IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 24, 2018, accepted January 11, 2019, date of publication January 18, 2019, date of current version March 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2893624

Matrix Differential Decomposition-Based Anomaly Detection and Localization in NFV Networks

JING CHEN[1], MING CHEN[1], XIANGLIN WEI[2], AND BING CHEN[1]
[1] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
[2] Nanjing Telecommunication Technology Research Institute, Nanjing 210007, China

Corresponding author: Ming Chen (mingchen@nuaa.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61772271 and Grant 61379149.

ABSTRACT Network function virtualization (NFV) is a promising network paradigm that enables the design and implementation of novel network services with lower cost, increased agility, and faster time-to-value. However, network anomalies caused by software malfunction, hardware failure, mis-configuration, or cyber attacks can greatly degrade the performance of NFV networks. A few matrix decomposition-based methods have shown their effectiveness in finding the existence of network-wide anomalies. However, a little attention has been paid to multiple anomalies detection and anomaly devices localization. To bridge this gap, in this paper, we propose a matrix differential decomposition (MDD)-based anomaly detection and localization algorithm for NFV networks. First, an NFV network prototype is built to investigate the property of NFV networks, and the effectiveness of traditional anomaly detection methods is evaluated. Second, we detail the MDD-based Anomaly DEtection and Localization (MADEL) algorithm. Finally, a series of experiments are conducted on three different NFV networks to evaluate the performance of the proposed algorithm. Experimental results show that the MADEL algorithm could effectively detect and localize different types of network anomalies.

INDEX TERMS Network function virtualization, anomaly detection, localization, matrix differential decomposition.

I. INTRODUCTION

Through decoupling network functions from the physical devices on which they run, Network Function Virtualization (NFV) could greatly facilitate agile network design and deployment with low cost [1]–[4]. However, network anomalies, which are defined as the exceptional patterns in network traffic deviating from the normal profile of the network dynamics, facing the Internet also pose great challenges to NFV networks [5]–[9]. Furthermore, softwarization makes NFV networks vulnerable to various cyber attacks. In the event of an anomaly, a NFV network may have many adverse effects, such as the decline in quality of Service (QoS), the deterioration of user experience, and even communication disruption, resulting in huge economic losses and adverse social effects. Most of the network anomalies caused by software malfunction, hardware failure, mis-configuration, or cyber-attacks, are accompanied by unusual or significant growth of the network traffic, which leads to network congestion, increased Round Trip Time (RTT), and enlarged packet loss rate. Therefore, accurate detection and localization of network anomalies are critical for ensuring the stable and efficient operation of NFV networks.

Network anomaly detection (NAD) has been a long-lasting challenge due to many reasons. First, the intrinsic dynamic nature of network traffic makes it hard to distinguish between abnormal and normal behaviors [10], [11]. Second, observing or measuring the features, such as traffic matrix, for NAD accurately is a difficult and resource-intensive task. Third, the normal profile of the network is constantly changing. In recent years, different NAD schemes have been put forward for the Internet, such as statistical-based NAD, classification-based NAD, clustering and outlier-based NAD, soft computing-based NAD, and knowledge-based NAD etc. In contrast, in NFV networks, NAD is still in its infancy. Compared with single device or link monitoring-based anomaly detection methods, matrix or network area-based detection algorithms are more favorable to the maintainers of NFV networks since they can better capture the abnormal

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

behavior of the network as a whole. However, traditional matrix-based anomaly detection methods, such as Principal Components Analysis (PCA) [12]–[15] cannot be used to determine the number of anomalies and localize the anomaly devices although they are good at finding out single point anomalies.

To solve these problems, this paper presents a NAD method, which can simultaneously detecting and locating multiple anomalies in a NFV network based on the round-trip delay (RTT) matrix measured between multiple advantage points in the network. First of all, a NFV network prototype is established to present our data collecting process and the performance of existing PCA-based NAD algorithm is investigated based on collected data. Secondly, we introduce a matrix differential decomposition (MDD)-based Anomaly DEtection and Localization (MADEL) algorithm for NFV networks. Thirdly, a series of experiments are conducted on three different types of NFV networks to evaluate the performance of the proposed algorithm. Experimental results have shown that MADEL algorithm could effectively detect and localize different types of network anomalies.

Our contributions in this paper include: first, three different types of NFV prototype networks are constructed, and the low rank property of RTT matrix is validated; second, a matrix differential decomposition based anomaly detection and localization algorithm is put forward, which can find out the existence and locations of multiple network anomalies ; third, a series of experiments are conducted on three NFV prototypes with diverse parameter settings to evaluate the performance of the proposed MADEL algorithm, and experimental results validate its accuracy and efficiency.

The rest of the paper is organized as follows: Section II summarizes the relevant work; Section III presents the construction of our NFV network prototype, and investigates the performance of PCA-based NAD algorithm based on collected RTT matrix; Section IV details the design of MADEL algorithm; Section V introduces the experimental settings and results; Section VI gives a brief conclusion.

II. RELATED WORK

In recent years, many new and different trends in networks have been promoted, among which NFV has attracted significant attention from both industry and academia [16], [17]. However, NFV is still a developing technology, facing many challenges and issues in practical applications. Despite a large amount of researches on anomaly detection in traditional networks, anomaly detection and localization in NFV environment remain a challenge.

In traditional networks, there are many methods on anomaly detection. Barford *et al.* have proposed a performance anomaly detection and localization algorithm using active measurement, and sent probes to all links in the network in a certain period of time [12]. Wang *et al.* [18] have presented an improved classification detection algorithm based on the active measurement of network performance anomalies, which takes into account the number of the detected links, reducing the network link load and optimizing the process of link selection strategy to obtain the global optimal. For statistics-based anomaly detection, the most classic is to use PCA by analyzing traffic matrix. Lakhina *et al.* [19] first proposed a PCA-based anomaly detection algorithm, obtaining link flow data and separating the flow matrix space into normal and abnormal subspace, using Q statistics to calculate the threshold to detect network anomalies. In the field of network-wide anomaly detection, there are also many researches on machine learning-based detection algorithms. Ahmed *et al.* [20] have introduced two machine learning algorithms for network anomaly detection: One-Class Neighbor Machine (OCNM) and the recursive Kernel-based Online Anomaly Detection algorithms (KOAD). Liu *et al.* [21] have put forward a fuzzy c-means clustering algorithm, which uses fuzzy clustering to process data naturally, and has a better result in the network intrusion detection. Xie *et al.* [22], [23] have proposed the use of dynamically measured data to form a conventional tensor, which takes advantage of the user domain and the time domain factor matrix to form a three-dimensional tensor to achieve network measurement. However, the formation and integration of three-dimensional tensors has high algorithm complexity and is not suitable for network anomaly detection and localization with high real-time performance. The round-trip delay method proposed by Qian *et al.* [24] can only detect whether an anomaly exists in the network, but it cannot handle multiple anomalies situation.

Current anomaly localization methods can be roughly divided into two categories: localization methods based on end to end path probe and wavelet transform. A link weight iterative method is proposed by Barford *et al.* [12], and a node marking method is introduced by Xia *et al.* [25]. On the other hand, Barford *et al.* [26] have adopted discrete wavelet localization, and a diffusion wavelet-based localization is presented by Tian *et al.* [27]. Most of the localization methods based on end to end path probe are used to localize link anomalies, are effective for centralized network architecture, require high real-time performance of the network, have large detection frequency and large extra link load. However, the network traffic statistics method based on wavelet transform requires a large amount of traffic data set, which is not suitable for real-time anomalies detection.

