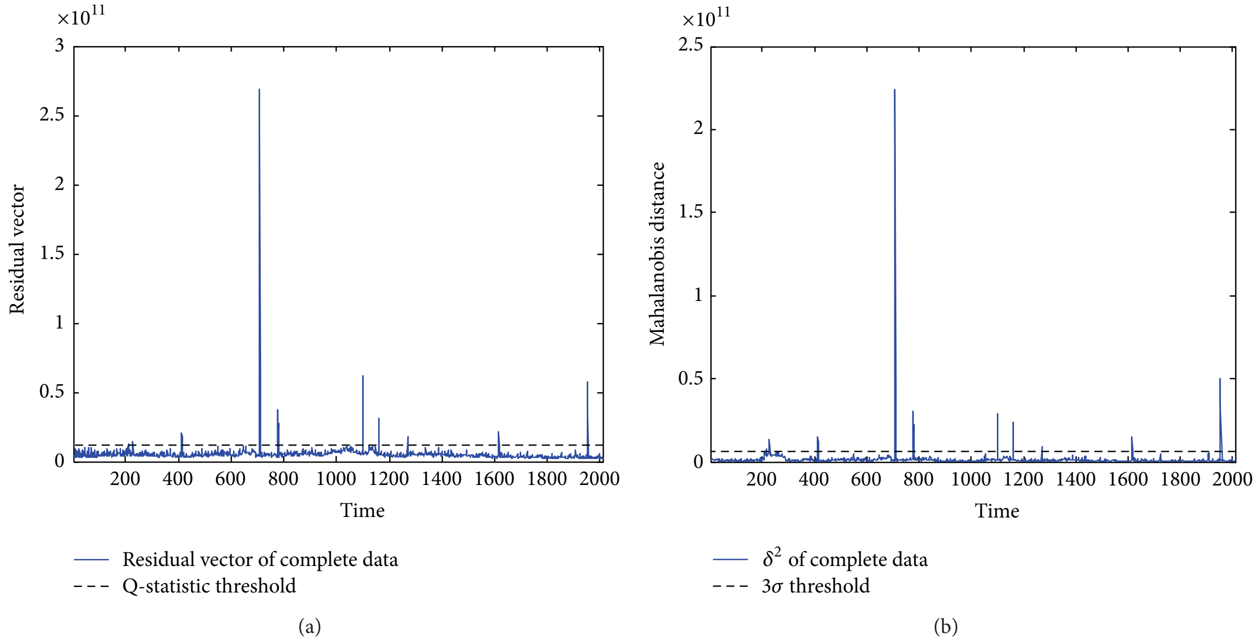


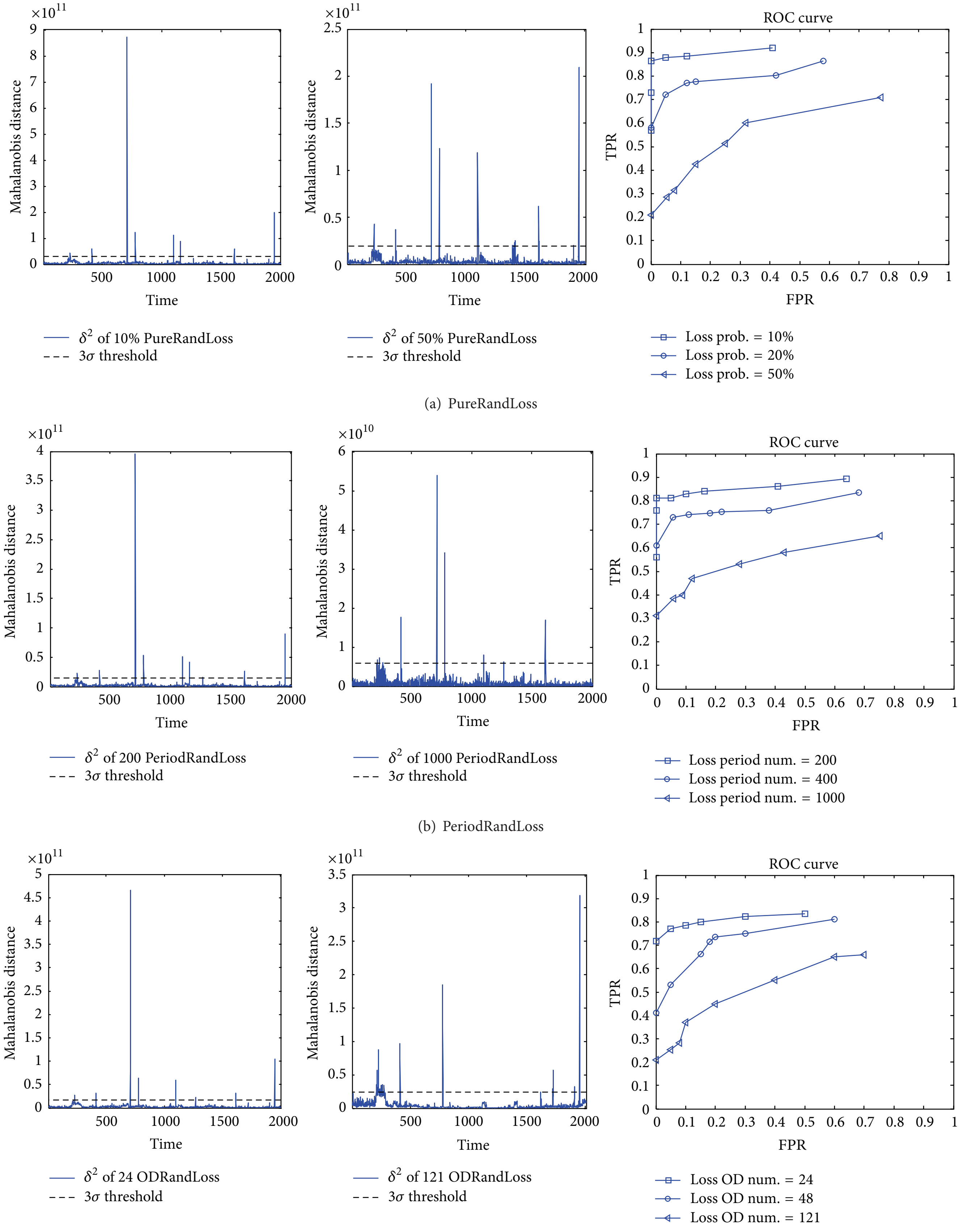
FIGURE 16: Test results of real network data with complete data, PCA (a) and RMPCM (b).

users are American universities and scientific institutions. We choose the data from Abilene data sets dating from 15 December 2003 to 21 December 2003 for its completeness and convenience for reference. The raw data is NetFlow data from 11 PoP nodes, the access point and exit point of each flow are obtained on the basis of BGP and ISIS routing table, and then the volume of OD flows and traffic matrix can be calculated, as shown in Table 8. Detection performance evaluation with data loss and sensitivity analysis were conducted with the data set.

