Computer Networks 56 (2012) 2049–2067

Contents lists available at SciVerse ScienceDirect

Computer Networks

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

Structural analysis of network traffic matrix via relaxed principal component pursuit ☆

Zhe Wang [a,b,*], Kai Hu [a], Ke Xu [a,b], Baolin Yin [a,b], Xiaowen Dong [c]

[a] School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[b] State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China
[c] Signal Processing Laboratories (LTS4/LTS2), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 23 February 2011
Received in revised form 24 February 2012
Accepted 27 February 2012
Available online 6 March 2012

Keywords:
Network measurement
Traffic matrix structural analysis
Robust principal component analysis
Relaxed principal component pursuit
Accelerated proximal gradient algorithm

ABSTRACT

The network traffic matrix is widely used in network operation and management. It is therefore of crucial importance to analyze the components and the structure of the network traffic matrix, for which several mathematical approaches such as Principal Component Analysis (PCA) were proposed. In this paper, we first argue that PCA performs poorly for analyzing traffic matrix that is polluted by large volume anomalies, and then propose a new decomposition model for the network traffic matrix. According to this model, we carry out the structural analysis by decomposing the network traffic matrix into three sub-matrices, namely, the deterministic traffic, the anomaly traffic and the noise traffic matrix, which is similar to the Robust Principal Component Analysis (RPCA) problem previously studied in [13]. Based on the Relaxed Principal Component Pursuit (Relaxed PCP) method and the Accelerated Proximal Gradient (APG) algorithm, we present an iterative approach for decomposing a traffic matrix, and demonstrate its efficiency and flexibility by experimental results. Finally, we further discuss several features of the deterministic and noise traffic. Our study develops a novel method for the problem of structural analysis of the traffic matrix, which is robust against pollution of large volume anomalies.

1. Introduction

1.1. The Internet traffic data

The Internet traffic data is considered as a significant input for network operation and management. To monitor and analyze traffic data efficiently is one of the most important problems in the research field of network measurements. In general, there are two levels of traffic data: the packet-level data and the flow-level data. With the rapid growth of the Internet scale and link transmitting capability, on the one hand, it is usually infeasible to collect and process the complete packet-level data; On the other hand, the coarse flow-level data obtained by packet sampling often contains enough network information and has become increasingly popular in recent studies. As an example, one type of widely used flow-level traffic data is the IP flows collected in each router by the Netflow protocol. Roughly speaking, each IP flow is a sequence of packets sharing the same Source/Destination IP addresses, Source/Destination port numbers and transport protocol during certain time intervals. However, in large scale networks, the volume of IP flow data is still too huge for storage and processing. For instance, the one-month IP flow data of the GEANT backbone network is about 150 GB [1], which makes important applications such as anomaly detection impractical. Therefore, it is necessary to further compress

☆ This work is supported by the National Natural Science Foundation of China (Grant No. 61073013) and the Aviation Science Foundation of China (Grant No. 2010ZA04001).

* Corresponding author at: School of Computer Science and Engineering, Beihang University, Beijing 100191, China.
E-mail addresses: wangzhe@cse.buaa.edu.cn (Z. Wang), hukai@buaa.edu.cn (K. Hu), kexu999@gmail.com (K. Xu), yin@nlsde.buaa.edu.cn (B. Yin), xiaowen.dong@epfl.ch (X. Dong).

1389-1286/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2012.02.017

the IP flow data, mainly by means of flow sampling and aggregation.

The network traffic matrix is computed by IP flow aggregation, which records how much data is transmitted between each Original-Destination (OD) pair during different time intervals. For each OD pair (k_1, k_2), the original point k_1 and the destination point k_2 are both the Points of Presence (PoP) in the network, and we aggregate all the IP flows which enter the network at k_1 and exit at k_2 to represent the OD flow corresponding to (k_1, k_2). The research topics in traffic matrix analysis mainly include: (1) To estimate a traffic matrix accurately; (2) To generate synthetic traffic matrix; (3) To utilize a traffic matrix effectively for measurement applications, such as anomaly detection and routing optimization. These topics require a deep understanding of the components and structure of the traffic matrix. In this paper, we carry out structural analysis of the traffic matrix by studying different traffic components that constitute the traffic matrix.

1.2. Traffic matrix and its structural analysis

Considering the network traffic with p OD flows during t time intervals, the corresponding traffic matrix X is a $t \times p$ non-negative matrix. For each integer $1 \leqslant j \leqslant p$, the jth column X_j of X is the traffic time series for the jth OD flow; for each integer $1 \leqslant i \leqslant t$, the ith row $X(i, :)$ of X is the traffic snapshot of all the OD flows during the ith time interval. According to the datasets adopted in this paper, we suppose $t > p$. In real network measurements, the Netflow protocol consumes much CPU resources; some PoP routers might not support Netflow; and the IP flow data might get lost during the transmission. These limitations make the collection of a complete traffic matrix a challenging task, therefore many estimation algorithms by means of indirect measurements (such as link traffic data collected by the SNMP protocol) were proposed in the literature. In recent studies, the error of the third generation estimation algorithms have been decreased to below 10%. In this paper, however, we do not concentrate on the estimation problem. Instead, we perform our experiments based on real world datasets of traffic matrices. These datasets are collected from the Abilene networks (in the US) and the GEANT networks (in Europe), which are available from [2,3], respectively.

In general, a traffic matrix is a combination of different classes of network traffic. In network operation and management, people usually need information on all classes of traffic, such as the deterministic traffic and the anomaly traffic. In this paper, we study the structural analysis problem of a traffic matrix, which is to accurately decompose the traffic matrix into sub-matrices that correspond to different classes of traffic, hence explore in detail various features of the network traffic.

The most widely used approach for traffic matrix analysis is Principal Component Analysis (PCA) and its variants. For example, Lakhina et al. [4] first introduced the PCA method in the studies of traffic matrices, and they found that traffic matrices can be well approximated by a few principal components that correspond to the largest singular values of the matrices. Therefore, they claimed that traffic matrices usually have *low effective dimensions*. They further introduced the concept of *eigenflow* and the eigenflow classification method, discussed the distribution pattern of different eigenflow classes, and proposed a method to decompose each OD flow time series according to the classification results. These ideas were further developed in their later work [5], in which they presented the PCA-subspace method for volume anomaly detection, and decomposed the link traffic matrix into two sub-matrices that correspond to a normal subspace and an anomaly subspace, respectively. During each time interval, the norm of the traffic volume that corresponds to the anomaly subspace was compared with the Q-statistic threshold, whose result was used to infer the existence of anomalies in the network.

After Lakhina's studies, many researchers have enriched the PCA-based methods for traffic matrix analysis. Huang et al. [6] proposed a distributed PCA method for volume anomaly detection, which considered the trade-off between the detection accuracy and the data communication overhead. Zhang et al. [7] extended the classical PCA method and argued that large volume anomalies should be extracted via both spatial and temporal approaches, which were later named as *Network Anomography*. Based on the fact that traffic matrices often have low effective dimensions, Soule et al. [8] proposed a new principal component method for traffic matrix estimation, and the experiments demonstrated that their approach has a lower estimation error compared to most of the previous methods, such as the tomogravity method and the route change method.

However, recent studies have shown some limitations of the PCA-based methods. Ringberg et al. [10] experimented the PCA-based anomaly detection method, suggesting that its efficiency is very sensitive to the choice of parameters, such as the number of principal components in the normal subspace, the value of detection threshold, and the level of traffic aggregation. In addition, they found that large volume anomalies might pollute the normal subspace, which could lead to high false positive rate in anomaly detection. Ohsita et al. [11] argued that the traffic matrix estimated by network tomography is not a proper input for the PCA-based anomaly detectors. Since most estimation methods are designed for an anomaly-free traffic matrix, the estimation error might increase when the network traffic contains large volume anomalies. Instead, they suggested estimating the increased traffic matrix and obtained a high attack-detection rate. Their research also indicated the strict requirements of input traffic matrix for the PCA-based methods. More recently, Rubinstein et al. [12] proved that attackers could significantly degrade the performance of the PCA-based anomaly detectors simply by adding chaff volumes before real attacks, and designed an anomaly detector that is more robust against these attacks.

1.3. Main contributions of this paper

As mentioned above, although it has been extensively studied before, PCA-based methods still have limitations for traffic matrix analysis and related applications. One important drawback is that, when the traffic matrix is

corrupted by large volume anomalies, the resulting principal components will be significantly skewed from those in the anomaly-free case. This prevents the subspace-based methods from accurately decomposing the total traffic into normal traffic and anomaly traffic, and decreases the efficiency of PCA-based anomaly detectors. However, to our knowledge there are only a few existing methods for analyzing a traffic matrix with large volume anomalies. This is going to be the focus of this paper, where we address the problem of structural analysis of polluted traffic matrices. The main contribution of our paper is twofold:

(1) As the basic assumption behind the subspace-based methods is that each eigenflow can be exactly classified, it is an interesting question whether those classification method still perform well for a polluted traffic matrix. Specifically, (i) is there an eigenflow that satisfies more than one classification criterion of the eigenflow classes? (ii) is there an eigenflow that satisfies no classification criterion of the eigenflow classes? (iii) does the distribution pattern of eigenflows maintain for a polluted traffic matrix? We discuss these problems in Section 2, where we use PCA for the structural analysis of real world traffic matrices, which usually have large volume anomalies.
(2) As the PCA-based structural analysis performs poorly when the traffic matrix contains large volume anomalies, it is necessary to provide a new analysis tool that is suitable for polluted traffic matrices. In Sections 3 and 4, we propose a new decomposition model for the traffic matrix based on empirical network traffic data, and formalize the mathematical definition of the structural analysis problem. This motivates us to discover the equivalence between structural analysis and the Robust Principal Component Analysis (RPCA) problem. We then design a decomposition algorithm based on the Relaxed Principal Component Pursuit (Relaxed PCP) method, which is suitable for solving the RPCA problem. Using this algorithm, we are able to obtain a proper traffic decomposition for the polluted traffic matrices in our experiments. Finally, we analyze several properties of the sub-matrices from the decomposition of the traffic matrix in detail in Section 5.