In contrast, anomaly detection in NFV networks has attracted limited attention so far. Pavlidis *et al.* [28] have proposed a management-plane monitoring and anomaly detection services in a monitoring architecture. These services applied within customized analytics module learning mechanisms based on average entropy values from normal traces, alerting if entropy values deviate from established references threshold. Without the need of any metric thresholds, Kourtis *et al.* [29] has proposed an automatic identification of an anomaly in an NFV service, as a significant deviation from its normal operation. Blaise *et al.* [30] have introduced a VNF service chain anomalies detection method based on the Markov chain to identify the correctness of the

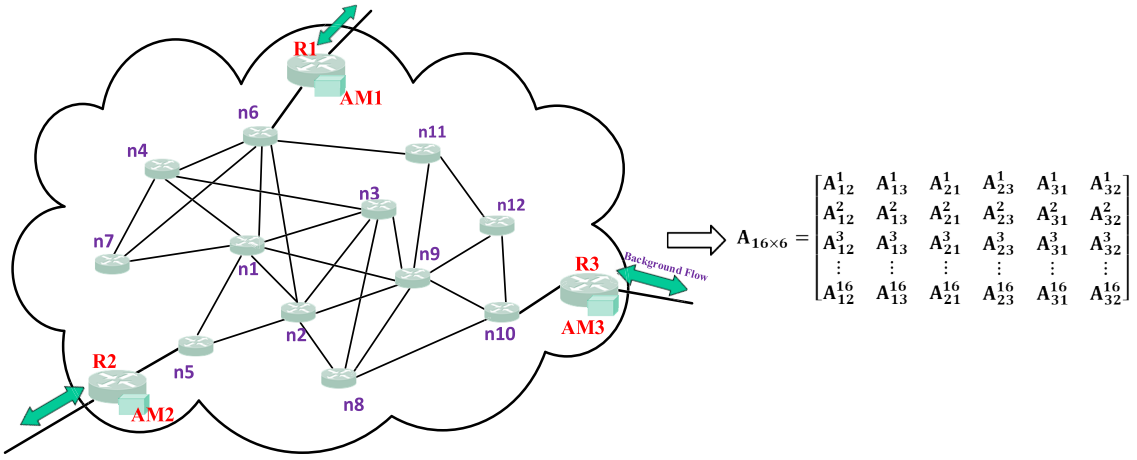


FIGURE 1. The topology of the constructed NFV network prototype (left), and the RTT matrix obtained at the end of the 16th timeslot (right).

service chaining request by observe whether exists abnormal behavior. Sampaio *et al.* [31] have presented reinforcement learning to promote resilience in SDN and NFV, whose policies for dealing with anomalies are defined based on rewards for each action.

Compared with the above research, this paper extends the network anomaly to include network performance anomalies based on RTT matrix in the NFV networks. Moreover, our algorithm can determine the number and location of anomalies and localize the devices that generate anomaly traffic.

III. PROTOTYPE CONSTRUCTION AND PROBLEM STATEMENT

A. PROTOTYPE CONSTRUCTION AND RTT MATRIX MEASUREMENT

To investigate the effectiveness of the PCA-based NAD algorithm and state the problem to be solved in this paper, we build a NFV network prototype based on Linux Container (LXC), and its topology is generated using BRITE [32] as shown in Fig.1. 15 routers following a power law distribution are included, in which three of them ($R1$, $R2$, and $R3$) are Autonomous System (AS) border routers and the other 12 nodes ($n1$-$n12$) are internal routers. The bandwidth of each link is set to be 10Mbps, and Open Shortest Path First (OSPF) is adopted as the routing algorithm. We inject a Poisson flow into this network every 5 seconds as the background flow, and each flow is composed of UDP packets with an intensity of 1 Mbps. To facilitate RTT measurement and collection, three RTT measurement virtual network function (VNF) programs $AM1$, $AM2$, and $AM3$ are deployed at $R1$, $R2$, and $R3$ respectively.

The operating period is divided into separated timeslots. In each timeslot, each VNF program (i.e. $AM1$, $AM2$, or $AM3$) measures the RTT values between itself and the other two programs deployed at other border routers. Generally speaking, for a NFV network with n border routers ports, the measured RTT values in each period are arranged into a $1\times\mathrm{P}$ vector, where $\mathrm{P} = n\times(n-1)$ is the number of RTT values between all border routers in each timeslot. All the RTT values obtained in T timeslots constitute a RTT matrix $\mathbf{A}_{\mathrm{T}\times\mathrm{P}}$. The i-th row represents all RTT values measured in the i-th timeslot, and the j-th column represents the RTT time series measured between the j-th border routers pairs. For the constructed prototype with 3 border routers, 6 RTT values can be obtained at each timeslot. For instance, a 16×6 RTT matrix $\mathbf{A}_{16\times6}$ can be obtained at the end of the 16th timeslot (as shown in the right hand of Fig. 1). In $\mathbf{A}_{16\times6}$, an element A_{ij}^{t} represents the RTT from the i-th border router to the j-th border router at the t-th timeslot.

B. RTT MATRIX ANALYSIS

RTT matrix describes the performance of the NFV network in both spatial and temporal dimensions, and the method of measuring the RTT value is easy to deploy. Moreover, the RTT matrix expresses the performance of the NFV network in a unified way, shielding the topology and technical complexity of the network. Furthermore, it has been validated that RTT matrix is sensitive to network performance changes and is suitable for reflecting the anomalies caused by network attacks or network congestion. PCA-based NAD algorithm is one of the typical methods that can detect network anomalies based on the RTT matrix [19].

The effectiveness of PCA-based NAD algorithm relies on the fact that the RTT matrix is a low rank matrix, which is caused by the similarity in the RTT matrix. To be specific, for a border router, the RTT data measured in adjacent timeslots has temporal dependence, which leads to the correlation or similarity between the rows in the RTT matrix. On the other hand, the measurement flows from border routers in the same timeslots may cross common links or devices in the NFV network, which leads to spatial correlation or similarity between columns in the RTT matrix. If there are a large

number of related vectors in the RTT matrix, it will be a low rank matrix [33].

Singular value decomposition (SVD) method can be adopted to verify whether a RTT matrix has the low rank property. First, $\mathbf{A}_{\mathrm{T}\times\mathrm{P}}$ is decomposed as: $\mathbf{A}=\mathbf{U}\sum\mathbf{V}^{\mathbf{T}}$. Here, $\mathbf{U}$ is a T×T unitary matrix, $\sum$ is a semi-definite T×P diagonal matrix, and the i-th diagonal element λ_i, $1 \leq i \leq T$ is the singular value of matrix $\mathbf{A}$. $\mathbf{V}$ is a P×P unitary matrix, $\mathbf{V}^{\mathbf{T}}$ is an order conjugate transposed matrix of $\mathbf{V}$. Let the rank of $\mathbf{A}$ be $rank(\mathbf{A})$. It can be known that $rank\ (\mathbf{A}) = r \ll min\{N, T\}$. If the sum of the variance contributions of the first k $(k < r)$ singular values is approximately equivalent to the sum of the variance contribution rates of all singular values, we say matrix $\mathbf{A}$ is a low rank matrix.

To validate the low rank property of the RTT matrix, we conduct experiments on the constructed NFV network prototype through injecting two anomalous flows targeted at $R3$, whose flow rate is about 50Mbps and is much higher than background flows, into $R1$. To be specific, from the first timeslot to the 30th timeslot, no anomalous flow is injected; from the 31th timeslot to the 80th timeslot, the first anomalous flow is continuously injected; from the 61th timeslot to the 80th timeslot, the second anomalous flow is continuously injected. Based on these settings, three datasets are collected. Dataset1 is the RTT matrix collected between the first timeslot and the 20th timeslot, i.e. no anomalous flows exist in Dataset1. Dataset2 is the RTT matrix collected between the 30th and the 50th timeslot, i.e. one anomaly is in Dataset2. Dataset3 is the RTT matrix collected between the 60th and the 80th timeslots, and it contains two anomalies.

Then, we calculate the ratio of the cumulative variance contribution rate of the first k singular values for the three Datasets, which is defined as: $P\,(k) = \frac{\sum_{i=1}^{k}\lambda_i^2}{\sum_{i=1}^{r}\lambda_i^2}, 0 \leq rank\ (A) = r \leq P$. The results are shown in Fig. 2. We can see that the cumulative variance contribution rate of first three singular values in this three datasets all exceed the 95% threshold, which means its first $k = 3$ singular values can capture most characteristics of the network. In other words, the RTT matrix has a low rank characteristic regardless of whether there is an anomaly in the matrix.