5.3.2. Anomaly Detection with Data Loss. When conducting anomaly detection with data loss, data set P was chosen to implement four kinds of data loss mechanisms discussed in Section 3.2.1. In order to clearly identify the different performance of RMPCM in different data loss mechanisms and evaluate the factors influencing the performance, nearly the same loss rate was assigned to PureRandLoss, PeriodRandLoss, and ODRandLoss; the influence of different volume of missing piece on the detection performance was analyzed under PieceRandLoss. In order to eliminate accidents, 10 experiments were conducted, and mean values were calculated.

The true anomalies in the real network traffic data could be hardly obtained exactly, and RMPCM and PCA generated very similar alerts with the complete data set P (as shown in Figure 16), so the benchmark was set to be the detecting result of RMPCM with complete data in order to test the performance of RMPCM under the situations of data loss.

In the experiment of PureRandLoss, three kinds of loss rates were selected which accounted for 10%, 20%, and 50% of total data; in PeriodRandLoss the missing periods were set to be 200, 400, and 1000, as the total periods were 2010; data loss rates in here were close to 10%, 20%, and 50%; in ODRandLoss we chose half adjacent data of a certain column to be empty because of the algorithm limitation that column of traffic matrix could not be empty entirely, lost OD numbers were set to be 24, 48, and 121, and the loss rate per OD was 50%; therefore, the total loss rates of ODRandLoss were still 10%, 20%, and 50%. Parts of the results and ROC curves under three kinds of loss mechanisms, respectively, are displayed in Figures 17(a), 17(b), and 17(c). Three tests above indicate the relatively high accuracy of RMPCM: when the loss rate is 10%, the effect on anomaly detectors is slight, and the detection performance is very similar to that of RMPCM and PCA with complete data; as the loss rate increasing the performance begins to decline, even when the loss rate achieves 50%, test results are still applicable. The tests also demonstrate that ODRandLoss has the biggest impact on the performance of anomaly detectors, followed by PeriodRandLoss and PureRandLoss. The total data loss remains the same when applying these three methods, but the performance of anomaly detectors deteriorates regularly, and it is highly probable that this is due to the increase