2. PCA for the structural analysis of polluted traffic matrix

2.1. The classical PCA method

PCA is widely used in high dimensional data analysis, where the redundant high dimensional data can be approximated by a low dimensional representation. In our study, we consider each row vector of the traffic matrix $X \in \mathbb{R}^{t\times p}$ as a data point in $\mathbb{R}^p$, thus X contains t data points. Following the common approach in [4,5], we normalize each OD flow vector (columns of X) to have zero mean before performing PCA:

$$X_j = X_j - \mathrm{mean}(X_j) \quad j = 1,2,\ldots,p. \tag{1}$$

PCA can be viewed as a coordinate transformation process, where the data points have been transformed from the original coordinate system to a new coordinate system. All the unit vectors of the new coordinate system are represented as $\{v_i\}_{i=1}^p$, and v_i is called the ith principal component vector. The first principal component vector v_1 captures the maximum variance (energy) of the original traffic matrix X:

$$v_1 = \underset{\|v\|=1}{\mathrm{argmax}} \|Xv\|. \tag{2}$$

For each integer $k \geqslant 2$, suppose we have obtained the first $k-1$ principal component vectors, the kth principal component vector v_k then captures the maximum variance of the residual traffic matrix, which is the difference between the original traffic matrix X and its mappings onto the first $k-1$ principal component vectors:

$$v_k = \underset{\|v\|=1}{\mathrm{argmax}} \left\| \left(X - \sum_{i=1}^{k-1} X v_i v_i^T \right) v \right\|. \tag{3}$$

Following this progress, all the principal component vectors are defined iteratively. It is easy to show that $\{v_i\}_{i=1}^p$ form an orthogonal basis of $\mathbb{R}^p$. Thus the traffic matrix can be decomposed as:

$$\begin{aligned} X &= X[v_1\ v_2\ \ldots\ v_p][v_1\ v_2\ \ldots\ v_p]^T = \sum_{i=1}^{p} X v_i v_i^T \\ &= \sum_{i=1}^{p} \|Xv_i\| \frac{Xv_i}{\|Xv_i\|} v_i^T = \sum_{i=1}^{p} \|Xv_i\| u_i v_i^T, \end{aligned} \tag{4}$$

where

$$u_i = \frac{Xv_i}{\|Xv_i\|} \quad i = 1,2,\ldots,p \tag{5}$$

is a unit vector in $\mathbb{R}^t$, which is called the ith eigenflow corresponding to v_i [4].

Following basic matrix theory, we can show that the principal component vectors $\{v_i\}_{i=1}^p$ are the eigenvectors of the matrix X^TX, sorted by the corresponding eigenvalues $\{\lambda_i\}_{i=1}^p$ in a descending order:

$$X^TXv_i = \lambda_i v_i \quad i = 1,2,\ldots,p, \tag{6}$$

where $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0$. Furthermore, $\sigma_i = \sqrt{\lambda_i}$ is called the i-th singular value of X. Therefore, $\{v_i\}_{i=1}^p$ can be found by computing the eigenvectors of X^TX.

Since $\|Xv_i\| = \sqrt{v_i^TX^TXv_i} = \sqrt{\lambda_i v_i^T v_i} = \sigma_i$, Eq. (4) can be rewritten as:

$$X = \sum_{i=1}^{p} \sigma_i u_i v_i^T. \tag{7}$$

Eq. (7) is called the Singular Value Decomposition (SVD) of X. By Eckart–Young theorem [19], for each integer $1 \leqslant r \leqslant p$, $A_r = \sum_{i=1}^{r} \sigma_i u_i v_i^T$ is the best rank-r approximation of X:

$$A_r = \underset{\mathrm{rank}(A)\leqslant r}{\mathrm{argmin}} \|X - A\|_F, \tag{8}$$

where $\|\cdot\|_F$ denotes the Frobenius matrix norm. Each data point $X(s,:)^T \in \mathbb{R}^p$ $(1 \leqslant s \leqslant t)$ can then be approximated as

$$X(s,:)^T \approx \sum_{i=1}^{r} \sigma_i u_i(s) v_i, \tag{9}$$

where $\{\sigma_i u_i(s)\}_{i=1}^{r}$ are the first r coefficients of $X(s,:)^T$ under the new coordinate system. Therefore, PCA can be viewed as a technique for dimensionality reduction.

2.2. Eigenflows of the polluted traffic matrix

In the following experiments, we adopt two widely used traffic matrix datasets, one from the Abilene network and the other from the GEANT network. Abilene is a Internet2 backbone network with 12 PoPs (144 OD flows), and GEANT is a pan-European research network with 23 PoPs (529 OD flows). In this paper, the minimal time interval in the Abilene dataset is 5 min; the minimal time interval in the GEANT dataset is 15 min, since the flow files are written in 15 min. In fact, there exists flow data with finer time scale. For example, the minimal time interval of another GEANT traffic dataset is precise to 1 s in [28]. The Abilene dataset contains 24 weeks' traffic records from March 1, 2004 to September 10, 2004, and each week's data is represented by a traffic matrix. Here we select the traffic matrices that correspond to the first 8 weeks (denoted as X01–X08). For the GEANT dataset, traffic records are not complete for certain days. Since most of the researchers study the weekly traffic matrix, we choose a subset of the GEANT dataset: the four consecutive weeks' traffic matrices from March 28, 2005 to April 24, 2005 (denoted as Y01–Y04). Each traffic matrix in the Abilene dataset consists of 2016 rows (time steps), while each GEANT traffic matrix consists of 672 rows.

Notice that for each of these traffic matrices, there are a small number of OD flow time series that contain a large percentage of zero entries. This usually means that these OD flows are not stable, therefore we delete them in the experiments. For each traffic matrix in the Abilene dataset, we delete the 23 OD flows whose source or destination PoP is "ATLA-M5" (thus the number of OD flows actually used is 121); For each traffic matrix in the GEANT dataset, we delete those OD flows that have more than 50% zero entries (the number of OD flows actually used is between 457 and 483). Table 1 summarizes the datasets used in our experiments, and Table 2 (the first column) shows the number of OD flows actually used for each traffic matrix.

In Lakhina's original study [4], all the eigenflows of a traffic matrix can be classified into three types: First, the eigenflows that exhibit distinct periodical patterns are called *d-eigenflow* (for "deterministic"), since they reflect the diurnal activity in the network traffic, as well as the difference between weekday and weekend activities; Second, the eigenflows that represent strong, short-lived spikes are called *s-eigenflow* (for "spike"), as they capture the occasional traffic bursts and dips which are usually reported; Third, the eigenflows that roughly have a stationary and Gaussian behavior are called *n-eigenflow* (for "noise"), since they capture the remaining random variation that arises due to multiplexing of many individual traffic sources.

In this paper, we follow the classification criteria in [4] for both d-eigenflow and s-eigenflow. The original classification criterion for n-eigenflow in [4] is to compare the qq-plot of eigenflow's distribution with the normal distribution; However, it is not considered as a quantitative method. Therefore, we use another classification criterion from the Kolmogorov–Smirnov (K–S) test instead. Suppose u_j is an eigenflow of the traffic matrix X, we classify it according to the following three criteria:

(1) d-eigenflow: Let H denote the set of period parameters measured in hours. For each element $h \in H$, we compute the Fourier power spectrum $\widetilde{u}_j(h)$ of u_j:

$$\begin{cases} \widetilde{u}_j(h) = \left|\sum_{k=0}^{t-1} u_j(k+1)\cdot \exp(-\omega k i)\right|^2 / t \\ \omega = 2\pi/T \\ T = 60h/t_0 \end{cases},$$

where T is the period of Fourier transform and t_0 is the length of time interval (measured in minutes). In this paper, $H = \{k\}_{k=1}^{10} \bigcup \{2k\}_{k=6}^{25}$. If $\{12, 24\} \bigcap \text{argmax}_{k\in H}\{\widetilde{u}_j(k)\} \neq \phi$, u_j satisfies the criterion of d-eigenflow and we classify it as d-eigenflow;

(2) s-eigenflow: Let σ denote the standard deviation of u_j. If u_j has at least one entry outside the interval [mean $(u_j) - 5\sigma$, mean $(u_j) + 5\sigma$], u_j satisfies the criterion of s-eigenflow and we classify it as s-eigenflow;

(3) n-eigenflow: We use the K–S test to verify the normal distribution of u_j. If the null hypothesis (Normal Distribution) is not rejected at 5% significance level, u_j satisfies the criterion of n-eigenflow and we classify it as n-eigenflow.

In order to evaluate the completeness (each eigenflow has to be classified into at least one class) and orthogonality (the same eigenflow must not be classified into more than one class at the same time) of eigenflow classification, we further define the following two concepts:

- indeterminate eigenflow: eigenflows that satisfy more than one classification criterion;
- non-determinate eigenflow: eigenflows that satisfy no classification criterion.

Next, we apply PCA to compute the principal component vectors and eigenflows of each traffic matrix. Following the classification criteria proposed above, we summarize the classification results of eigenflows in Table 2. Here we

Table 1
Datasets used in the experiments.

Name	#OD flows/actually used	Time interval	Time steps	Peroid
Abilene	144/121	5 min	2016	8 weeks
GEANT	529/457–483	15 min	672	4 weeks

Table 2
Eigenflow classification using PCA.

Traffic matrix name (# OD flows used)	# Satisfy d-eigenflow	# Satisfy s-eigenflow	# Satisfy n-eigenflow	# non-determinate eigenflow	# Indeterminate eigenflow	# Classified eigenflow	Unclassified energy rate (%)
X01 (121)	10	42	87	10	26	85	83.08
X02 (121)	9	46	80	10	22	89	82.59
X03 (121)	14	78	47	12	29	80	11.08
X04 (121)	10	62	74	5	28	88	94.67
X05 (121)	8	73	62	9	30	82	46.09
X06 (121)	15	52	83	9	33	79	0.32
X07 (121)	6	63	75	9	30	82	9.39
X08 (121)	10	53	71	11	24	86	98.49
Y01 (483)	17	54	469	5	59	419	76.67
Y02 (465)	5	66	453	6	65	394	77.70
Y03 (465)	7	47	454	5	48	412	43.19
Y04 (457)	12	62	444	5	66	386	82.42

define the *unclassified energy rate* as the percentage of energy captured by the principal component vectors that correspond to either indeterminate or non-determinate eigenflows. Since the energy captured by one principal component vector is proportional to the square of the corresponding singular value, we have:

$$\text{unclassified energy rate} = \frac{\sum_{k\in UEID}\lambda_k}{\sum_{i=1}^{p}\lambda_i}, \tag{10}$$

where the union of unclassified eigenflow ID (*UEID*) is

$$UEID = \{k|u_k \text{ is indeterminate or non-determinate}\}. \tag{11}$$

Based on the classification results of the twelve weekly traffic matrices described above, we make the following observations on traffic matrices that are possibly polluted by large volume anomalies:

(1) One the one hand, only a small number of eigenflows satisfy the classification criterion of d-eigenflow (usually less than 20), and most of them correspond to large singular values; On the other hand, there is a considerable number of eigenflows satisfying the classification criterion of s-eigenflow and n-eigenflows. The proportion of each eigenflow class varies from one traffic matrix to another. These are similar to Lakhina's experimental results presented in [4].
(2) The PCA-based eigenflow classification method shows serious limitations in terms of classification completeness and orthogonality. Specifically, a large number of eigenflows cannot be exactly classified into one eigenflow class. Furthermore, some unclassified eigenflows correspond to large singular values, and the unclassified energy rate is larger than 70% for seven of the twelve traffic matrices. These results are not consistent with those in Lakhina's study [4], in which the non-determinate eigenflows do not exist and the indeterminate eigenflows only contribute little energy. Although the authors in [4] argued that their classification method could be enhanced by heuristic mechanisms, our experiments show that some unclassified eigenflows are essentially different from all the eigenflow classes, hence the classification results cannot be clearly improved only by changing parameters or adopting heuristic algorithms.
(3) For each traffic matrix in the Abilene dataset (X01–X08), the first six eigenflows (eigenflows corresponding to the six largest singular values) often contain some instances satisfying the classification criterion of s-eigenflow. This does not happen in Lakhina's study [4], where the first six eigenflows are exactly classified as d-eigenflows. For the GEANT traffic matrices, the first six eigenflows do not satisfy the criterion of s-eigenflow in general, which can be explained by the fact that the anomaly volumes in the GEANT networks are not as large as that in the Abilene networks.