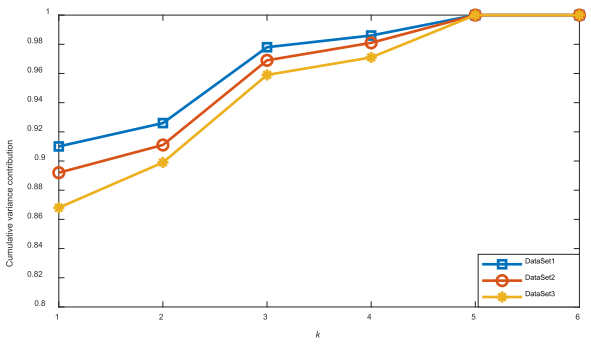


FIGURE 2. The cumulative variance contribution rate of the first k singular values for three collected datasets.

C. PCA-BASED NAD ALGORITHM AND PROBLEM STATEMENT

To derive the principal component of $\mathbf{A}$, we first calculate the covariance matrix $\mathbf{C}=\mathbf{A}^{\mathbf{T}}\mathbf{A}$ and the eigenvalues as well as eigenvectors of $\mathbf{C}$. Let the i-th eigenvalue be λ_i, and its corresponding eigenvector be v_i, then $A^T A v_i = \lambda_i v_i$, $i = 1, \cdots, p$. Afterwards, the principal component selection process is to sequentially select the corresponding orthogonal feature vectors according to the descending order of the feature values, i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Then, we can calculate the cumulative variance contribution rate of the first k principal component of $\mathbf{A}$. The cumulative contribution rate of the first r principal elements is denoted by $\alpha_r = \frac{\sum_{i=1}^{r}\lambda_i}{\sum_{i=1}^{p}\lambda_i}, 0 \leq r \leq p$. If the cumulative variance contribution rate of the first K principal elements exceeds a threshold c_0, which is usually set to be $c_0 = 85\%$, we can say that the first K principal components could express the most significant characteristics of the matrix. Finally, $\mathbf{A}$ will be projected to the selected K principal components to obtain mutually orthogonal vectors $U_i = \frac{Av_i}{\sqrt{\lambda_i}}, i = 1, \cdots, K$, and a new matrix $\mathbf{Av}$ could be derived. We know that U_i is the weighted sum of all column vectors of the RTT matrix $\mathbf{A}$ with the weight v_i, i.e. the i-th principal component of the RTT matrix. Using the idea of the subspace, we can divide the new space into two parts: one is normal subspace (denoted as S), i.e., the set of the first r 'normal' principal components; the other is 'anomalous' subspace, i.e., the set of the remaining K-r abnormal principal component (denoted as $\tilde{M}$). Then, we project the newly acquired $\mathbf{Av}$ to these two subspaces. Let the projection value of a RTT row vector x in the normal and abnormal subspaces be $\hat{x}$ and $\tilde{x}$ respectively, we have $x = \hat{x} + \tilde{x}$. Arranging the set $\{v_i\}_i^r$ according to S into a matrix $\mathbf{P}_{p\times r}$ in order, then $\hat{x} = PP^T x$ and $\tilde{x} = (I - PP^T)x$.

The anomalies are judged based on a metric named Square Prediction Error (SPE), which is defined as: $\mathrm{SPE} \equiv ||\tilde{x}||^2 = ||(I - PP^T)x||^2$ [24]. If $SPE \geq \delta_\alpha^2$, an anomaly is raised, and δ_α^2 is a SPE threshold with confidence $1 - \alpha$.

PCA-based NAD algorithm can be used to detect network anomalies in the three datasets collected in Section III-B. Fig.3 shows the detection results. In Fig. 3(a), (b), and (c), an anomaly is declared if the SPE statistic value exceeds the 99% confidence (the red dashed line in Fig. 3). It can be seen that, PCA-based NAD algorithm finds that there are anomalies in Dataset2 and Dataset3, and no anomaly in Dataset1. However, this method can neither determine the number of anomalies nor localize the anomalies in Dataset2 and Dataset 3.

IV. MATRIX DIFFERENTIAL DECOMPOSITION-BASED ANOMALY DETECTION AND LOCALIZATION

This section firstly analyzes the problem. Then, we detail the matrix differential decomposition-based anomaly detection and localization algorithm.

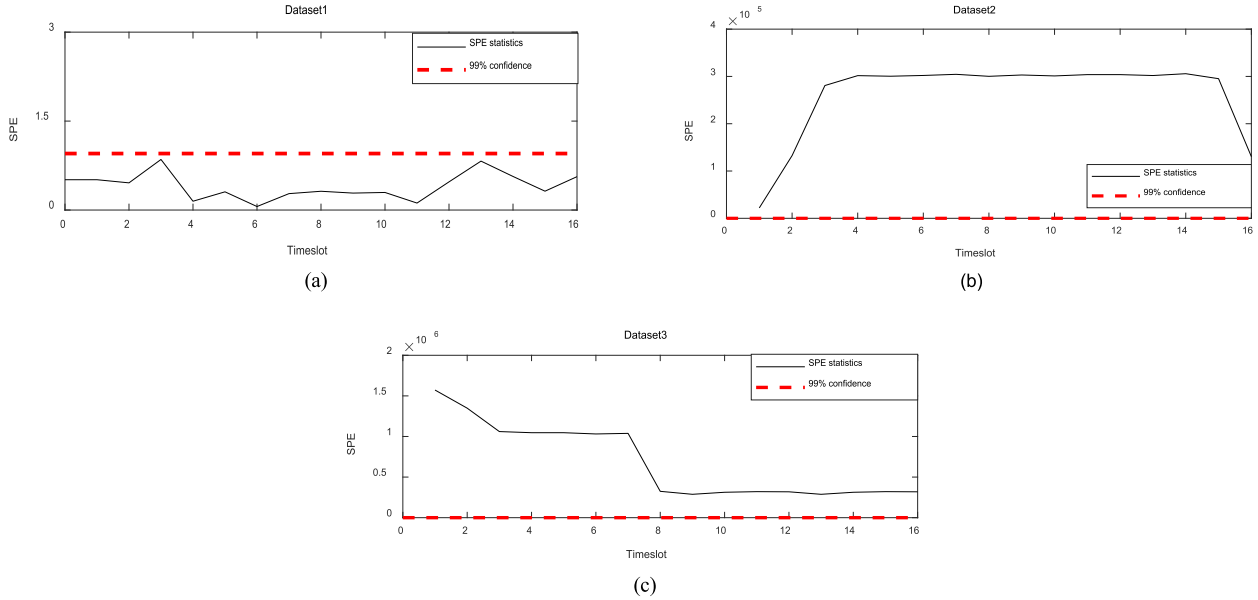


FIGURE 3. Detection results of PCA-based NAD algorithm on three datasets. (a) Detection results on Dataset 1. (b) Detection results on Dataset 2. (c) Detection results on Dataset 3.

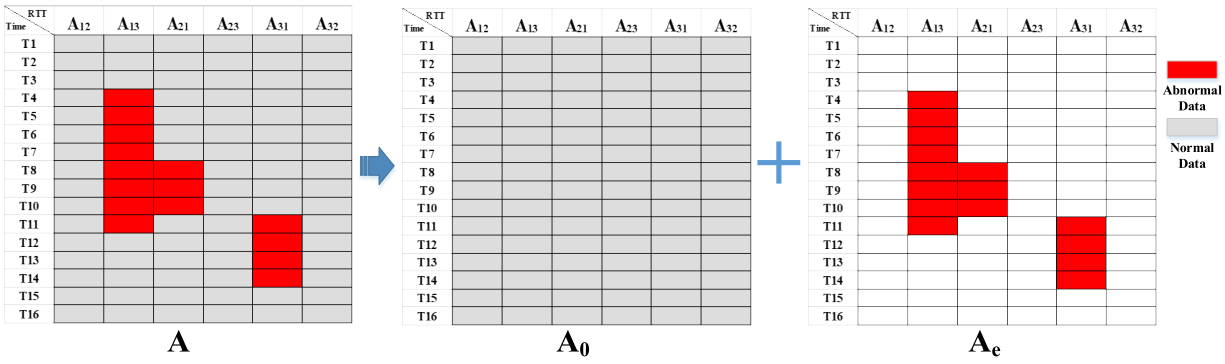


FIGURE 4. An example of the RTT matrix decomposition.