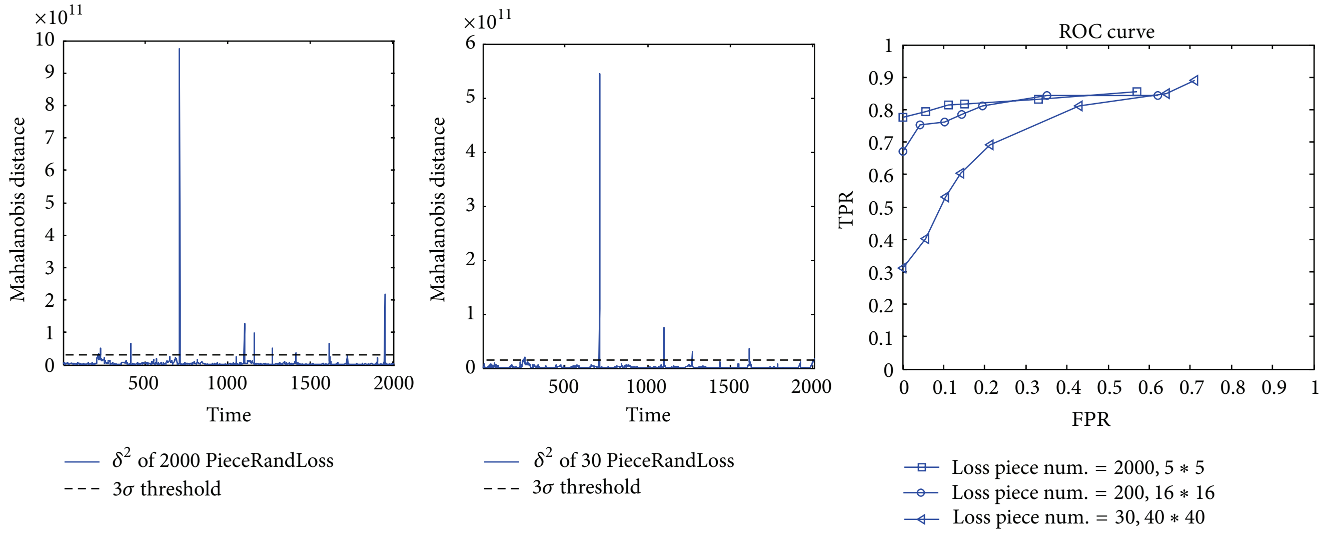


(a) PureRandLoss

(b) PeriodRandLoss

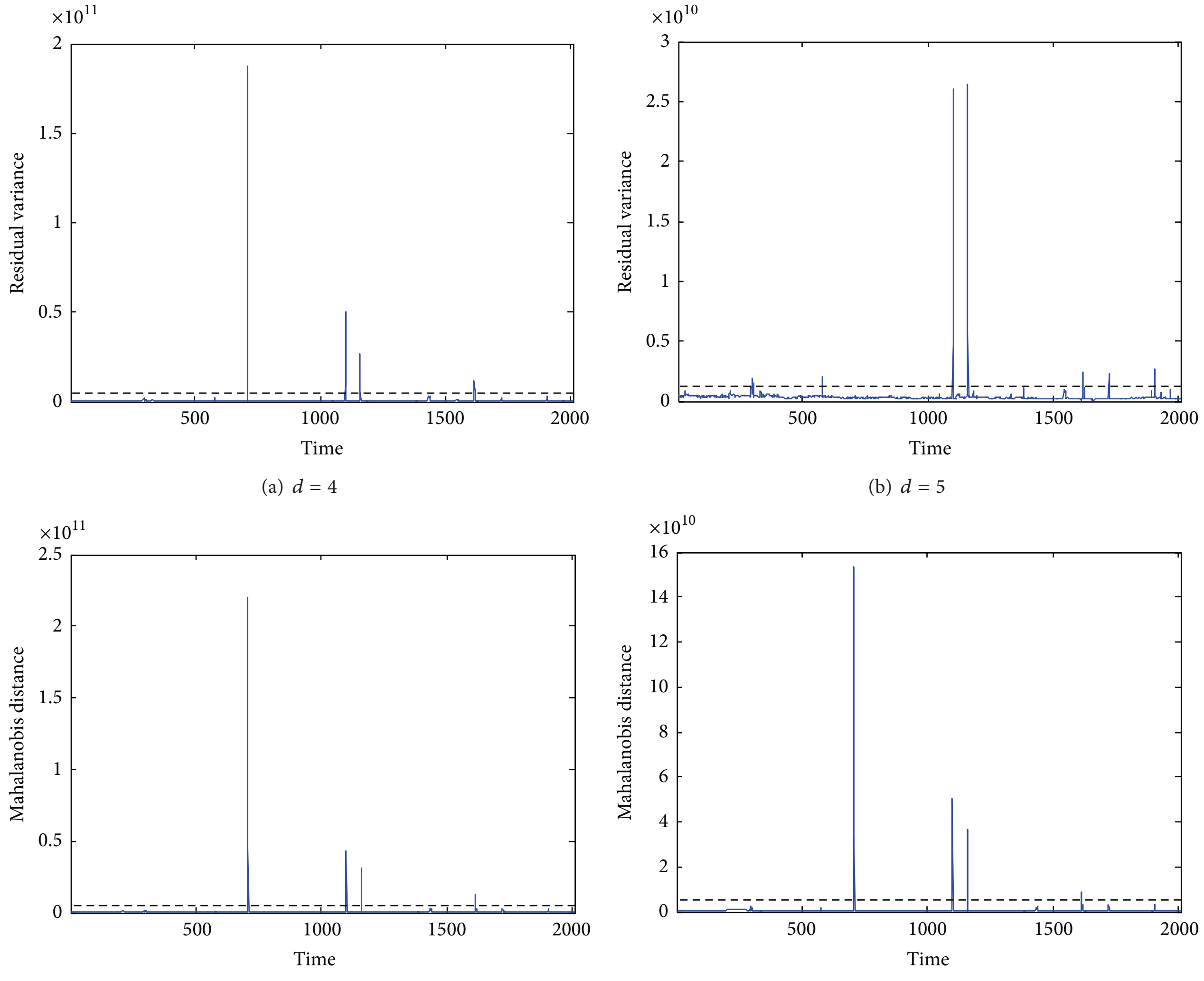
(c) ODRandLoss

Figure 17: Continued.



(d) PieceRandLoss

Figure 17: Test results of real network data under four loss mechanisms.



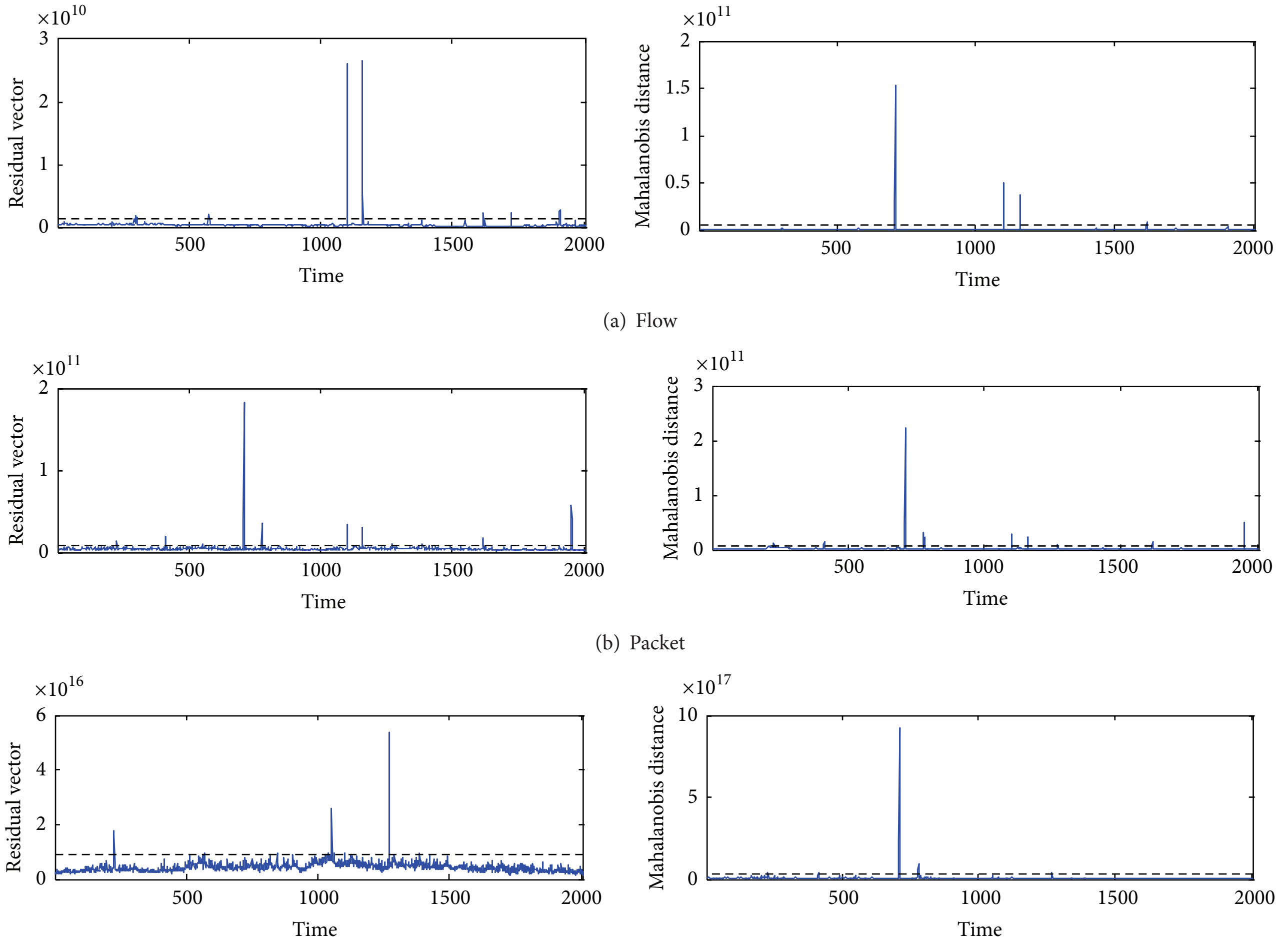
(a) $d = 4$

(b) $d = 5$

(c) $d = 4$

(d) $d = 5$

Figure 18: Comparison of sensitivity to the change of intrinsic dimensions of PCA (a, b) and RMPCM (c, d).