Here we present the experimental result for the traffic matrix X01 as a case study. Fig. 1 shows the classification result of the 121 eigenflows (sorted from large to small by the corresponding singular values), where each indeterminate eigenflow appears simultaneously in more than one classes whose classification criteria it satisfies. The first six eigenflows and their Fourier power spectra are shown in blue in Figs. 2 and 3, respectively. Each pair of red lines in Fig. 2 is the $+5\sigma/-5\sigma$ boundary for the corresponding eigenflow, which is used for the inference of s-eigenflow. It is clear that five out of six eigenflows in Fig. 2 satisfy the classification criterion of s-eigenflow. However, for the first four eigenflows, their Fourier power spectra all achieve the maximum values when the period parameter is equal to 24 h, suggesting that they satisfy the criterion of d-eigenflow. Therefore, we can view each of the first four eigenflows as a hybrid time series mixed with the deterministic diurnal pattern and the short-lived anomaly pattern, which is quite different from the three pre-defined eigenflow classes. Changing parameters or using heuristic algorithms in classification would not help much in this case. If we classify these four eigenflows as d-eigenflow, the energy of anomaly traffic will be significantly underestimated, which might increase the false negative rate of anomaly detection algorithms; On the contrary, classifying them as s-eigenflow will lead to underestimation of normal network traffic, which could

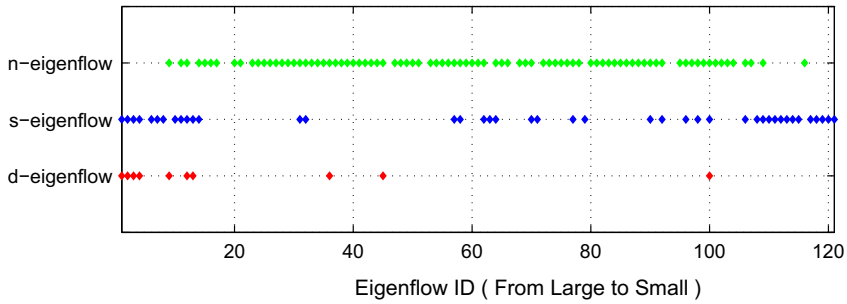


Fig. 1. Eigenflow classification result for the traffic matrix X01.

prevent us from correctly decomposing the deterministic traffic component from a polluted traffic matrix.

From the discussions above, we can see that PCA-based method has limitations in eigenflow classification: (i) for both the Abilene dataset and the GEANT dataset, a large proportion of eigenflows could not be exactly classified into one eigenflow class; (ii) for the Abilene dataset, some eigenflows corresponding to the six largest singular values satisfy the classification criterion of s-eigenflow, which would be problematic for us to isolate the deterministic traffic trend or to detect the anomaly traffic events. This motivates us to propose a new model for the structural analysis of traffic matrix in the next section.

3. Relaxed principal component pursuit for the structural analysis of polluted traffic matrix

3.1. The decomposition model of traffic matrix

In this section we propose a new decomposition model for the traffic matrix data, and discuss the mathematical nature of the structural analysis problem. Next, based on the optimization process corresponding to PCA, we explain intuitively the limitations of PCA-based method for analyzing a polluted traffic matrix. This helps interpret the experimental results presented in Section 2.2 more deeply.

According to the empirical measurement data, we suppose that there exist three classes of network traffic: the deterministic traffic which shows diurnal pattern; the anomaly traffic which appears rarely but involves large peak-like or block-like volumes; the noise traffic that has small magnitude but appears in every OD flow during all the time intervals. Formally speaking, we propose to decompose each traffic matrix into three sub-matrices:

$$X = A + E + N, \tag{12}$$

where A, E and N represent the deterministic traffic matrix, the anomaly traffic matrix and the noise traffic matrix, respectively. Each class of traffic has its own features, based on which we make the following hypotheses:

(1) The deterministic traffic is mainly contributed by periodical traffic. This means that the periodical traffic time series of different OD flows have similar periods and phases, and they mainly differ in magnitudes. This implies that A should be a low-rank matrix.

(2) The anomaly traffic does not show up frequently, which implies that E should be a sparse matrix. However, its nonzero entries may have large magnitudes.

(3) We assume that N is a random matrix consists of independent zero-mean Gaussian random variables. Each column of N (the noise traffic time series of an OD flows) can be viewed as Gaussian random variables which have the same variance. Since the noise variance is usually proportional to the scale of the corresponding OD flow, and different OD flows may have very different scales, the Gaussian random variables corresponding to different columns may have different variances.

Eq. (12) and hypotheses (1)–(3) constitute our decomposition model for the traffic matrix. In practical measurements, only the traffic matrix X can be observed, thus the mathematical nature of the structural analysis problem is to exactly decompose X into three components, namely A, E and N in our decomposition model.

Consider a simple case where we neglect the noise traffic matrix N in (12). Suppose that $rank(A) = r_0$. According to Section 2.1, $A_{r_0} = \sum_{i=1}^{r_0} \sigma_i u_i v_i^T$ is the best rank-r_0 approximation of X. Then, one natural question is to ask whether $A = A_{r_0}$. If this is true, the PCA method would achieve a decomposition of X that is consistent with our decomposition model ($A_{r_0} = A$, hence $X - A_{r_0} = E$). However, it has not been shown that A is the solution to the optimization problem (8). Considering the magnitude of nonzero entries in E, our discussion is further divided into the following two cases:

(1) If the anomaly traffic only has entries with small magnitude, the Frobenius matrix norm $\|E\|_F = \|X - A\|_F$ would be small, thus the deterministic traffic matrix A is most likely to be the solution to problem (8). In this case $A = A_{r_0}$, and these two matrices have the same eigenflows. As the deterministic traffic is mainly contributed by diurnal traffic, the first r_0 eigenflows of A are usually d-eigenflows (since $rank(A) = r_0$). Following the definition of A_{r_0}, these eigenflows are also the first r_0 eigenflows of X. Therefore, when the anomaly traffic has entries with small magnitude, the PCA method performs well in eigenflow classification.

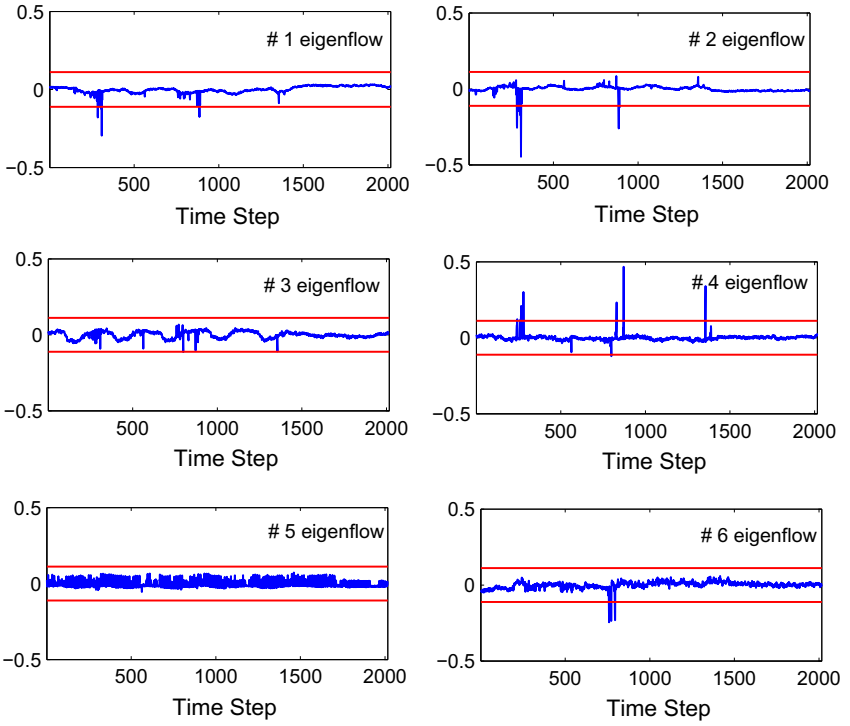


Fig. 2. The first six eigenflows of the traffic matrix X01.

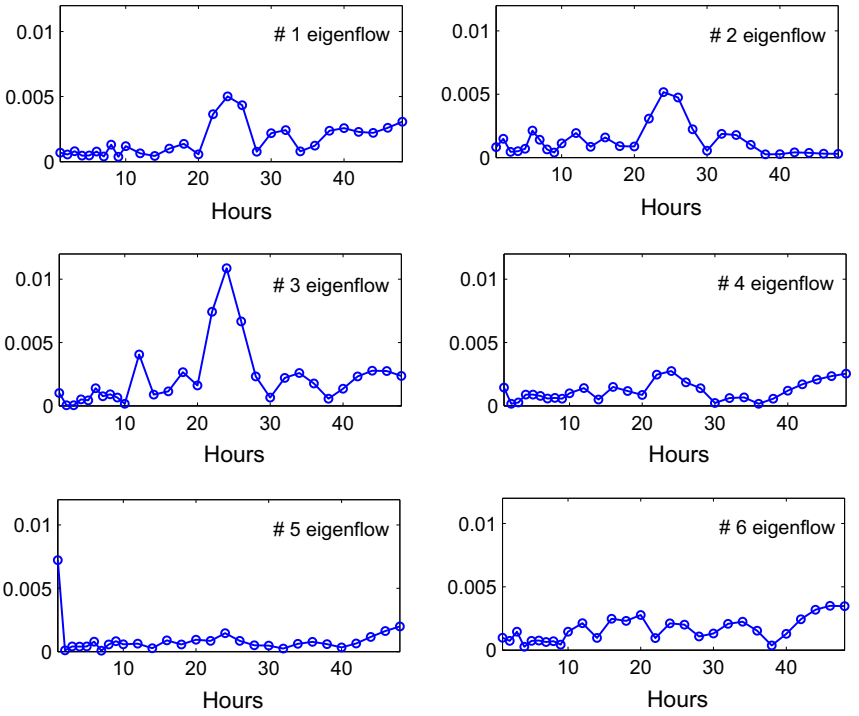


Fig. 3. Fourier power spectra of the first six eigenflows of X01.

(2) If the anomaly traffic has entries with large magnitude, even though E is a sparse matrix, its Frobenius matrix norm $\|E\|_F = \|X - A\|_F$ would be large. As a result, A is usually not the solution to problem (8), and A_{r_0} contains a large amount of anomaly traffic. Therefore, some of the first r_0 eigenflows of X may satisfy the classification criterion of s-eigenflow.

The discussions above explain intuitively the experimental results presented in Section 2.2. The PCA-based eigenflow classification can be considered as a special method for traffic matrix decomposition. However, when the traffic matrix is polluted by large volume anomalies, PCA cannot achieve a complete and orthogonal eigenflow classification, and the PCA-based traffic matrix decomposition is inconsistent with the proposed decomposition model.