A. PROBLEM ANALYSIS

Assume that anomalies could occur at one or more locations in the network, and cause fluctuations in the values of the corresponding values in the RTT matrix. Moreover, it is assumed that abnormal values caused by anomalies account for a small proportion of matrix $\mathbf{A}$. To detect anomalies, we can decompose matrix $\mathbf{A}$ into two matrices, i.e. a reference matrix $\mathbf{A_0}$ and a differential matrix $\mathbf{A}_e$. The RTT values in normal condition are anticipated to be included in $\mathbf{A_0}$, and the anomaly values in $\mathbf{A}_e$, which is expected to be a sparse matrix since abnormal values only account for a small proportion. Take the RTT matrix $\mathbf{A}$ obtained in the NFV network prototype constructed in Section III as an example, Fig. 4 shows the decomposition process of matrix $\mathbf{A}$, in which red rectangles represent the values impacted by anomalies.

This decomposition problem could be formulated as:

$$\begin{aligned} &\min_{A_0,A_e} \quad \|(A - A_e) - A_0\|_F \\ &s.t.\ rank\ (A_0) \leq k, \quad \|A_e\| \leq E \end{aligned} \tag{1}$$

This is a constrained optimization problem, which aims at minimize the difference between $\mathbf{A} - \mathbf{A}_e$ and $\mathbf{A}_0$ on the condition that $\mathbf{A}_0$ is a low-rank matrix, and there are at most E non-zero values in matrix $\mathbf{A}_e$. The Frobenius norm of the matrix ($\| \|_F^2$) is used to minimize the squared error constraint, it can reduce the error caused by the L2 normal form ($\| \|_2$) when the number of anomalies is small and thus to improve the accuracy of anomaly detection. Then, we can detect and localize the anomalies based on matrix $\mathbf{A}_e$.

B. DIFFERENTIAL DECOMPOSITION ALGORITHM

Based on the above analysis, we put forward the matrix differential decomposition (MDD)-based Anomaly DEtection and Localization (MADEL) algorithm that contains four steps, as shown in Algorithm 1:

(1) Differential processing. As shown in lines 1-5 in Algorithm 1, a differential matrix $\mathbf{A}_e = \mathbf{A} - \mathbf{A}_0$ is obtained, where $\mathbf{A}$ and $\mathbf{A}_0$ are the RTT matrix contains anomalous data and reference RTT matrix whose data collected in a normal

Algorithm 1 MADEL Algorithm

Input:

measured RTT matrix: $\mathbf{A}$; reference RTT matrix: $\mathbf{A_0}$; number of border routers: N; abnormal frequency threshold: f_0

Output:

Detected Anomalies Pairs Set: ***APS***, Localized Anomalies Set: ***LAS***

1: $\mathbf{A_e} = \mathbf{A} - \mathbf{A_0}$ //Differential processing
2: ***for*** each $A_{e_{ij}}$ in $\mathbf{A_e}$ //Anomalies Detection Phase
3: ***if*** $(|A_{e_{ij}}| > c_0)$ ***then*** $A_{e_{ij}} = |A_{e_{ij}}|$, add j to β
4: ***else*** $A_{e_{ij}} = -1$
5: ***end***
6: $\widetilde{A_e} = A_e(\alpha, \beta) \in R^{T \times d}$ //Abnormal data extraction
7: ***for*** each β_i in β // Determination the number of anomalies
8: ***do*** $f_{(P_{k1},P_{k2})} = f_{(P_{k2},P_{k1})} = f_{(P_{k1},P_{k2})}+1$, where P_{k1} and P_{k2} are the source and the destination of the β_i-th column of matrix $\widetilde{A_e}$, add (P_{k1}, P_{k2}) to P
9: ***if*** $(f_{(P_{k1},P_{k2})} > f_0)$ ***then*** add (P_{k1}, P_{k2}) to ***APS***
10: ***end***
11: $|$***APS***$|$ is the number of anomalies
12: ***for*** each (P_{k1}, P_{k2}) in ***APS*** // Anomalies localization Phase
13: *IPS* $\leftarrow$ Overlap IPs on the traceroute paths from P_{k1} to P_{k2} and from P_{k2} to P_{k1}
14: *count_times* $\leftarrow$ Sub-paths contains the source and destination whose response times remarkably exceeds the average one-hop delay in *IPS*
15: Sort *count_times* and localize the first sub-path (x_1, x_2), add (x_1, x_2) into ***LAS***
16: ***end***
17: ***LAS*** is the localization results of anomalies
18: return ***APS*** and ***LAS***

network environment respectively. For each element $A_{e_{ij}}$ in $\mathbf{A}_e$, if $|A_{e_{ij}}| < e_0$ (e_0 is a small constant), we assign a negative value to it, i.e. $A_{e_{ij}} = -1$; otherwise, $A_{e_{ij}} = |A_{e_{ij}}|$.

(2) Abnormal data extraction. From line 3 to line 6, according to the correlation between RTT matrix rows and the properties of the non-negative matrix and block matrix [34], we can know that an anomaly will affect a certain range of device pairs. So, we can find the non-negative anomaly sub-matrix from the differential sub-matrix $\widetilde{A_e} = A_e(\alpha, \beta) \in R^{T \times d}$, $\alpha_i = \{1, \cdots, T\}$, $\beta_i \subseteq \{1, \cdots, n^2\}$, $d \leq n^2$, where α is all the row label set of $\mathbf{A}_e$, and β is the column label set selected by $\mathbf{A}_e$. The selection of β satisfies that at least one element in each column is non-negative, and β reveals correlation information for all anomalous device pairs.

(3) Determine the number of anomalies. From line 7 to line 10, each element β_i, $0 \leq i \leq |\beta|$, in β contains an anomalous border routers pair information, and the corresponding device pair is added to the column label set P. The frequency of each pair (P_{k1}, P_{k2}), $0 \leq k \leq |P|$ in the set P is counted, and $(P_{k1}, P_{k2}) = (P_{k2}, P_{k1})$. If a pair's frequency exceeds a predefined threshold f_0, it is added to the Anomalies Pairs Set (*APS*). Line 11 shows that the size of the *APS* is the number of detected anomalies, and the pairs in *APS* are the potential anomalous border routers and will be called suspicious border routers pairs.

(4) Anomaly localization. As shown in lines 12-18, each suspicious border routers pair in the *APS* represents a path with an anomaly. For each suspicious path, traceroute is adopted to get the path information by probing back and forth, and all the IP addresses and hop counts on the two paths for each pair could be recorded. For instance, for a pair of routers (P_{k1}, P_{k2}), we traceroute the paths both from P_{k1} to P_{k2} and from P_{k2} to P_{k1} respectively. We record each overlapped IP address (device) of these two paths in a set IPS. Furthermore, we also record all the devices (by addresses) whose response times remarkably exceeds the average one-hop delay in the IPS, count and sort the occurrence times of each sub-path accordingly (denoted as *count_times*), where sub-path is the source and destination of the anomalous hop. Then the sub-path with the highest *count_times* will be treated as the anomaly localization results. i.e. the anomaly devices.