(a) Flow

(b) Packet

(c) Byte

Figure 19: Comparison of sensitivity to the change of traffic measures of PCA (left) and RMPCM (right).

of structured loss of data. The volume of missing piece in PureRandLoss is smallest, while volume of missing piece in ODRandLoss is largest.

In order to verify the impact of structured loss on the detection performance, we conducted experiment with PieceRandLoss, and the total volume of loss is set to be the same while the volume of each missing piece is set at different sizes, which are $5 * 5$, $16 * 16$, and $40 * 40$, and the numbers of missing pieces were 2000, 200, and 30, respectively; the volume of total loss accounts for 20% of whole data. Parts of the results and ROC curves in Figure 17(d) indicate that the performance of anomaly detector declines significantly when the volume of structured loss increases while the total loss rate remains the same. However, RMPCM achieves about 70% TPR with less than 20% FPR when applying PieceRandLoss overall, which means the performance could satisfy the requirement.

5.3.3. Sensitivity Analysis. The authors of [22] point out that the detection algorithm based on subspace construction via PCA is too sensitive to the change of intrinsic dimensions and traffic measures; therefore, the same issues should also be considered with RMPCM. The test will be conducted in two scenarios: one is to analyze the sensitivity to the change of intrinsic dimension d; the other is to analyze the sensitivity to the change of traffic measures. The results of RMPCM are to be compared with the algorithm based on subspace construction via PCA.

Real network data set F in Table 8 was selected to conduct the sensitivity to the change of intrinsic dimension analysis. Intrinsic dimension d varies depending on the accumulative variance contribution of principal components which is defined beforehand. The experiment results are displayed in Figure 18. Figures 18(a) and 18(b), respectively, represent detection curves when PCA model's intrinsic dimension d was set to be 4 and 5, as it can be seen that the detection results vary largely; Figures 18(c) and 18(d) represent corresponding curves of RMPCM, and the detection results agree with each other. The further procedures are taken to test d varying from 2 to 10 when using RMPCM, and test results are almost identical, which suggest that RMPCM is highly robust.

Real network data sets B, P, and F in Table 8 were employed to analyze the sensitivity to the change of traffic measures. The accumulative variance contribution of principal components was set to be the same value of 0.85 when using different traffic measures. The experiment results are shown in Figure 19. The detection results in three kinds of measures (B, P, F) are completely different by applying PCA. While using RMPCM the baselines of three curves are similar to each other, and three test results are highly correlative. Although benchmarks of anomalies in real network data sets could not be acquired, the statistical data in three kinds of measures (flow, packet, and byte) has interrelationship; therefore corresponding anomaly detection results should somewhat coincide. PCA produces too much conflict, which indicates it is much sensitive to the change of traffic measures, while RMPCM generates more satisfying results and shows a strong robustness.

To summarize, RMPCM is convenient for practical implementation for its lower sensibility and higher robustness to the change of parameters such as intrinsic dimensions and traffic measures.

6. Discussion

Issues like parameter selection, distribution characters of data source, and so forth will be discussed further in this section.

6.1. Intrinsic Dimension. The Internet traffic matrix has this characteristic of low dimensionality. The intrinsic dimension d of the matrix needs to be determined at the beginning in the process of modelling normal traffic by applying RMPCM. The scree plot of data sets of B, P, and F in Table 8 is drawn using PCA, as is shown in Figure 20. The accumulative variance contribution threshold is set to 0.85, and then intrinsic dimension of these three data sets are all close to 5, which is far from their dimension D ($D = 121$) of traffic matrix. We then go further to Abilene traffic data in the following three weeks and find that their intrinsic dimension d is considerably stable. What is more, the experiment in Section 5.3.3 also shows that RMPCM performance keeps high stability when d varies from 2 to 10, so we chose $d = 5$ in this paper. In fact, when implementing RMPCM in real situation, if network does not change a lot, the intrinsic dimension of traffic matrix doesn't need to be changed, and there is no need to determine the intrinsic dimension every time before detecting. This significantly reduces the complexity of computation.

6.2. Anomaly Detection Threshold. When network traffic is in normal status, the distribution of the squared Mahalanobis distance of traffic samples is close to Gaussian distribution. In order to verify this, the normal probability plot of P in Table 8 is drawn by normplot of Matlab, as is shown in Figure 21. Samples are mainly distributed on the normal line, which are concentrated in the middle and sparse on both ends. A Jarque-Bera test is employed in order to further verify its normality. Since the samples' squared Mahalanobis distance almost follows Gaussian distribution, "3σ" control chart is taken for anomaly detection in Section 4.3.

6.3. Distribution Characters of Data Source. Conventional network-wide traffic anomaly detection algorithms usually assume that the traffic matrix elements are drawn from a Gaussian or Gaussian-like distribution [13]. We conduct statistical analysis for real network data sets in Table 8 and draw the normal probability plot. We find that the data (traffic volume) exhibit a highly skewed distribution, many of which are diverged from the normal line, and the density of samples scattered on two ends is higher than Gaussian distribution (Figure 22 illustrates some of drawn results). Frequency histograms with superimposed normal density curves are then plotted (as partially shown in Figure 23), which indicates that real traffic volume's heavy-tailed feature is more noticeable compared with Gaussian distribution. This is naturally one of the reasons why Gaussian distribution is replaced by multivariate t-distribution when modeling the traffic matrix.