3.2. Robust principal component analysis and principal component pursuit

Following the decomposition model proposed in Section 3.1, the structural analysis problem of the traffic matrix is to accurately decompose the original traffic matrix X into a deterministic traffic matrix A, an anomaly traffic matrix E, and a noise traffic matrix N. This is similar to the Low-rank Matrix Recovery problem, which is also known as *Robust Principal Component Analysis (RPCA)* [13].

Recently, developments in the theory of *Compressive Sensing* [24,20] have attracted wide attentions in the field of information science. Compressive Sensing theory states that: If the signal has a sparse representation under some orthonormal bases or dictionaries, it can then be recovered by far fewer samples or measurements than that are needed using traditional methods. Partially motivated by this claim, Ma et al. proposed the *Principal Component Pursuit (PCP)* method for the RPCA problem. They studied the approximate algorithms for PCP, and applied it to different real world applications such as video background modeling [13], face alignment [15], and web document corpus analysis [16]. Following the definitions in [13,14,17], we briefly describe the RPCA problem and the PCP method as follows. We assume that X, A, E, N are real matrices in $\mathbb{R}^{t\times p}$; $\Lambda(\cdot)$ denotes the support set of a matrix, which is the union of non-zero positions of the matrix.

Problem 1 (*standard RPCA problem*). Suppose that $X = A + E$, where A and E are two unknown matrices. Assume that A is a low-rank matrix and E is a sparse matrix, the standard RPCA problem is to recover A and E from X.

The authors in [17] suggested that the standard RPCA problem can be formulated as the following optimization problem:

$$\min_{A,E} rank(A) + \gamma\|E\|_0 \quad s.t.\ A + E = X, \tag{13}$$

where $\|\cdot\|_0 = |\Lambda(\cdot)|$ is the degree of the support set, which is also called the l_0-norm of the matrix; γ is a positive parameter that balance the two competing terms. Problem (13) consists of two sub-problems, namely the low-rank matrix completion problem and the l_0-norm minimization problem. Unfortunately, both of them are NP-hard, which makes problem (13) intractable in polynomial time.

On the one hand, thanks to the developments in Compressive Sensing theory, a series of previous works suggested the equivalence between the l_0-norm minimization and the l_1-norm minimization problems; On the other hand, recent research on the matrix completion problem [21] studied the matrix nuclear norm $\|\cdot\|_*$ (for a matrix $A \in \mathbb{R}^{t\times p}$, the nuclear norm $\|A\|_* = \sum_{k=1}^{p}\sigma_k(A)$ is defined as the sum of its singular values $\{\sigma_k(A)\}_{k=1}^{p}$), and indicated that the two optimization problems, namely, the matrix rank minimization problem and the problem of minimizing the matrix nuclear norm, usually produce similar results. More importantly, both the l_1-norm minimization and the nuclear norm minimization problems are convex optimization problems hence can be solved efficiently.

The PCP method is used for solving a hybrid optimization problem that consists of l_1-norm minimization and nuclear norm minimization: To relax the objective function in (13), we replace the l_0-norm with the l_1-norm, and the rank with the matrix nuclear norm $\|\cdot\|_*$, respectively. Candes et al. [13] proved that, under surprisingly broad conditions, "almost all" matrices of the form $X = A + E$, namely, matrices that are the sums of a low-rank matrix A and a sparse matrix E, can be exactly decomposed into these two components by solving the following convex optimization problem:

$$\min_{A,E}\|A\|_* + \lambda\|E\|_1 \quad s.t.\ A + E = X, \tag{14}$$

where $\lambda > 0$ is a regularization parameter. In addition, they showed that $\lambda = 1/\sqrt{\max(t,p)}$ is a proper choice for the parameter that is independent from A and E, and we follow this choice throughout the rest of the paper.

The standard RPCA problem assumes that X is strictly equal to the sum of a low-rank matrix and a sparse matrix. However, in many real world applications, observational data often contains certain level of noise, and it usually pollutes almost all the entries of the matrix. One of the most common perturbations is the Gaussian white noise, which leads to the generalized RPCA problem with Gaussian noise:

Problem 2 (*generalized RPCA problem with Gaussian noise*). Suppose that $X = A + E + N$, where A, E and N are unknown matrices. Assume that A is a low-rank matrix, E is a sparse matrix, and N is a random matrix whose entries follow i.i.d. zero-mean Gaussian distributions with $\|N\|_F < \delta$ for some positive δ. The generalized RPCA problem is to recover A and E from X under the perturbation of N.

For Problem 2, Zhou et al. [14] generalized the PCP method to propose the Relaxed PCP method, which is to solve the following optimization problem:

$$\min_{A,E}\|A\|_* + \lambda\|E\|_1 \quad s.t.\ \|X - A - E\|_F \leqslant \delta. \tag{15}$$

They proved that, under the same conditions as that PCP requires, for any realization of the Gaussian noise satisfying $\|N\|_F < \delta$, the solution to the generalized RPCA problem (15) gives a stable estimation of A and E with high probability.

The assumptions in the generalized RPCA problem are similar to the hypotheses on the traffic matrix in our decomposition model (12). In fact, we can multiply the columns of the traffic matrix by some constants to make sure that the Gaussian random variables in the noise traffic matrix have the same variance. Meanwhile, this multiplication preserves the rank of the deterministic traffic matrix and the sparsity of the anomaly traffic matrix. Therefore, the Relaxed PCP method can be used for solving our structural analysis problem.

3.3. The accelerated proximal gradient algorithm

The Relaxed PCP method used for solving the constrained optimization problem (15) is usually computationally expensive. A more efficient way is to solve an equivalent unconstrained optimization problem instead, using algorithms such as Iterative Thresholding (IT) [13], Augmented Lagrange Multiplier (ALM) and Accelerated Proximal Gradient (APG) [14]. In this paper, we adopt the APG algorithm, which solves the following unconstrained minimization problem:

$$\min_{A,E} \mu\|A\|_* + \mu\lambda\|E\|_1 + \frac{1}{2}\|X - A - E\|_F^2, \tag{16}$$

where $\frac{1}{2}\|X - A - E\|_F^2$ is the penalty function, and $\mu > 0$ is a Lagrangian parameter. It has been proved in [14] that with some proper choices of $\mu = \mu(\delta)$, the solution to (16) is equivalent to the solution to (15).

As mentioned above, the choice of the regularization parameter λ follows that in [13,14]:

$$\lambda = \frac{1}{\sqrt{\max(t,p)}}. \tag{17}$$

For the Lagrangian parameter μ, it is chosen as $\sqrt{2\max(t,p)}\sigma$ and $(\sqrt{t} + \sqrt{p})\sigma$ in [14,22], respectively, where σ is the variance of Gaussian noise matrix N. These choices are motivated by neglecting the effect of the sparse matrix E: if we set $E = 0$ in problem (16), the APG algorithm which solves this problem boils down to the *Singular Value Thresholding* algorithm with total sampling [14,23]. In our case, since the anomaly traffic matrix might contribute a large proportion of energy, these choices are not suitable. Therefore we present a new choice of μ in this paper. Considering the case when $A = 0$, problem (16) boils down to:

$$\min_{A,E} \mu\lambda\|E\|_1 + \frac{1}{2}\|X - E\|_F^2. \tag{18}$$

If we consider X and E as two column vectors of dimension $t \times p$, problem (18) becomes the *Basis Pursuit Denoising* problem first introduced in [25,26]. As E is a sparse vector, we follow [25] to choose $\mu\lambda = \sigma\sqrt{2\log(tp)}$. Since λ is chosen as in (17), we compute μ accordingly as:

$$\mu = \sigma\sqrt{2\log(tp)\max(t,p)}. \tag{19}$$

For each OD flow time series $X_j \in \mathbb{R}^t$ $(1 \leqslant j \leqslant p)$, we need to estimate the variance σ_j of the Gaussian noise traffic. This is a well-studied signal processing problem. We adopt the estimation method proposed in [27]: Given an orthonormal wavelet basis, and let $W_j = \{a_k^j\}_{k=1}^{t/2}$ denote X_j's wavelet coefficients at the finest scale, σ_j is estimated as the median absolute deviation of W_j divided by 0.6745:

$$\sigma_j = \frac{1}{0.6745}\text{median}\left\{\left|a_k^j - \text{median}(W_j)\right|\right\}, \tag{20}$$

where median $(\cdot)$ denotes the median value of a vector. This estimation method is motivated by the empirical fact that, wavelet coefficients at the finest scale are, with few exceptions, essentially pure noise. In this paper, we adopt the Daubechies-5 wavelet basis.

Now we are ready to present the proposed Algorithm 1 (see Appendix) for traffic matrix decomposition, which is partially based on the noisy-free version of the APG algorithm in [18].

4. Experiments

We decompose the twelve traffic matrices described in Section 2 using Algorithm 1 (X01–X08 are from the Abilene dataset and Y01–Y04 are from the GEANT dataset). The detailed experimental results are summarized in Table 3, where each row corresponds to the decomposition result for one traffic matrix. From left to right, the columns of Table 3 represent:

(1) Name of the Traffic matrix.
(2) The rank of the deterministic traffic matrix.
(3) The rank of the original traffic matrix.
(4) The ratio of (2) to (3), to evaluate the relative low-rank degree of the deterministic traffic matrix.
(5) The l_0-norm of the anomaly traffic matrix.
(6) $t \times p$, where t and p are the number of rows and columns of the traffic matrix, respectively.
(7) The ratio of (5) to (6), to evaluate the relative sparsity level of the anomaly traffic matrix.
(8) The ratio of Frobenius norm of the noise traffic matrix to that of the corresponding original traffic matrix, to evaluate how much energy is contained in the noise traffic matrix.
(9) Number of iterations implemented in the APG algorithm for each traffic matrix. In this paper, the tolerance parameter for the stopping criterion is set to 10^{-6}.
(10) Computational time of the implementation of the APG algorithm for each traffic matrix (in seconds). In all the experiments, we use a commercial PC with 2.0 GHz Intel Core2 CPU and 2.0 GB RAM.

All the traffic matrices in our experiments are decomposed into three sub-matrices. According to the results in Table 3, these sub-matrices indeed satisfy the hypotheses of the decomposition model (12):

1. The ranks of the deterministic traffic matrices in the Abilene dataset are less than 13; for the GEANT dataset, the ranks are less than 31. The rank of each deterministic traffic matrix is less than 11% of the rank of the corresponding original traffic matrix. Therefore, all the decomposed deterministic traffic matrices are typical low-rank matrices.
2. In both datasets, the l_0-norm of one anomaly traffic matrix does not exceed 40000. For each anomaly traffic matrix, less than 17% entries are non-zero entries. Therefore, all the anomaly traffic matrices are typical sparse matrices.
3. For all the twelve traffic matrices used in our experiments, the ratio of the Frobenius norm between the noise traffic matrix and the original traffic matrix is less than 0.18. Hence, noise traffic matrices usually contribute only a small proportion of the total energy.