The time complexity of the MADEL algorithm is decided by two factors, i.e. MDD decomposition and APS set calculation based on frequency threshold. The complexity of the RTT non-negative anomaly sub-matrix $\widetilde{A_e}$ decomposition is $O(tn)$, where t is the number of measurement timeslots and n is the number of border routers. The time complexity of the label set based on the frequency threshold is $O(n)$. Therefore, the time complexity of the MADEL algorithm is $O(tn)$. In contrast, the complexity of the traditional PCA-based NAD algorithm is $O(tn^2)$ [35], and the tensor-based three-dimensional data anomaly detection and decomposition method's time complexity is approximated as $O(kn^3)$ (k is a parameter) [23]. Moreover, the MADEL algorithm only uses two-dimensional data to reflect the spatial-temporal characteristic of the RTT values, which greatly reduces the

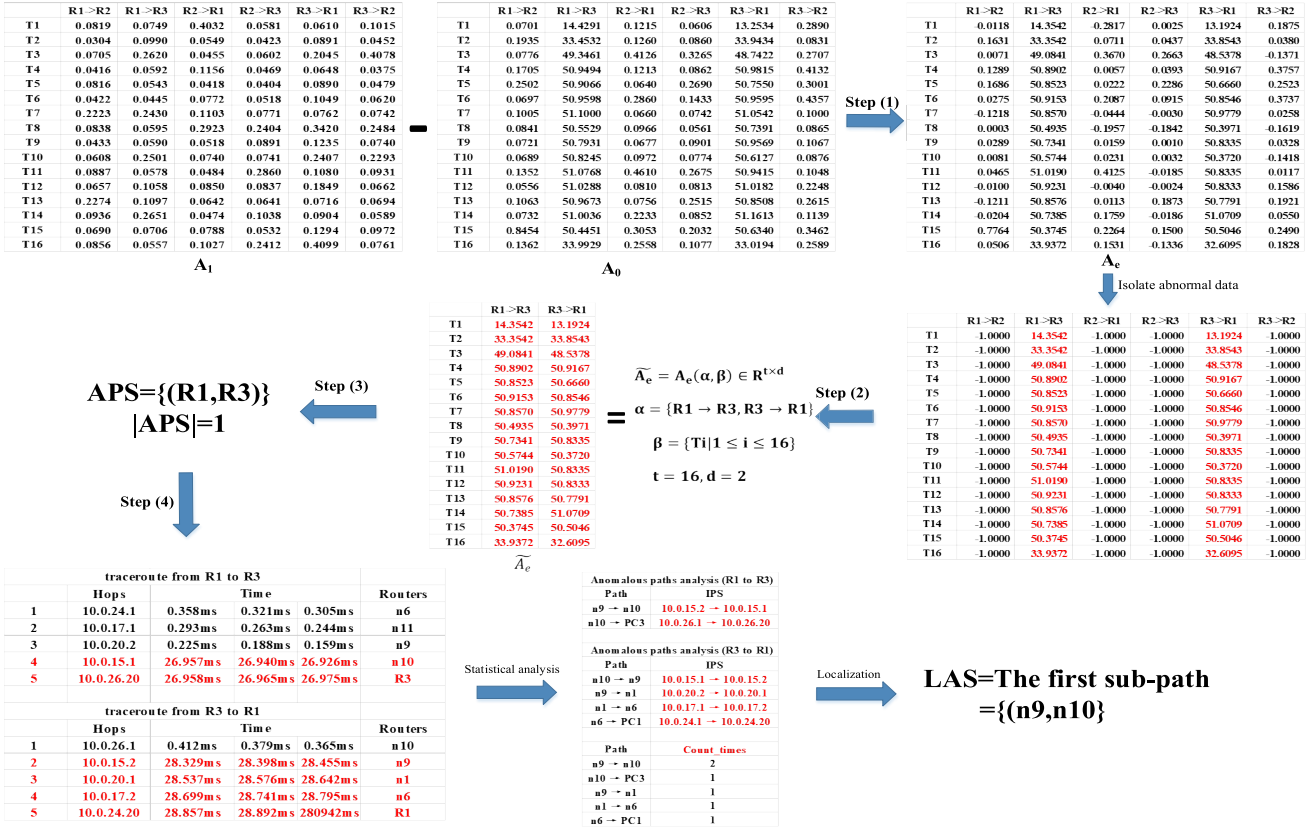


FIGURE 5. An example of MADEL algorithm's working flow.

space complexity. Furthermore, PCA-based NAD algorithm can only detect the existence of anomalies but cannot localize to anomaly devices.

C. AN EXAMPLE

Here, the NFV network prototype is adopted to show the workflow of MADEL algorithm. The length of each timeslot is set to be 8 seconds. From the beginning of the 20*th* timeslot, we inject abnormal flows into the internal router $n9$ for one timeslot. After 16 timeslots, a 16×6 RTT matrix $\mathbf{A}$ is obtained as shown in Fig. 5. In the step (2) shown in the right middle of Fig. 5, the abnormal data is isolated by the differential matrix. Thus, $APS = \{(R1, R3)\}$, and there exist $|APS| = 1$ anomaly in this network, and the anomalous border routers are $R1$ and $R3$ respectively. The traceroute results of $R1$ to$R3$ and $R3$ to $R1$ are shown in the tables in Fig. 5. Combining with network topology and statistical analysis information, the IP address of anomalous devices can be localized, i.e. $LAS = \{(n9, n10)\}$ in Fig. 5. This is comply with the parameter setting. Therefore, MADEL algorithm can accurately detect the anomaly in the experimental environment and provide the localization information of the anomaly.

V. EXPERIMENTS AND RESULTS ANALYSIS

In order to evaluate the performance of MADEL algorithm, a series of experiments are conducted. This section firstly introduces the parameter settings, and then analyzes the experimental results.

A. PERFORMANCE METRICS

To evaluate the performance of the anomaly detection and localization method, the following metrics are adopted:

True Positive Rate (*TPR*): the proportion of anomalies that are correctly diagnosed.

True Negative Rate (*TNR*): the proportion of non-anomalies that are correctly diagnosed.

False Negative Rate (*FNR*): the proportion of anomalies that are not identified.

False Positive Rate (*FPR*): the proportion of non-anomalies that are wrongly identified as anomalies.

Correct Localization Rate (*CLR*): the proportion of network devices with anomalies that are correctly localized.

Let the anomalies injected into the network be IAS (Injected Anomalies Set) and the number of injected anomalies is |IAS|, the non-anomalies routers set is recorded as NAS (Non-Anomalies Set) and |NAS| is the number of non-anomalies. Assume the detected anomalous border routers pairs information by the MADEL algorithm is stored in APS, the localized anomalies set is LAS, and the number of detected anomalies is |APS|. According to the above

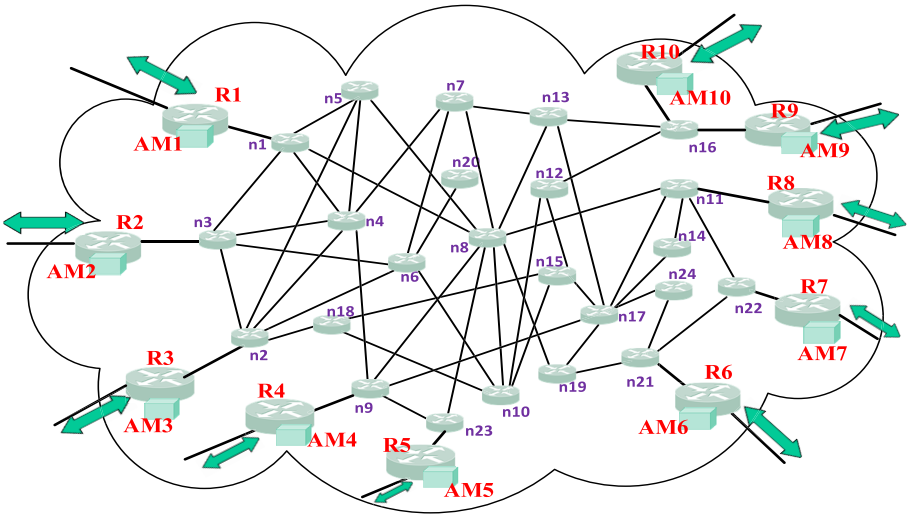


FIGURE 6. The topology of the NFV network with increased size.

definitions, we know that:

$$TPR = \frac{|APS \cap IAS|}{|IAS|}, \quad FNR = \frac{|IAS - APS \cap IAS|}{|IAS|}$$
$$FPR = \frac{|APS \cap NAS|}{|NAS|}, \quad TNR = \frac{|NAS - APS \cap NAS|}{|NAS|}$$
$$CLR = \frac{|IAS \cap LAS|}{|IAS|} \tag{2}$$

B. RESULTS AND ANALYSIS

The example in the above section has shown the effectiveness of the MADEL algorithm on the constructed prototype. Here, we first increase the size of the NFV network to investigate the performance of the MADEL algorithm in large-scale networks. Then, a real-world topology is adopted to further show its accuracy.