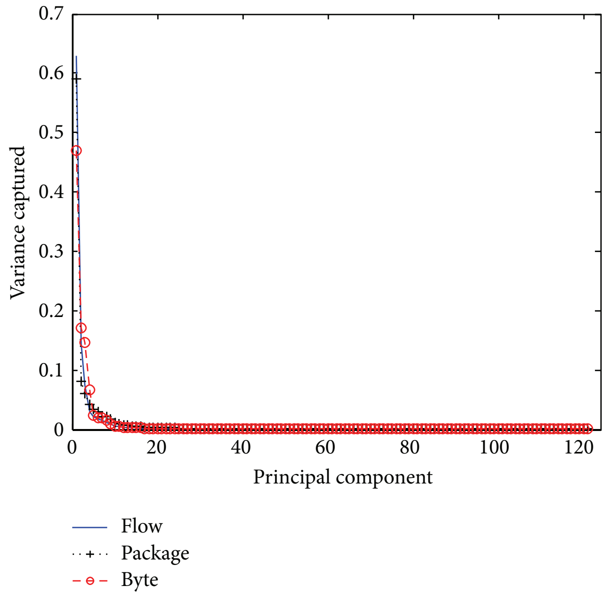


FIGURE 20: Scree plot of data sets in Table 8.

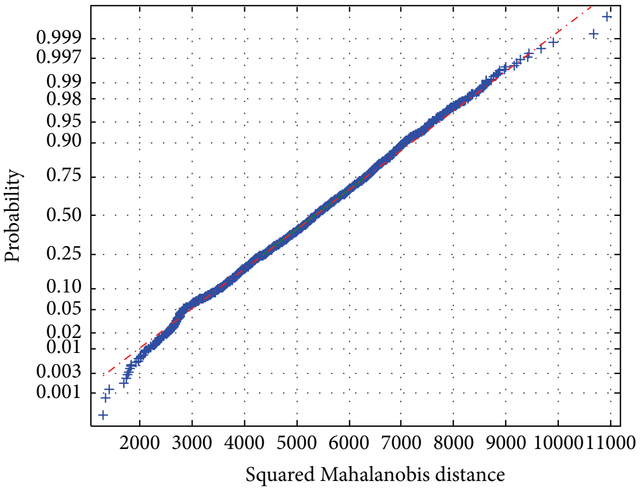


FIGURE 21: Normal probability plot of the squared Mahalanobis distance of normal traffic (normplot test).

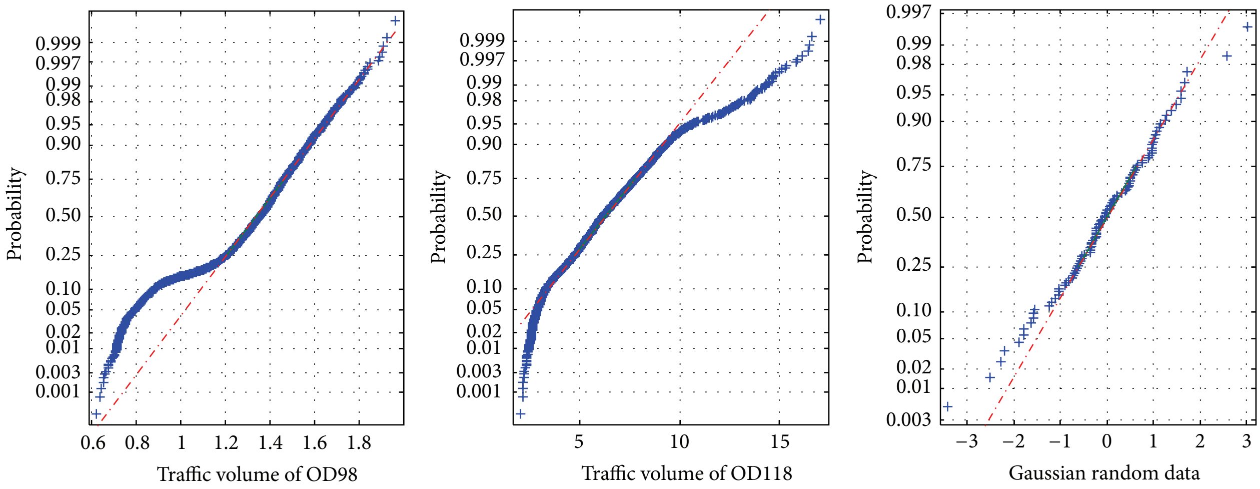


Figure 22: Comparison of normal probability plot of real traffic volume and Gaussian random data.