Table 3
Traffic matrix decomposition results using Relaxed PCP.

Traffic matrix	Rank (A)	Rank (X)	Rank (A)/rank (X)	$\|E\|_0$	$t \times p$	$\|E\|_0/(t \times p)$	$\|N\|_F/\|X\|_F$	# Iteration	Computation time (s)
X01	10	121	0.0826	32766	243936	0.1343	0.1718	86	24
X02	11	121	0.0909	34713	243936	0.1423	0.1483	209	56
X03	12	121	0.0992	37280	243936	0.1528	0.0824	254	70
X04	11	121	0.0909	30519	243936	0.1251	0.0395	663	168
X05	10	121	0.0826	30878	243936	0.1266	0.1173	86	31
X06	10	121	0.0826	31133	243936	0.1276	0.0562	118	44
X07	13	121	0.1074	37463	243936	0.1536	0.0887	170	64
X08	12	121	0.0992	39287	243936	0.1611	0.0564	155	60
Y01	31	483	0.0642	31505	324576	0.0971	0.1114	168	350
Y02	28	465	0.0602	28575	312480	0.0914	0.1330	171	390
Y03	30	465	0.0645	28651	312480	0.0917	0.1229	146	312
Y04	30	457	0.0656	29119	307104	0.0948	0.0752	219	489

In addition, the number of iterations and the computational time needed in the implementation of the APG algorithm are quite acceptable. Specifically, for all the traffic matrices in the Abilene dataset, the computational time is less than three minutes; the average computational time for the GEANT traffic matrices is longer, but still less than nine minutes.

To further understand the experimental results, we focus on the decomposition results of three traffic matrices: X01, X04 and Y01. Fig. 4 shows the traffic matrix X01 (upper left) in the Abilene dataset, and the three sub-matrices decomposed by the Relaxed PCP method. We can see that the deterministic traffic matrix (upper right) contributes most of the energy of the total network traffic, and the columns (corresponding to the deterministic traffic time series) show clear diurnal pattern, especially for OD flows with large magnitude. Most of the large volume anomalies in X01 are short-lived and well isolated in the anomaly traffic matrix (bottom left). This matrix is indeed quite sparse, without distinct periodical traffic or noise traffic. Furthermore, most of the entries in the noise traffic matrix (bottom right) have small magnitudes, therefore the noise traffic indeed contributes little energy to the total network traffic. For most of the OD flows, the variances of the noise traffic are proportional to their mean volume, that is, an OD flow with large magnitude usually has large noise traffic. However, there also exist some OD flows of moderate magnitude which contain very large noise traffic, which will be discussed in Section 5.2. In summary, using the Relaxed PCP proposed described in Section 3.2, we achieve a proper decomposition for the traffic matrix X01, where all the sub-matrices satisfy the corresponding hypotheses in the decomposition model (12).

For comparison, we then decompose X01 using the PCA-based subspace method. Recall that in Figs. 2 and 3, the first four eigenflows satisfy the criteria of both d-eigenflow and

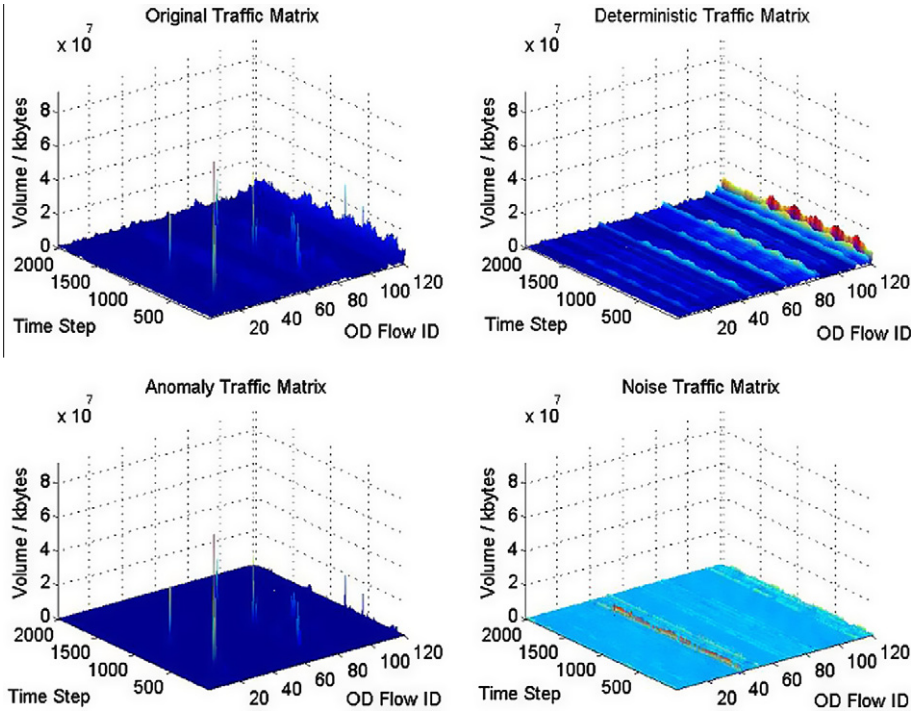


Fig. 4. The decomposition result of X01 by Relaxed PCP.

s-eigenflow, and they contribute most of the energy of X01. Therefore, the key is the classification results for these four eigenflows. Suppose that for $i \in \{1, 2, 4\}$, we classify the first i eigenflow (s) as d-eigenflow, and the rest (if exists) in the first four eigenflows as s-eigenflow. The normal traffic matrix (projection onto the normal subspace) is then generated by the first i d-eigenflow (s), and the residual traffic matrix is the difference between the original traffic matrix and the normal traffic matrix. In other words, the normal traffic matrix and the residual traffic matrix constitute a decomposition of X01 For each choice of i, the resulting normal traffic matrix and residual traffic matrix are shown in each row of Fig. 5. We can see that for $i = 1$, most of the large anomaly traffic is isolated in the residual traffic matrix, while the normal traffic matrix only captures partially the deterministic traffic in X01. In other words, the residual traffic matrix also contains a large proportion of the diurnal traffic. For $i = 2$, although more diurnal traffic is present in the normal traffic matrix, this matrix still contains large anomaly traffic. Thus we cannot efficiently identify large volume anomalies from the residual traffic matrix. For $i = 4$, since the fifth eigenflow is not a s-eigenflow and the sixth is one, the normal traffic matrix is generated based on the first five eigenflows of X01. In this case, the decomposition is even worse, since most of the anomaly traffic is contained in the normal traffic matrix. This again demonstrates that, for traffic matrices with large volume anomalies, PCA is not a suitable method for the structural analysis problem. Furthermore, we believe that its performance cannot be significantly improved only by changing parameters or using heuristic mechanisms.

Fig. 6 shows the traffic matrix X04 in the Abilene dataset, as well as three sub-matrices decomposed by the Relaxed PCP method (they are arranged in the same way as Fig. 4). A significant difference between X04 and X01 is that the former one contains long-lived large volume anomalies. Therefore, the anomaly traffic in X04 contributes a larger proportion of the total energy than that in X01. Using the same ways to compute parameters λ and μ in Algorithm 1, our experimental result in Fig. 6 shows that the Relaxed PCP method can also exactly decompose a traffic matrix with long-lived large volume anomalies.

Fig. 7 displays the decomposition result of the traffic matrix Y01 in the GEANT dataset. In general, the result is

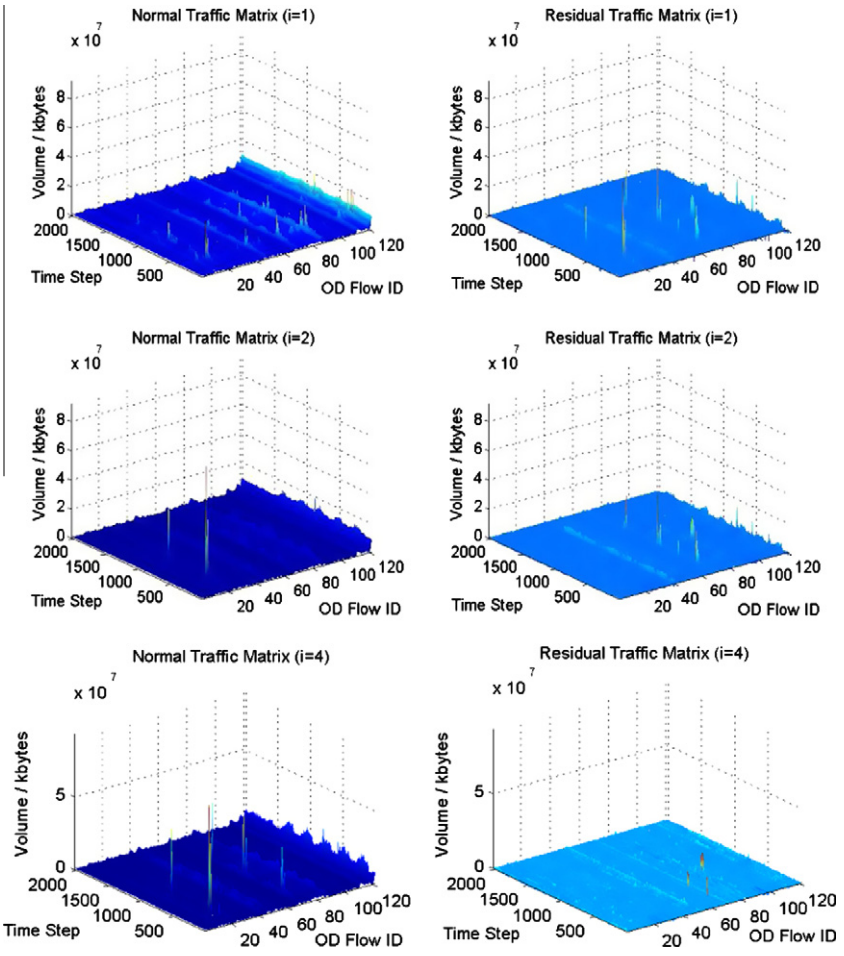


Fig. 5. The decomposition result of X01 by PCA.