The topology of the network with increased size is shown in Fig. 6. This NFV network contains 34 routers following a power law distribution, in which ten of them ($R1$-$R10$) are AS border routers, and the other 24 nodes ($n1$-$n24$) are internal routers. Four different types of anomaly injection methods, i.e. S1-S4, are adopted here. In scenario S1, two independent anomalies are injected into two separate internal routers (e.g. $n3$ and $n14$ in Fig. 6); in scenario S2, two anomalies are injected into two internal routers which are close to each other (e.g. the two ports adjacent to $n17$ and $n24$ in Fig. 6); in scenario S3, two anomalies are injected into two different ports of the same router (e.g. the two different ports of $n6$ in Fig. 6); in scenario S4, three anomalies are injected into $n9$, $n13$, and $n17$, respectively.

MADEL algorithm is adopted to detect and localize the anomalies in these four scenarios. We average all results over 200 experiments. Higher *TPR* (*TNR, CLR*) and smaller *FPR* (*FNR*) mean better detection performance. The detection and localization results are shown in TABLE 1.

TABLE 1. Detection and localization results of the MADEL algorithm in four different scenarios.

Scenario	TPR	TNR	FNR	FPR	CLR
S1	99%	99.03%	1%	0.97%	99.91%
S2	97%	98.33%	3%	1.67%	95.92%
S3	96.75%	97.88%	3.25%	2.12%	95.64%
S4	96.16%	97.61%	3.84%	2.39%	93.31%

The results in TABLE 1 show that the detection and localization accuracy of MADEL algorithm is robust to the injection models. Moreover, if the number of anomalies is small and the correlation between abnormal positions is low, MADEL algorithm performs much better.

In order to further evaluate MADEL algorithm in other anomalous scenarios, we first define a metric named anomalies ratio as: $\varepsilon = \frac{anomalies\_num}{n}$, where *anomalies_num* represents the number of anomalies and n is the total number of AS border routers in the NFV network. Here, we assume that $anomalies\_num \leq n$ to avoid the situation that anomalous traffic dominates the network. The performance of MADEL algorithm on the network shown in the Fig. 6 is illustrated in Fig. 7. With the increase of anomaly ratio, the *TPR*, *TNR*, and *CLR* of the MADEL algorithm decrease from 1 to 0.82, 0.86, and 0.82 respectively. This is because multiple anomalies cause overlapping of anomalous paths, which reduces the accuracy of detection and localization. However, MADEL algorithm's *TPR*, *TNR*, and *CLR* are still larger than 0.8, indicating that MADEL algorithm has a good effect on multiple anomaly detection and localization.

We also implement another NFV network based on the topology of the Corporation for Education Network Initiatives in California (CENIC) network [36], whose backbone is

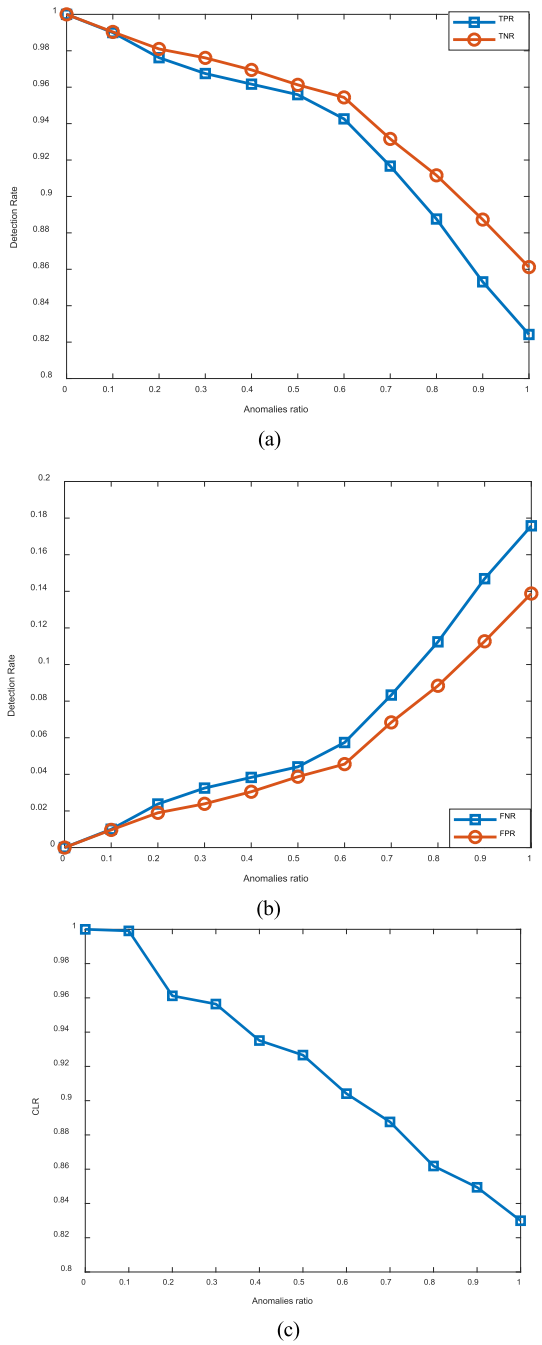


FIGURE 7. The performance of the MADEL algorithm with different anomalies ratio. (a) The *TPR* and *TNR* of MADEL algorithm. (b) The *FNR* and *FPR* of MADEL algorithm. (c) The *CLR* of MADEL algorithm.

composed of a sparse mesh routers connected by high-speed links. With different anomaly injection methods, we have found that the average values of the MADEL algorithm's *TPR*, *TNR* and *CLR* are always no less than 80%. This means that MADEL algorithm can adapt to different NFV networks.

C. PARAMETER INFLUENCES

1) THE IMPACT OF THE DEGREE OF NETWORK ANOMALIES

We investigate the performance of the MADEL algorithm with different sizes of the injected anomalies, we build three types of anomalous flows. In the NFV network shown in Fig. 6, at the 10^{th} timeslot, we inject one anomalous Poisson flow with the average rate of 5 Mbps into the network, and it lasts for 10 timeslots; at the 30^{th} timeslot, another anomalous Poisson flow lasting for 10 timeslots with the average rate of 20 Mbps is injected into the same internal routers; at the 50^{th} timeslot, at the same locations, we inject the last anomalous Poisson flow with the average rate of 50 Mbps, which also lasts for 10 timeslots. We carry out the anomaly injection algorithms in these three anomalous Poisson flows scenarios over 200 experiments. Experimental results are shown in Fig. 8. The results in Fig. 8(a) show that with the increase of the intensity of anomalous flows, the larger anomalous data values in the differential matrix because of the sudden increase of the anomalous flow leads to an increase in network congestion. Fig. 8(b)-(d) also shows that MADEL algorithm's *TPR*, *TNR* and *CLR* values increase with the increase of the injection rate of the anomalous flows. In other words, MADEL algorithm performs better when the volume of the injected anomaly traffic is larger, and the performance of MADEL algorithm will be roughly the same when the rate of injected flows exceed 20 Mbps but not dominate the traffic pattern in the network as assumed in Section IV.

2) THE IMPACT OF NETWORK STATE CONVERGENCE TIME

To investigate the impact of network state convergence time on the performance of MADEL algorithm, in the extended NFV network introduced in in Fig. 6, Anomalies Ratio is set to be 0.2, i.e. two anomalous Poisson flows with the average rate of 50 Mbps are injected into internal routers $n3$ and $n14$ respectively at the beginning, and both of them last for 20 timeslots. In this scenario, we collect data at each timeslot and use MADEL algorithm to detect and localize anomalies, and we average all results over 200 experiments. During the anomalies injection process, we need to avoid serious anomalous paths overlap with existing anomalies.