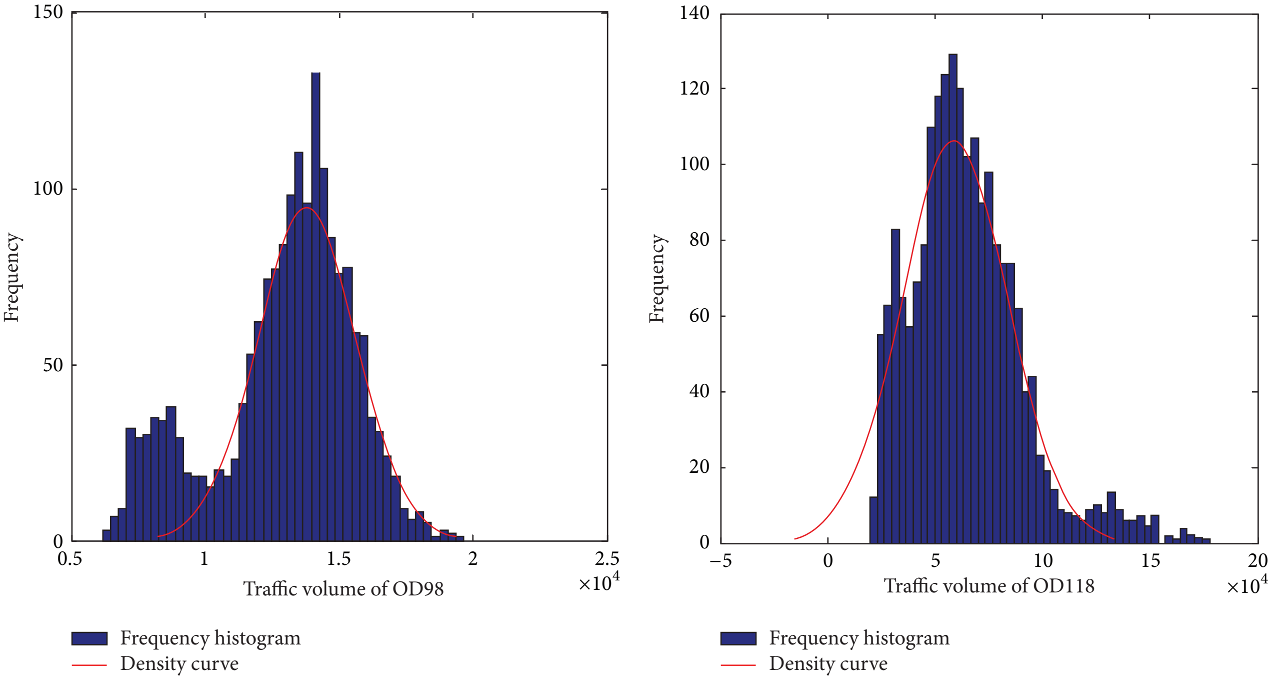


Figure 23: Frequency histograms with superimposed normal density curves of real traffic volume.

6.4. False Positives and False Negatives. Compared with other methods listed in the paper, RMPCM achieves a better performance on reducing false positives and false negatives in the experimental scenarios. One reason is that it is more accurate for RMPCM to describe the traffic data. Conventional traffic anomaly detection methods usually assume that the traffic data are drawn from a Gaussian or Gaussian-like distribution. We conduct statistical analysis for real network data sets, and it indicates that real traffic volume's heavy-tailed feature is more noticeable compared with Gaussian distribution. So we use the multivariate t-distribution as prior distribution when modeling the traffic data. Another reason is that introducing a latent variable probabilistic model based on t-distribution can relieve the noise interference and achieve a better performance. In addition, RMPCM method is within a probabilistic framework, making it attain advantages over traditional method of nonprobabilistic model (such as subspace construction via PCA (Lakhina et al.)) in handling missing data.

However, RMPCM also suffers from false positives and false negatives to some degree. This is due to the dynamic nature of the network traffic flows, which inevitably leads to some deviation when describing the change of traffic by applying this algorithm based on baseline. Moreover, false positives and false negatives will occur under the circumstances in which anomalous traffic generated in

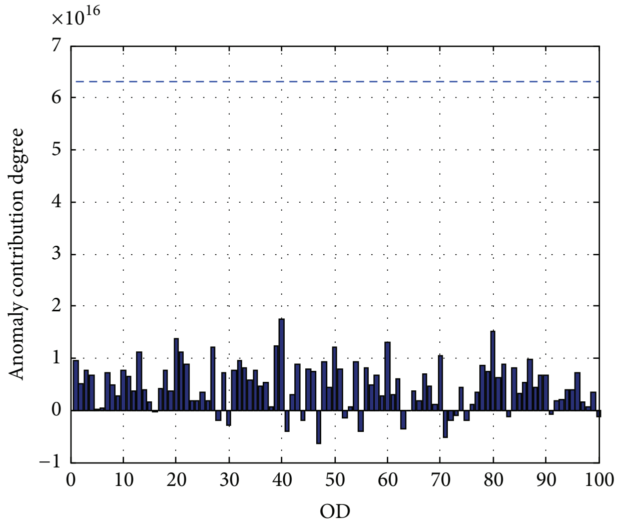


Figure 24: Anomaly location histogram with 5 anomalous OD settings.

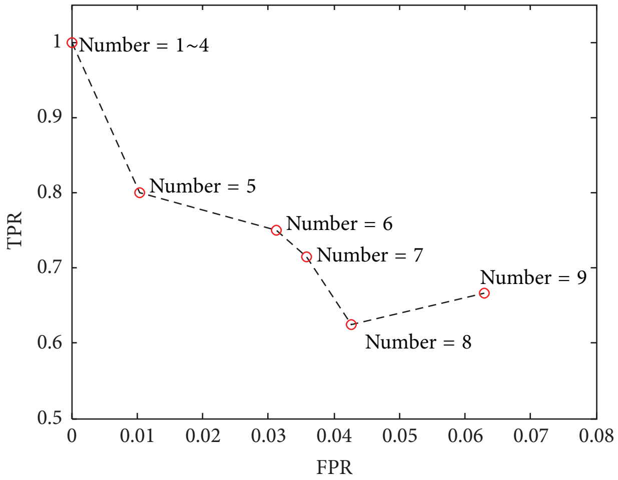


Figure 25: Anomaly location results with 1~9 anomalous OD settings.

the experiments is excessively small or large along with the interference of background traffic. But on the whole the false positives and false negatives of RMPCM can meet the needs of engineering practices.

6.5. Problems on Anomaly Localization. The anomaly localization in scenarios with more anomalous OD is discussed here. The experimental environment is DETERLab with 10 PoP nodes the same as in Section 5.2. The benchmark is set up with the scenario where DDoS attacks took place from 1~9 nodes to one node. When the number of anomalous OD is from 1 to 4, the results agree with the setting. When the number of anomalous OD is 5, the preset anomalous OD are OD20, OD40, OD50, OD60, and OD80, but the localization results are OD20, OD39, OD40, OD60, and OD80, as shown in Figure 24. During the experiment, it is found that the accuracy of anomaly localization decreases with the increasing number of anomalous OD, as shown in Figure 25. This problem will be researched further in depth in the future. There are several preliminary considerations for this taking place.

(a) With the increasing number of anomalous OD, the volume of anomalous traffic also increases significantly, if the volume of anomalous traffic is excessively large which exceeds the anomalous observations tolerance of the proposed approach, and then it is highly likely to impact the detection performance and location results of the proposed approach.

(b) In addition, with the increasing number of anomalous OD and the volume of anomalous traffic, the impact on every single OD becomes relatively smaller, which brings negative impact on the accuracy of anomalous OD localization.