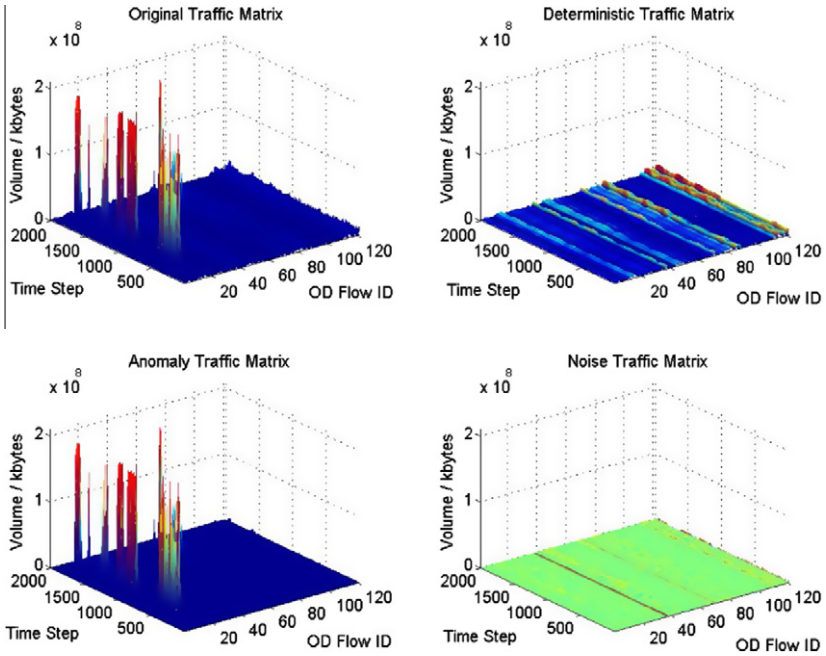


Fig. 6. The decomposition result of X04 by Relaxed PCP.

similar to that of X01 and X04. However, since the GEANT traffic matrices usually contain more unstable OD flows than the Abilene traffic matrices, the periodical traffic pattern shown in the resulting deterministic traffic matrix is less obvious compared to results of X01 and X04.

5. Discussions

According to the traffic matrix decomposition model (12), we decompose the traffic matrix into three submatrices, which correspond to three classes of network traffic. Based on the experimental results obtained, we now have further discussions on the deterministic traffic matrix and the noise traffic matrix in this section (We do not discuss the anomaly traffic matrix since it may vary significantly for different input traffic matrices).

5.1. Non-periodical traffic in the deterministic traffic matrix

As shown in Section 4, for each traffic matrix in our experiments (X01–X08 and Y01–Y04), the deterministic traffic matrix decomposed by Algorithm 1 has a low rank compared to the corresponding OD flow number. In most cases, columns of the deterministic traffic matrix (deterministic traffic time series) display significant diurnal pattern. However, there also exist several columns that contain traffic changes, which are quite different from the periodical traffic. This observation is quite obvious for the Abilene traffic matrices X03 and X07.

As an example, Fig. 8 displays X03 and the decomposed deterministic traffic matrix. Clearly, we observe a few long-lived traffic changes in the deterministic traffic matrix. These traffic changes affect a few columns (traffic time series) with long-lived growth or decline in terms of traffic volume, and such growths and declines usually happen during the same time intervals. In particular, we illustrate in Fig. 9 eleven affected time series, which have the same source router "WASH". As we can see, one traffic growth and ten traffic declines all happen during the time intervals [1150; 1450], which share the same starting and ending time. In fact, more than 20 time series in the deterministic traffic matrix are significantly affected, but their source and destination routers do not present clear distribution laws. These traffic changes have not been reported in the previous studies. Therefore, it seems that the deterministic traffic matrix may contain non-periodical traffic changes, which are usually combinations of long-lived traffic growths and declines during the same time intervals. These changes can hardly be judged as any of the well known volume anomalies such as DoS/DDoS, flash crowd, alpha, outages and ingress/egress shift [9]. Since the Abilene traffic dataset only records OD flows' coarse-gained byte counts during every five-minute time interval, and we do not have more detailed information about the network when these traffic changes happen, it is difficult to explain the reason for these long-lived traffic changes. We leave this for future work.

In addition, we illustrate in Fig. 10 the sum of the aforementioned eleven OD flow time series in X03, as well as the sum of the corresponding deterministic traffic series (eleven columns of the deterministic traffic matrix of X03), both with the same source router "WASH". We can see that the sum of the OD flows contains some short-lived large traffic growths during the time intervals [1150, 1450], while these needle-like traffic growths cannot be observed in the sum of the deterministic traffic series. In fact, the latter sum presents typical periodical pattern during the whole week. This shows that, although individual deter-

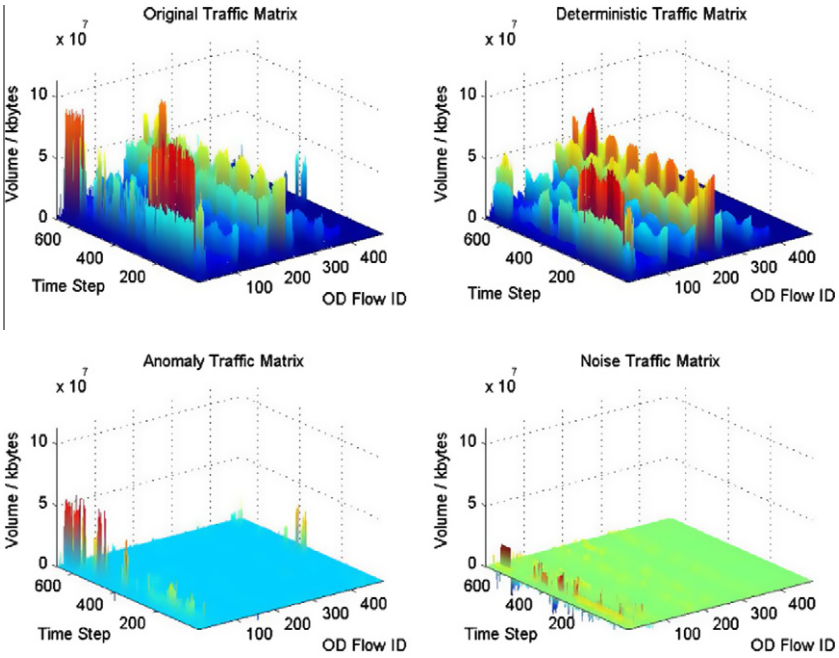


Fig. 7. The decomposition result of Y01 by Relaxed PCP.

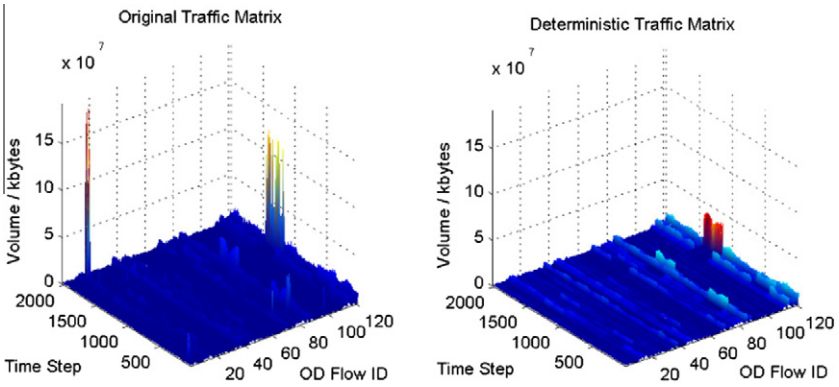


Fig. 8. The traffic matrix X03 (left) and the associated deterministic traffic matrix (right).

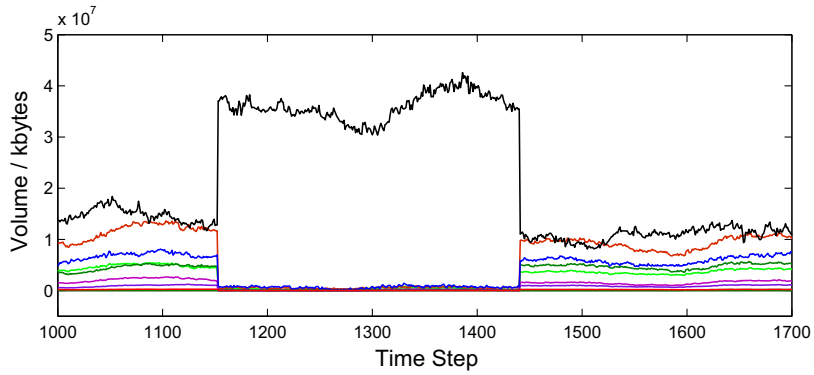


Fig. 9. Eleven columns of X03 (deterministic traffic time series) with the same source router "WASH".

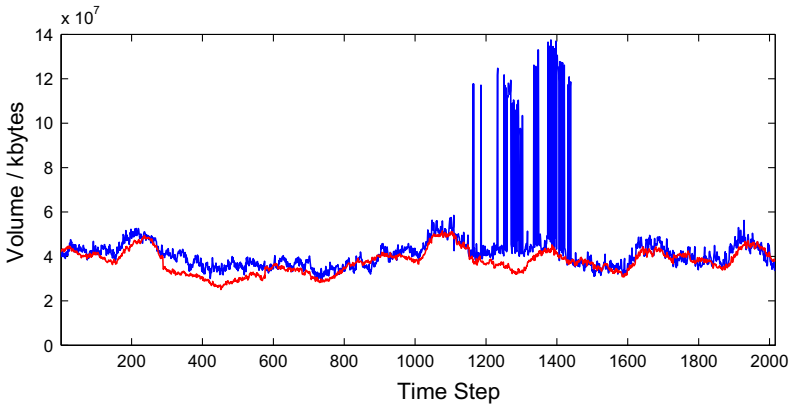


Fig. 10. The sum of eleven OD flow time series (blue) and the sum of the corresponding deterministic traffic series (red), both with the same source router "WASH". (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

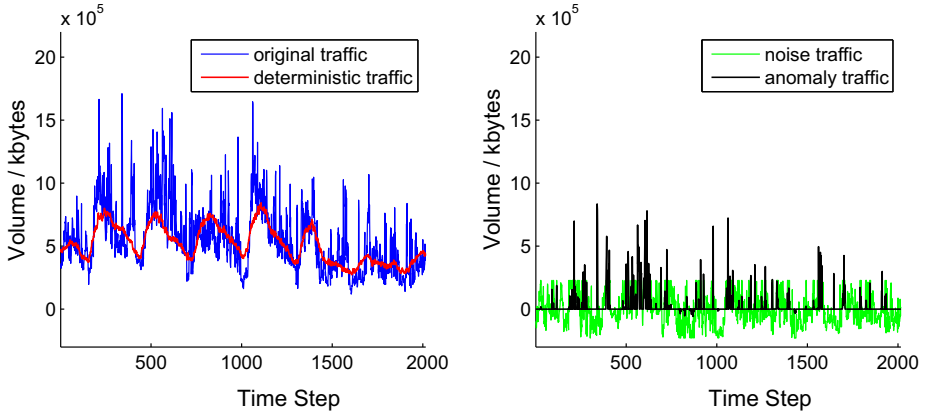


Fig. 11. The decomposition of OD flow time series No. 50 in X01. Blue: original traffic; Red: deterministic traffic; Black: anomaly traffic; Green: noise traffic. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

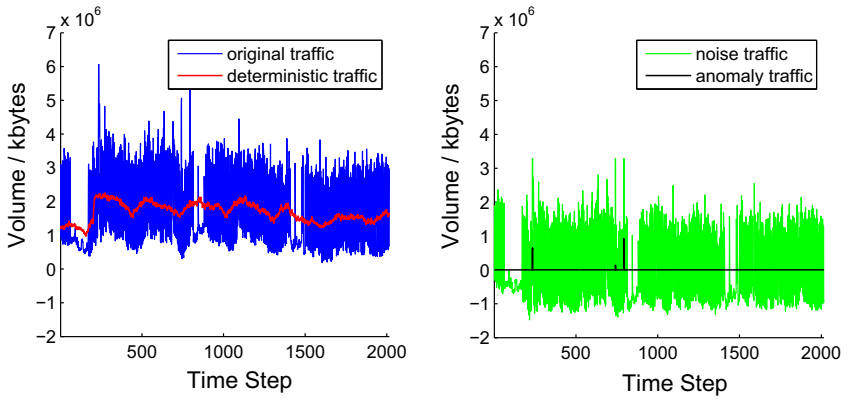


Fig. 12. The decomposition of OD flow time series No. 51 in X01. Colors indicate the same classes of traffic as that in Fig. 11.

ministic traffic series with the same source router may contain significant traffic changes, the sum of them tends to show expected patterns. As a result, if we consider the total network traffic with the source router "WASH" (which is the sum of eleven OD flows), the anomaly traffic component can be well decomposed by the Relaxed PCP method.