Fig. 9 shows MADEL algorithm's *TPR*, *TNR* and *CLR* values when different amount of data are collected for detection. To be specific, each Timeslot's corresponding *TPR*, *TNR* and *CLR* values are derived based on the condition that we use the RTT matrix collected at the first T timeslots for anomaly detection. We can observe that the *TPR*, *TNR* and *CLR* values are approximately 0.99 after the 7^{th} timeslot. This is due to the fact that after the 7^{th} timeslot, the state of the network tends to be more stable than the burst arrival period. Therefore, we should collect RTT data no less than timeslot=7 in each period.

3) THE IMPACT OF NETWORK SIZE

In order to investigate the impact of network size and structural complexity on the performance of MADEL algorithm, we generate a larger scale network topology using BRITE with a power law distribution [32], and let other parameter settings intact. In specific, there are 30 border routers and 100 internal routers (denoted as Prototype2),

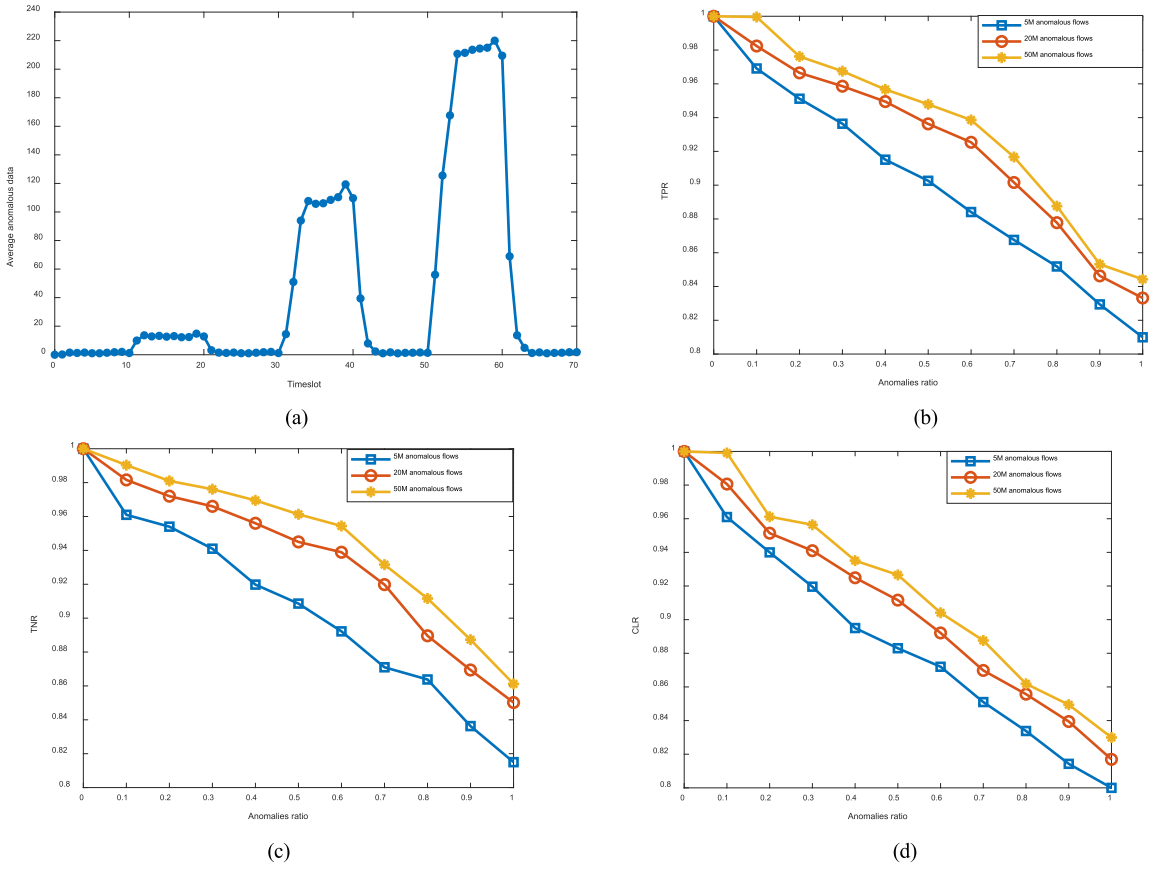


FIGURE 8. The influence of the degree of anomalous flows on the performance of MADEL algorithm. (a) Average anomalous data. (b) The *TPR* of MADEL algorithm. (c) The *TNR* of MADEL algorithm. (d) The *CLR* of MADEL algorithm.

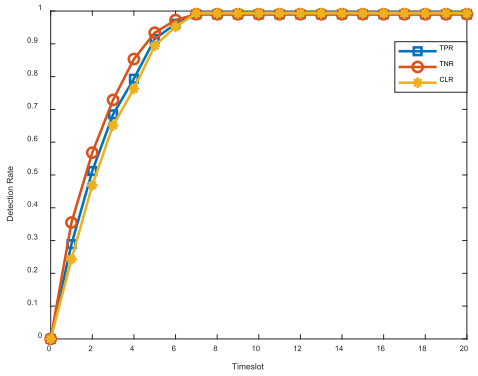


FIGURE 9. The influence of the convergence time on MADEL algorithm.

and four anomaly injection scenarios are adopted as introduced in Section V-B. The results are shown in TABLE 2. Moreover, compared Prototype2 with the network introduced in Section V-B (denoted as Prototype1), Fig. 10 shows the experimental results with different number of anomalies and various network sizes.

As can be seen from Table 2 and Fig. 10, MADEL algorithm's *TPR*, *TNR*, and *CLR* values on Prototype1 are higher than those values on Prototype2 when the anomalies ratio is less than 80%; when the anomalies ratio is larger than 0.8,

TABLE 2. Results of MADEL algorithm on Prototype 2.

Scenario	TPR	TNR	FNR	FPR	CLR
S1	97.24%	97.68%	2.76%	2.32%	99.91%
S2	96.65%	97.27%	3.35%	2.73%	95.48%
S3	95.26%	96.38%	4.74%	3.62%	94.39%
S4	94.95%	95.38%	5.05%	4.62%	91.96%

the detection and localization accuracy of MADEL algorithm on Prototype1 significantly decreases. This is because that Prototype1 has a smaller network size, and when the number of anomalies is too large, the range of anomalies influences will overlap, which will affect the accuracy of MADEL algorithm. Therefore, we can know that MADEL algorithm performs much better when the number of anomalies is less than the number of AS border routers and the number of anomalies only accounts for a small portion of the number of internal routers. The accuracy of detection and localization slightly decreases when the number of anomalies increases or when the correlation between multiple anomalies is strong. There is no significant difference between the *TPR*, *TNR* and *CLR* for different network scales and structural complexity.

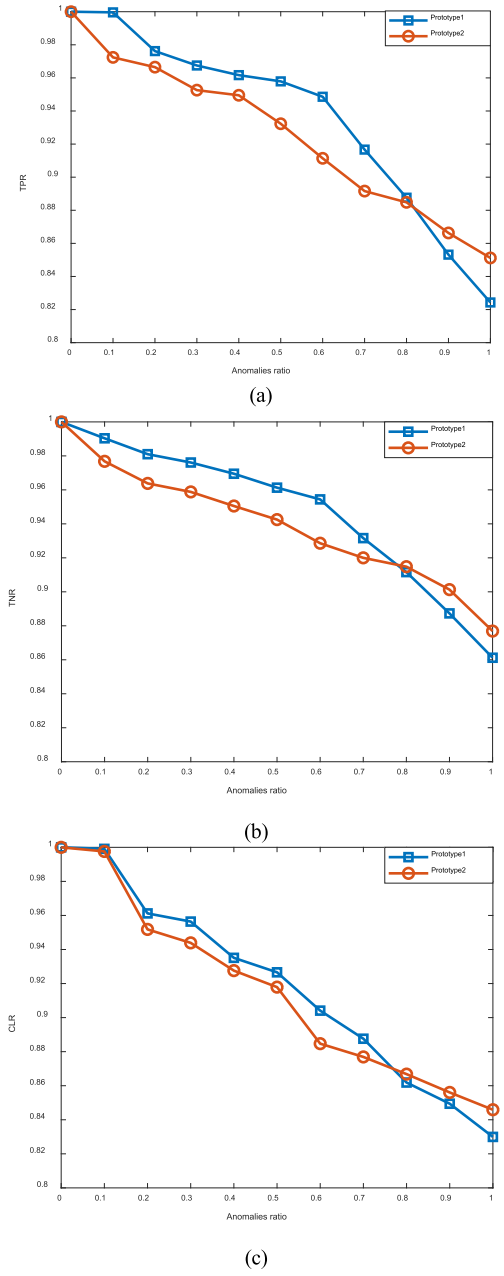


FIGURE 10. The performance of the MADEL algorithm with different number of anomalies and network sizes. (a) The *TPR* of two prototypes. (b) The *TNR* of two prototypes. (c) The *CLR* of two prototypes.