7. Conclusion

In conclusion, traditional network-wide anomaly detection methods have actual problems of performance reduction or being unavailable when noise interference or data loss takes place, and in order to solve these problems and advance anomaly detection and localization, we propose a network-wide approach based on robust multivariate probabilistic calibration model in this paper. The analysis conducted with simulations, DETERLab experiments, and real data from the Internet indicates that the performance of RMPCM is better than PCA and ANTIDOTE and has a better robustness. Regardless of data loss or noise interference, RMPCM demonstrates a stable performance and less sensitivity to the change of parameters. RMPCM can be also applied to locating anomalies. In the future, we will take our researches on more accurate and fine-grained anomaly localization and online RMPCM algorithm.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the National Grand Fundamental Research 973 Program of China under Grant no. 2012CB315901 and no. 2013CB329104; Science and Technology Commission of Shanghai Municipality Research Program under Grant no. 13DZ1108800; Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory Research Fund.

References

[1] D. Turner, K. Levchenko, S. Savage, and A. C. Snoeren, “A comparison of syslog and IS-IS for network failure analysis,” in *Proceedings of the 13th ACM Internet Measurement Conference (IMC ’13)*, pp. 433–440, ACM, Barcelona, Spain, October 2013.

[2] R. Vaarandi and M. Pihelgas, “Using security logs for collecting and reporting technical security metrics,” in *Proceedings of the*

33rd Annual IEEE Military Communications Conference (MILCOM '14), pp. 294–299, IEEE, Baltimore, Md, USA, October 2014.

[3] K. V. M. Naidu, D. Panigrahi, and R. Rastogi, "Detecting anomalies using end-to-end path measurements," in *Proceedings of the 27th IEEE Conference on Computer Communications (INFOCOM '08)*, IEEE, Phoenix, Ariz, USA, April 2008.

[4] P. Barford, N. Duffield, A. Ron, and J. Sommers, "Network performance anomaly detection and localization," in *Proceedings of the 28th Conference on Computer Communications (INFOCOM '09)*, pp. 1377–1385, IEEE, Rio de Janeiro, Brazil, April 2009.

[5] G.-Z. Cheng, D.-N. Cheng, and D.-J. Yu, "Network traffic detection based on multi resolution low rank model," *Journal on Communications*, vol. 33, no. 1, pp. 182–190, 2012.

[6] T. Guo, J.-L. Lan, Y.-F. Li, and Y.-M. Jiang, "Network traffic prediction with radial basis function neural network based on quantum adaptive particle swarm optimization," *Journal of Electronics and Information Technology*, vol. 35, no. 9, pp. 2220–2226, 2013.

[7] V. Yegneswaran, P. Barford, and J. Ullrich, "Internet intrusions: global characteristics and prevalence," *ACM SIGMETRICS Performance Evaluation Review*, vol. 31, no. 1, pp. 138–147, 2003.

[8] A. Lakhina, K. Papagiannaki, and M. Crovella, *Structural Analysis of Network Traffic Flows*, SIGMETRICS, New York, NY, USA, 2004.

[9] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *Proceedings of the ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '04)*, pp. 65–76, ACM Press, Portland, Ore, USA, August 2004.

[10] A. Soule, K. E. Salamatian, and N. Taft, "Combining filtering and statistical methods for anomaly detection," in *Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement (IMC '05)*, pp. 311–312, Boston, Mass, USA, 2005.

[11] L. Huang, M. Garofalakis, and J. M. Hellerstein, "Toward sophisticated detection with distributed triggers," in *Proceedings of the SIGCOMM Workshop on Mining Network Data (MineNet '06)*, 2006.

[12] Y.-K. Qian, M. Chen, L.-X. Ye, F.-R. Liu, S.-W. Zhu, and H. Zhang, "Network-wide anomaly detection method based on multiscale principal component analysis," *Journal of Software*, vol. 23, no. 2, pp. 361–377, 2012.

[13] D. Brauckhoff, K. Salamatian, and M. May, "Applying PCA for traffic anomaly detection: problems and solutions," in *Proceedings of the 28th Conference on Computer Communications (INFOCOM '09)*, pp. 2866–2870, IEEE, Rio de Janeiro, Brazil, April 2009.

[14] B. I. P. Rubinstein, B. Nelson, L. Huang et al., "Stealthy poisoning attacks on PCA-based anomaly detectors," in *Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '09)*, ACM Press, Seattle, Wash, USA, August 2009.

[15] B. I. P. Rubinstein, B. Nelson, L. Huang et al., "Antidote: Understanding and defending against poisoning of anomaly detectors," in *Proceedings of the 9th ACM SIGCOMM Internet Measurement Conference (IMC '09)*, pp. 1–14, Chicago, Ill, USA, November 2009.

[16] Y.-K. Qian and M. Chen, "Poison attack and defense strategies on PCA-based anomaly detector," *Acta Electronica Sinica*, vol. 39, no. 3, pp. 543–548, 2011.

[17] T. Ahmed, M. Coates, and A. Lakhina, "Multivariate online anomaly detection using kernel recursive least squares," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 625–633, IEEE, Anchorage, Alaska, USA, May 2007.

[18] Y.-K. Qian and M. Chen, "MOADA-SVR: a multivariate online anomaly detection algorithm based on SVR," *Journal on Communications*, vol. 32, no. 2, pp. 106–113, 2011.

[19] W. Chen, Y. Liu, and Y. Guan, "Cardinality change-based early detection of large-scale cyber-attacks," in *Proceedings of the 32nd IEEE Conference on Computer Communications (INFOCOM '13)*, pp. 1788–1796, IEEE, Turin, Italy, April 2013.

[20] Y.-K. Qian, M. Chen, Q. Hao, F.-R. Liu, and W.-Z. Shang, "ODC: a method for online detecting & classifying network-wide traffic anomalies," *Journal on Communications*, vol. 32, no. 1, pp. 111–120, 2011.

[21] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and internet traffic matrices," in *Proceedings of the ACM SIGCOMM Conference on Data Communication (SIGCOMM '09)*, pp. 267–278, ACM Press, Barcelona, Spain, August 2009.

[22] H. Ringberg, A. Soule, J. Rexford, and C. Diot, "Sensitivity of PCA for traffic anomaly detection," in *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pp. 78–89, ACM Press, 2007.

[23] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. 13, no. 2, pp. 222–232, 1987.

[24] W. Xu and X. Guo, "Nonparametric checks for varying coefficient models with missing response at random," *Metrika*, vol. 76, no. 4, pp. 459–482, 2013.

[25] W. Xu and L. Zhu, "Testing the adequacy of varying coefficient models with missing responses at random," *Metrika*, vol. 76, no. 1, pp. 53–69, 2013.