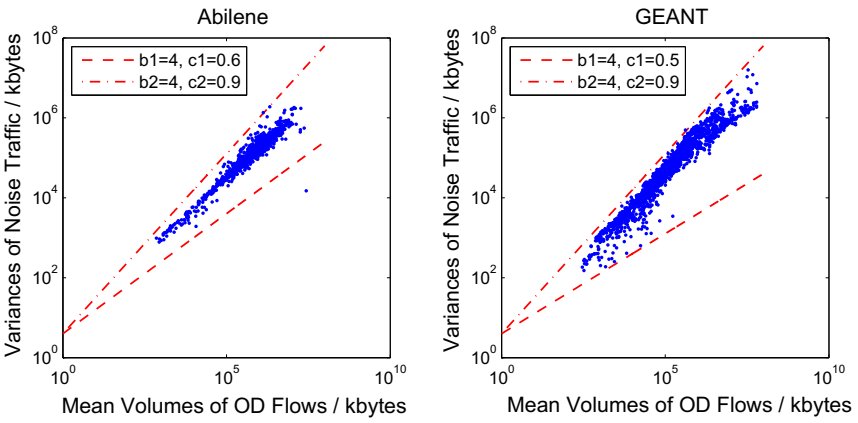


Fig. 13. The relationship between variance of the noise traffic and the mean volume of the corresponding OD flow. Left: the Abilene networks; Right: the GEANT networks.

5.2. Some features of the noise traffic matrix

5.2.1. The proportion of noise traffic in different OD flows

As can be seen from Table 3 in Section 4, noise traffic matrices contribute a small proportion of the total network traffic. However, we observe that the ratios of the noise traffic to the total traffic vary in different OD flows. For instance, Figs. 11 and 12 illustrate decompositions of two OD flows in the Abilene traffic matrix X01, namely OD flow No. 50 and No. 51, respectively. More specifically, for OD flow No. 50, the total[1] traffic time series (blue) is mainly contributed by the deterministic traffic time series (red) and the anomaly traffic time series (black), and the noise traffic time series (green) has much smaller average magnitude. Therefore, we conclude that the noise traffic is not an important component in OD flow No. 50. In fact, this is the case for most of the OD flows.

However, a small number of OD flows have different decomposition results. For example, for OD flow No. 51, the noise traffic time series has quite large average magnitude compared to the original traffic time series, therefore the noise traffic becomes a significant component for this OD flow hence should not be neglected in the analysis. Actually, this OD flow contains large-amplitude oscillations, which is not a common feature for all the OD flows, and it should be classified as the noise traffic in the network.

To summarize, although noise traffic time series are usually small in magnitude, they cannot be neglected in the analysis of a few OD flows which contain large oscillations.

5.2.2. The variance of noise traffic time series

The energy (variance) of the noise traffic may vary significantly in different OD flows. Suppose that X is a traffic matrix, and we compose it as $X = A + E + N$ by Algorithm 1. For each OD flow time series X_j (one column of X), we are interested in the relationship between variance $Var(N_j)$ of the noise traffic (which is estimated by the standard deviation of noise traffic time series N_j) and statistics of X_j. For example, Fig. 13 illustrates the relationship between $Var(N_j)$ and mean (X_j). Specifically, for each data point in the figure, the horizontal axis represents the mean value of an OD flow time series, and the vertical axis represents the variance of the noise traffic of the same OD flow. Here we have analyzed all the OD flows in our datasets, where time series of the same OD flow in different weeks are considered as different data points. Therefore, we have $121 \times 8 = 968$ data points for the Abilene dataset, and 1870 for the GEANT dataset.

It is clear from Fig. 13 that there is a strong positive correlation between the mean volume of OD flows and the variance of the corresponding noise traffic. In the log–log plot, the distribution of the data points follows a weak linear relationship, and such relationship is more noticeable for the Abilene dataset. Therefore, it is reasonable to assume that in most cases the variance of the noise traffic of an OD flow can be approximated by a power function of the mean volume of the OD flow, which can be written as

$$Var(N_j) \approx b\ \mathrm{mean}(X_j)^c, \tag{21}$$

where b and c are some positive parameters. Notice that there exist many mathematical methods for the estimation of the parameters b and c; However, this is beyond the scope of the current study and we leave it for future work. Instead, we propose empirical bounds for the variance of the noise traffic for the two datasets, which are two parameter pairs $(b1, c1)$ and $(b2, c2)$ satisfying

$$b1\ \mathrm{mean}(X_j)^{c1} \leqslant Var(N_j) \leqslant b2\ \mathrm{mean}(X_j)^{c2}. \tag{22}$$

As labeled in Fig. 13, for the Abilene dataset, the choices $b1 = b2 = 4$, $c1 = 0.6$ and $c2 = 0.9$ seem to work well for most of the data points except a few outliers; for the GEANT dataset, $b1 = b2 = 4$, $c1 = 0.5$ and $c2 = 0.9$ are the reasonable choices.

In addition, we have also analyzed the relationships between the variance of the noise traffic and several other statistics of the corresponding OD flow, such as the l_2-norm, the median value, and the variance of the OD flow.

[1] For interpretation of color in Figs. 1–14, the reader is referred to the web version of this article.

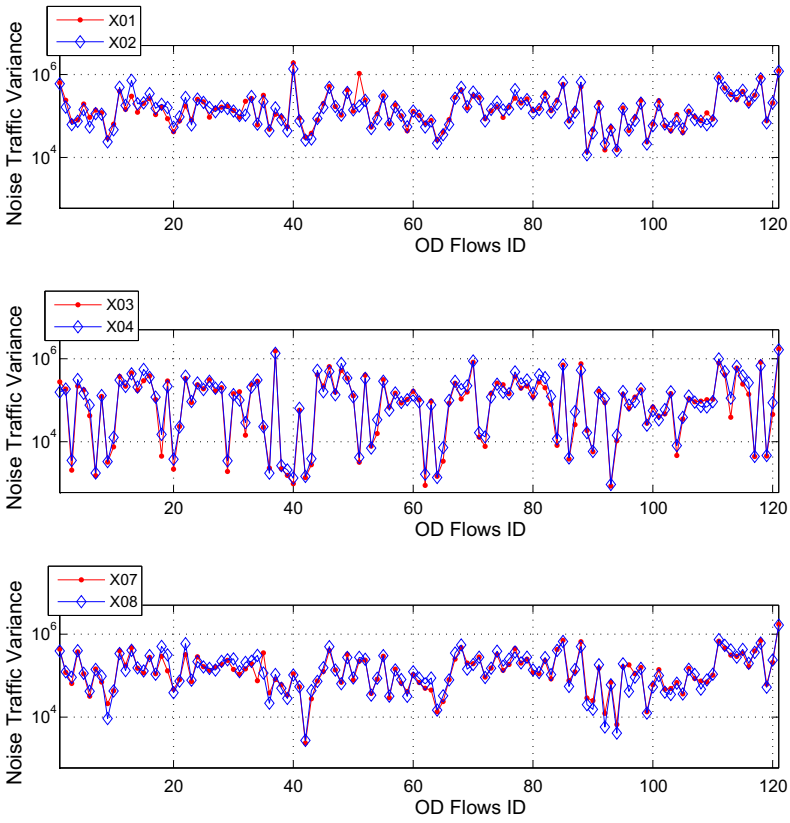


Fig. 14. The pair of variances of the noise traffic for two consecutive weeks.

For all of them, we have observed the positive correlation between the two as well, but not as significant as the correlation between the noise variance and the mean volume of the flow. In this case, it is less obvious to find an explicit mathematical model for the correlation as Eq. (21).

Finally, it is interesting to study the temporal stability of variances of the noise traffic during different weeks. Suppose we have two traffic matrices that record the traffic volume of the same network for two consecutive weeks. By decomposing them using the APG algorithm independently, we first obtain two noise traffic matrices, one for each week. Recall that each column vector of a noise traffic matrix represents the noise traffic time series of an OD flow. For each OD flow, we then compare the pair of variances of the corresponding noise traffic for the two consecutive weeks. Specifically, we choose traffic matrices for three pairs of consecutive weeks in the Abilene dataset:

- X01 (from March 1, 2004) and X02 (from March 8, 2004).
- X03 (from April 2, 2004) and X04 (from April 9, 2004).
- X07 (from May 8, 2004) and X08 (from May 15, 2004).

Since each Abilene traffic matrix contains 121 OD flows (columns), the variances of the corresponding noise traffic include 121 data points. Fig. 14 shows the variances of the noise traffic for the six chosen traffic matrices ($121 \times 6 = 726$ data points in total).

We observe the following features in Fig. 14: (1) For each traffic matrix, the variances of the noise traffic of different OD flows may vary significantly; (2) For traffic matrices of two consecutive weeks, the variances of the noise traffic of the same OD flow are similar in most cases; (3) For two traffic matrices that do not represent two consecutive weeks, the variances of the noise traffic of the same OD flow may vary significantly (Take the same OD flows in X01 and X03 as examples).

However, our observations are not sufficient to conclude that the variance of the noise traffic is strictly stable, the reason being that: (1) The analysis above is not comprehensive enough as our datasets do not contain traffic matrices for many consecutive weeks; (2) There also exist a few variance pairs in which one is obviously different from the other, although they correspond to the same OD flow for the two consecutive weeks. We plan to investigate the unstable variances of the noise traffic in future work.

6. Conclusions

In this paper, we focus on the structural analysis of the traffic matrix that has been polluted by large volume

anomalies. We first demonstrate that the PCA-based analysis method performs poorly for polluted traffic matrices. Next, we propose a new decomposition model for the traffic matrix that is more practical in the analysis of empirical network traffic data, and study the decomposition problem using the relaxed principal component pursuit method. Finally, we discuss the experimental results in more details for the deterministic and noise traffic matrix. The major findings in this paper are:

1. We experiment the classical PCA method for traffic matrix analysis. Different from the previous works [4,5], the traffic matrices that we analyze contain some large volume anomalies. In this case, our results show that the eigenflow classification is neither complete nor orthogonal, which suggests that PCA is unable to decompose accurately the traffic matrix into the normal traffic matrix and the large anomaly traffic matrix.
2. Based on the empirical network traffic data, a new decomposition model for the traffic matrix is proposed in Section 3.1. To the best of our knowledge, it is a novel way of formalizing the structure of the traffic matrix, which also provides a simple view of the traffic matrix analysis problem. Moreover, this model helps explain intuitively some of the limitations of the classical PCA method in our experiments.
3. According to the decomposition model of the traffic matrix, we show that the problem of traffic matrix decomposition is equivalent to the robust PCA problem, which has been extensively studied recently. Based on the relaxed principal component pursuit method and the accelerated proximal gradient algorithm, we develop an algorithm for the decomposition of traffic matrices that may contain large volume anomalies. The experimental results demonstrate the efficiency and flexibility of the proposed algorithm.
4. We discuss some detailed features of the deterministic traffic matrix and the noise traffic matrix. Firstly, we observe that the deterministic traffic matrix may contain non-periodical traffic changes, which are usually combinations of long-lived traffic growths and declines during the same time intervals. Secondly, although the noise traffic matrix contributes a small proportion of the total network traffic in general, the ratios of the noise traffic to the total traffic may vary significantly in different OD flows. Thirdly, we find that there exists significant positive correlation between the mean volume of OD flow and the variance of the noise traffic time series, and we further test the temporal stability of the variance of the noise traffic.

To summarize, this paper is a preliminary study on applying the Relaxed PCP method for network traffic analysis, whose efficiency and flexibility have been demonstrated in the experimental results. For future work, we plan to further optimize the Relaxed PCP method to make it adaptable to the network traffic data, and explore its applications in volume anomaly detection and data cleaning for the polluted traffic matrix.

Acknowledgments

We thank Professor Jinping Sun at Beihang University for his advices on the earlier draft of this paper. We also thank Professor Pascal Frossard at Ecole Polytechnique Fédérale de Lausanne (EPFL) for his help with the revised version of the paper. Finally, we are grateful to the anonymous reviewers for their constructive suggestions on the paper. X. Dong is supported by Nokia Research Center (NRC), Lausanne, Switzerland.

Appendix A

In this appendix, we present the APG algorithm for traffic matrix decomposition.

Algorithm 1. APG for traffic matrix decomposition

Input: the traffic matrix $X \in \mathbb{R}^{t\times p}$.

1. Compute the regularization parameter λ using (17).
2. Compute the Lagrangian parameter μ using (19) with $\sigma = 1$.
3. For each OD flow time series X_j, estimate the variance σ_j of its Gaussian noise component using (20).
4. Let $X = X/diag\{\sigma_j\}$.
5. Let $A_0 = A_{-1} = 0$, $E_0 = E_{-1} = 0$, $t_0 = t_{-1} = 1$, $S_1^A = S_1^E = 1$ and $k = 0$.
6. **while** not converged **do**

$$Y_k^A = A_k + \frac{t_{k-1}-1}{t_k}(A_k - A_{k-1});$$
$$Y_k^E = E_k + \frac{t_{k-1}-1}{t_k}(E_k - E_{k-1});$$
$$G_k^A = Y_k^A - \frac{1}{2}\left(Y_k^A + Y_k^E - X\right);$$
$$G_k^E = Y_k^E - \frac{1}{2}\left(Y_k^A + Y_k^E - X\right);$$
$$(U,S,V) = SVD\left(G_k^A\right);$$
$$A_{k+1} = U\mathcal{S}_{\frac{\mu}{2}}[S]V^T;$$
$$E_{k+1} = \mathcal{S}_{\frac{\lambda\mu}{2}}\left[G_k^E\right];$$
$$t_{k+1} = \frac{1+\sqrt{4t_k^2+1}}{2};$$
$$S_{k+1}^A = 2\left(Y_k^A - A_k\right) + \left(A_{k+1} + E_{k+1} - Y_k^A - Y_k^E\right);$$
$$S_{k+1}^E = 2\left(Y_k^E - E_k\right) + \left(A_{k+1} + E_{k+1} - Y_k^A - Y_k^E\right);$$
$$k = k + 1;$$

   end while
7. Let $X = X \cdot diag\{\sigma_j\}$.

Output:
$A = A_k \cdot diag\{\sigma_j\}$; $E = E_k \cdot diag\{\sigma_j\}$; $N = X - A - E$.

In Algorithm 1, $\mathcal{S}_\varepsilon[\cdot] : \mathbb{R}^{t\times p} \to \mathbb{R}^{t\times p}$ represents the soft-thresholding operator with parameter $\varepsilon > 0$. $\forall X \in \mathbb{R}^{t\times p}$, $\mathcal{S}_\varepsilon[X] \in \mathbb{R}^{t\times p}$ and it satisfies

$$\mathcal{S}_\varepsilon[X](i,j) = \begin{cases} X(i,j) - \varepsilon & \text{if } X(i,j) > \varepsilon \\ X(i,j) + \varepsilon & \text{if } X(i,j) < -\varepsilon \\ 0 & \text{otherwise} \end{cases}. \tag{23}$$

We choose the stopping criterion of Algorithm 1 as the one defined in [18], which terminates the iterations when the quantity $\left\|S^A_{k+1}\right\|_F^2 + \left\|S^E_{k+1}\right\|_F^2$ is less than a pre-defined tolerance parameter.

References

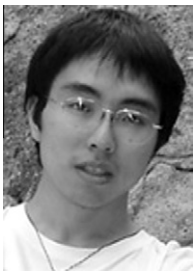
[1] S. Uhlig, B. Quoitin, J. Lepropre, S. Balon, Providing public intradomain traffic matrices to the research community, SIGCOMM Comput. Commun. Rev. 36 (1) (2006) 83–86.
[2] Abilene data, <http://www.cs.utexas.edu/yzhang/research/AbileneTM/>.
[3] Geant data, <http://totem.info.ucl.ac.be/dataset.html>.
[4] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E.D. Kolaczyk, N. Taft, Structural analysis of network traffic flows, SIGMETRICS Perform. Eval. Rev. 32 (1) (2004) 61–72.
[5] A. Lakhina, M. Crovella, C. Diot, Diagnosing network-wide traffic anomalies, SIGCOMM Comput. Commun. Rev. 34 (4) (2004) 219–230.
[6] L. Huang, X. Nguyen, M. Garofalakis, M. Jordon, A. Joseph, N. Taft, In-network PCA and anomaly detection, in: Proceedings of Neural Information Processing Systems (NIPS) 2006, December 2006.
[7] Y. Zhang, Z. Ge, A. Greenberg, M. Roughan, Network anomography, in: Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement (IMC '05), USENIX Association, Berkeley, CA, USA, 2005, pp. 317–330.
[8] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, C. Diot, Traffic matrices: balancing measurements, inference and modeling, SIGMETRICS Perform. Eval. Rev. 33 (1) (2005) 362–373.
[9] A. Soule, K. Salamatian, N. Taft, Combining filtering and statistical methods for anomaly detection, in: Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement (IMC '05), USENIX Association, Berkeley, CA, USA, 2005, pp. 331–344.
[10] H. Ringberg, A. Soule, J. Rexford, C. Diot, Sensitivity of PCA for traffic anomaly detection, SIGMETRICS Perform. Eval. Rev. 35 (1) (2007) 109–120.
[11] Y. Ohsita, S. Ata, M. Murata, Identification of attack nodes from traffic matrix estimation, IEICE Trans. Commun. E90-B (10) (2007) 2854–2864.
[12] B. Rubinstein, B. Nelson, L. Huang, A. Joseph, S. Lau, S. Rao, N. Taft, J. Tygar, ANTIDOTE: understanding and defending against poisoning of anomaly detectors, in: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference (IMC '09), ACM, New York, NY, USA, 2009, pp. 1–14.
[13] E. Candes, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, Arxiv preprint, arXiv:0912.3599, 2009.
[14] Z. Zhou, X. Li, J. Wright, E. Candes, Y. Ma, Stable principal component pursuit, in: Proceedings of IEEE International Symposium on Information Theory (ISIT) 2010, June 2010.
[15] Y. Peng, A. Ganesh, J. Wright, W. Xu, Y. Ma, RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010, June 2010.
[16] K. Min, Z. Zhang, J. Wright, Y. Ma, Decomposing background topics from keywords using principal component pursuit, in: Proceedings of ACM International Conference on Information and Knowledge Management (CIKM) 2010, October 2010.
[17] J. Wright, A. Ganesh, S. Rao, Y. Peng, Y. Ma, Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization, in: Proceedings of Neural Information Processing Systems (NIPS) 2009, December 2009.
[18] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, Y. Ma, Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix, in: Proceedings of IEEE 3rd International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), December 2009.
[19] G. Stewart, On the early history of the singular value decomposition, SIAM Rev. 35 (4) (1993) 551–566.
[20] E. Candes, M. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 21–30.
[21] E. Candes, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (6) (2009) 717–772.
[22] E. Candes, Y. Plan, Matrix completion with noise, Proc. IEEE 98 (6) (2010) 925–936.
[23] J. Cai, E. Candes, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optimiz. 20 (4) (2010) 1956–1982.
[24] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory. 52 (4) (2006) 1289–1306.
[25] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[26] D. Donoho, De-noising by soft-thresholding, IEEE Trans. Inform. Theory 41 (3) (1995) 613–627.
[27] D. Donoho, I. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 14 (6) (1994) 425–455.
[28] I. Cunha, F. Silveira, R. Oliveira, R. Teixeira, C. Diot, Uncovering artifacts of flow measurement tools, in: Proceedings of Passive and Active Measurement Conference. Seoul, Korea, April 2009.

Zhe Wang is currently a Ph.D. candidate in the School of Computer Science and Engineering, Beihang University. He received his B.S degrees in Mathematics from the School of Mathematics and System Science, Beihang University, in 2007. His research interests include network measurement and analysis, parallel computing.

Kai Hu is currently an Associate Professor in the School of Computer Science and Engineer of Beihang University. He received Ph.D. degree in 2001 from the School of Computer Science and Engineering, Beihang university. From 2001 to 2003, He was a Research Fellow in the School of Computer Engineering, Nanyang Technological University, Singapore. Dr. Kai Hu's research interests include high performance parallel computing, network technology, and embedded system.

Ke Xu is currently a Professor in the School of Computer Science and Engineer of Beihang University. He received a Ph.D. degree in computer science from Beihang University in 2000. His main research interests include algorithm and computational complexity, complex networks, network measurement and data mining.

Baolin Yin is currently a Professor in the School of Computer Science and Engineer of Beihang University. He graduated from the Department of Electronic Engineering, Beihang University, in 1973, and received his Ph.D. degree from the Department of Artificial Intelligence, the University of Edinburgh, in 1984. His research interests include distributed applications and image processing.



Xiaowen Dong has been working as a PhD student in the Signal Processing Laboratories (LTS4 / LTS2) at Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland since October 2009. His research interests include wavelets, spectral graph theory, and their applications to mobile and online social network analysis.

Before joining EPFL, Mr. Dong received his B.Eng. degree in Information Engineering from Zhejiang University, Hangzhou, China and his M.Sc. degree in Signal Processing from Institute for Digital Communications, The University of Edinburgh, Edinburgh, UK.