[26] 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 Internet Measurement Conference (IMC '10)*, pp. 451–464, ACM Press, Melbourne, Australia, November 2010.

[27] M. Svensén and C. M. Bishop, "Robust Bayesian mixture modelling," *Neurocomputing*, vol. 64, no. 1–4, pp. 235–252, 2005.

[28] C. Liu and D. B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM amd ECME," *Statistica Sinica*, vol. 5, no. 1, pp. 19–39, 1995.

[29] D. Peel and G. J. McLachlan, "Robust mixture modelling using the t distribution," *Statistics and Computing*, vol. 10, no. 4, pp. 339–348, 2000.

[30] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.

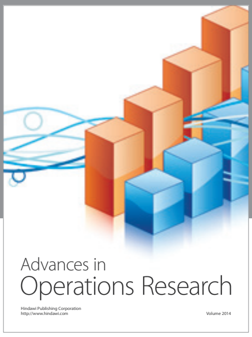
[31] R. J. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley, Chichester, UK, 1987.

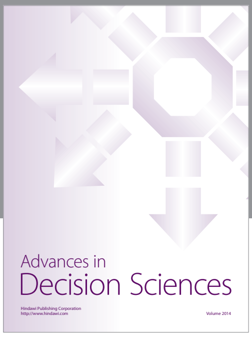
[32] A. Lakhina, M. Crovella, and C. Diot, "Characterization of network-wide anomalies in traffic flows," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement (IMC '04)*, pp. 201–206, ACM Press, New York, NY, USA, 2004.

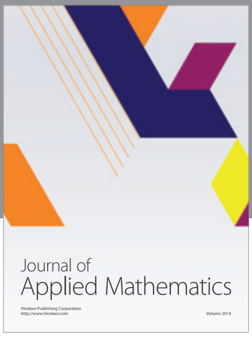
[33] T. Chen, J. Morris, and E. Martin, "Probability density estimation via an infinite Gaussian mixture model: application to statistical process monitoring," *Journal of the Royal Statistical Society, Series C: Applied Statistics*, vol. 55, no. 5, pp. 699–715, 2006.

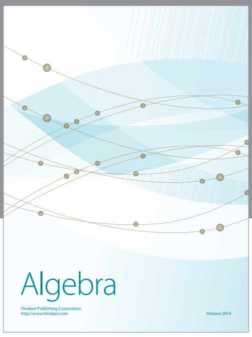
[34] I. Paredes-Oliva, X. Dimitropoulos, M. Molina, P. Barlet-Ros, and D. Brauckhoff, "Automating root-cause analysis of network anomalies using frequent itemset mining," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 467–468, 2011.

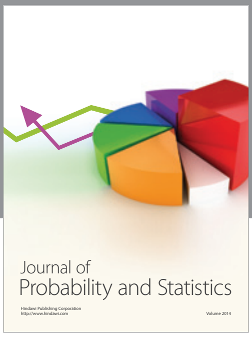
[35] T. Benzel, R. Braden, D. Kim et al., "Experiences with DETER: a testbed for security research," in *Proceedings of 2nd International Conference on Testbeds and Research Infrastructures for the for the Development of Networks and Communities (TridentCom '06)*, pp. 388–397, IEEE Press, 2006.


Advances in
Operations Research

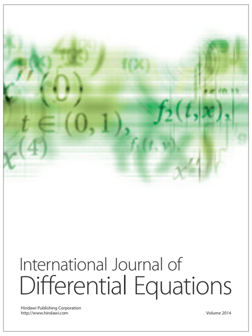

Advances in
Decision Sciences


Journal of
Applied Mathematics

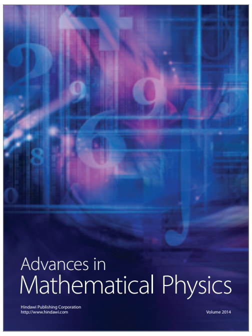

Algebra

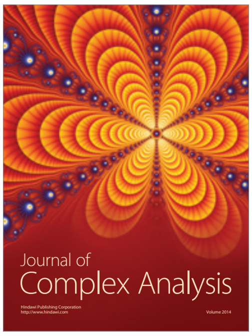

Journal of
Probability and Statistics

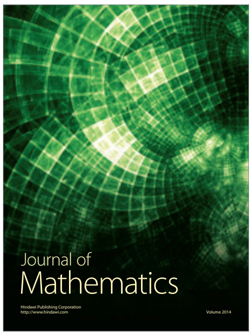
The Scientific
World Journal

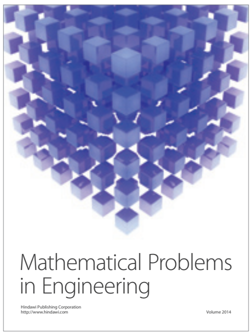

International Journal of
Differential Equations

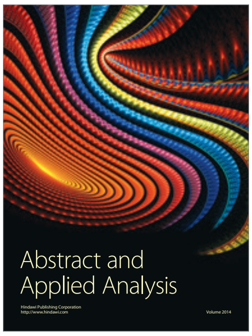
International Journal of
Combinatorics

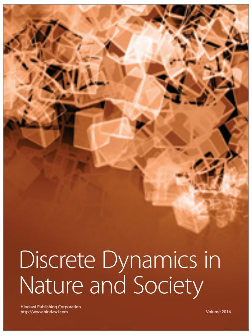

Advances in
Mathematical Physics

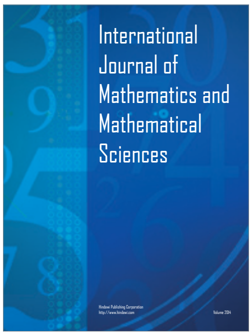

Journal of
Complex Analysis


Journal of
Mathematics


Mathematical Problems
in Engineering

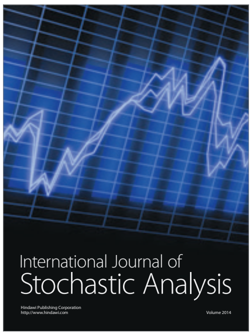

Abstract and
Applied Analysis


Discrete Dynamics in
Nature and Society


International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces


International Journal of
Stochastic Analysis

Journal of
Optimization