In summary, in different NFV network settings, MADEL algorithm can detect and localize multiple anomalies quickly, efficiently, and accurately.

VI. CONCLUSION

In order to overcome the shortcomings of current matrix-based network performance anomaly detection and localization methods, such as the inability to determine the number and location of multiple anomalies, this paper first constructs a Network Function Virtualization (NFV) network prototype, and evaluates the performance of Principal Components Analysis (PCA)-based Network anomaly detection (NAD) algorithm on the measured Round Trip Time (RTT) matrix, which is validated to be low rank. Secondly, a matrix differential decomposition (MDD)-based Anomaly DEtection and Localization (MADEL) algorithm is put forward, which can not only determine the number of multiple anomalies, but also localize the network devices that lead to the anomalies. Finally, a series of experiments are conducted on three typical NFV networks with diverse parameter settings. Experimental results show that the MADEL algorithm can accurately and effectively detect and localize multiple anomalies in the network, and has a strong adaptability. The work of this paper greatly reduces the computational overhead of the existing anomaly detection method, reduces the difficulty of deployment. In future, we plan to develop an online detection and localization algorithm, and extend it to the actual management application of the Internet.

REFERENCES

[1] J. de Jesus Gil Herrera and J. F. B. Vega, "Network functions virtualization: A survey," *IEEE Latin America Trans.*, vol. 14, no. 2, pp. 983–997, Feb. 2016.

[2] O. Krasko, H. Al-Zayadi, V. Pashkevych, H. Kopets, and B. Humeniuk, "Network functions virtualization for flexible deployment of converged optical-wireless access infrastructure," in *Proc. 14th Int. Conf. Adv. Trends Radioelectron., Telecommun. Comput. Eng. (TCSET)*, Feb. 2018, pp. 1135–1138.

[3] M. D. Ananth and R. Sharma, "Cost and performance analysis of network function virtualization based cloud systems," in *Proc. IEEE 7th Int. Adv. Comput. Conf. (IACC)*, Jan. 2017, pp. 70–74.

[4] T. Lin and Z. Zhou, "Robust virtual network function provisioning under random failures on network function enabled nodes," in *Proc. 10th Int. Workshop Resilient Netw. Design Modeling (RNDM)*, Aug. 2018, pp. 1–7.

[5] A. J. Gonzalez, G. Nencioni, A. Kamisiński, B. E. Helvik, and P. E. Heegaard, "Dependability of the NFV orchestrator: State of the art and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3307–3329, 4th Quart., 2018.

[6] K. K. Ramakrishnan, "Software-based networks: Leveraging high-performance NFV platforms to meet future communication challenges," in *Proc. IEEE 36th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2016, p. 24.

[7] A. Aljuhani and T. Alharbi, "Virtualized Network Functions security attacks and vulnerabilities," in *Proc. IEEE 7th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2017, pp. 1–4.

[8] N. Omnes, M. Bouillon, G. Fromentoux, and O. Le Grand, "A programmable and virtualized network & IT infrastructure for the Internet of Things: How can NFV & SDN help for facing the upcoming challenges," in *Proc. 18th Int. Conf. Intell. Next Gener. Netw.*, Feb. 2015, pp. 64–69.

[9] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 236–262, 1st Quart., 2016.

[10] D. Jiang, Z. Xu, P. Zhang, and T. Zhu, "A transform domain-based anomaly detection approach to network-wide traffic," *J. Netw. Comput. Appl.*, vol. 40, pp. 292–306, Apr. 2014.

[11] J. Liu and H. Tian, "Study on network anomaly localization techniques," in *Proc. 17th Int. Conf. Parallel Distrib. Comput., Appl. Technol. (PDCAT)*, Guangzhou, China, Dec. 2016, pp. 395–398.

[12] P. Barford, N. Duffield, A. Ron, and J. Sommers, "Network performance anomaly detection and localization," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 1377–1385.

[13] D. H. Hoang and H. D. Nguyen, "A PCA-based method for IoT network traffic anomaly detection," in *Proc. 20th Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2018, pp. 381–386.

[14] Y. Jin, C. Qiu, L. Sun, X. Peng, and J. Zhou, "Anomaly detection in time series via robust PCA," in *Proc. 2nd IEEE Int. Conf. Intell. Transp. Eng. (ICITE)*, Sep. 2017, pp. 352–355.

[15] E. Zyad, C. Khalid, and B. Mohammed, "Combination of R1-PCA and median LDA for anomaly network detection," in *Proc. Intell. Syst. Comput. Vis. (ISCV)*, Apr. 2017, pp. 1–5.

[16] M. Chen *et al.*, "Design and implementation of network test platform based on network function virtualization," *Chin. J. Comput.*, vol. 41, no. 9, pp. 2016–2028, 2018.

[17] M. Michalski, K. Cieslak, and M. Polak, "The system for large networks emulation with OSPF/BGP routers based on LXC," in *Proc. IEEE 16th Int. Conf. High Perform. Switching Routing (HPSR)*, Jul. 2015, pp. 1–4.

[18] G. Wang, Y. Qiao, X. Qiu, and L. Meng, "An improved network performance anomaly detection and localization algorithm," in *Proc. 14th Asia-Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Seoul, South Korea, Sep. 2012, pp. 1–4.

[19] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *Proc. ACM SIGCOMM*, 2004, pp. 219–230.

[20] T. Ahmed, B. Oreshkin, and M. Coates, "Machine learning approaches to network anomaly detection," in *Proc. 2nd Workshop Tackling Comput. Syst. Problems Mach. Learn.*, 2007, pp. 1–6.

[21] D. Liu, C.-H. Lung, I. Lambadaris, and N. Seddigh, "Network traffic anomaly detection using clustering techniques and performance comparison," in *Proc. 26th IEEE Can. Conf. Elect. Comput. Eng. (CCECE)*, Regina, SK, Canada, May 2013, pp. 1–4.

[22] K. Xie, C. Peng, X. Wang, G. Xie, and J. Wen, "Accurate recovery of internet traffic data under dynamic measurements," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Atlanta, GA, USA, May 2017, pp. 1–9.

[23] K. Xie *et al.*, "Fast tensor factorization for accurate Internet anomaly detection," *IEEE/ACM Trans. Netw.*, vol. 25, no. 6, pp. 3794–3807, Dec. 2017.

[24] Y. K. Qian, B. N. Li, and X. Luo, "Network anomaly detection method based on round-trip delay matrix subspace," *J. Nanjing Univ. Sci. Technol.*, pp. 215–224, 2015.

[25] C. Xia, Y. Shi, and Q. Zhao, "A new algorithm NA for IP traceback," *J. Comput. Res. Develop.*, vol. 41, no. 4, pp. 689–696, Apr. 2004.

[26] P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," in *Proc. ACM SIGCOMM Internet Meas. Workshop*, 2002, pp. 71–82.

[27] H. Tian, M. Roughan, Y. Sang, and H. Shen, "Diffusion wavelets-based analysis on traffic matrices," in *Proc. 12th Int. Conf. Parallel Distrib. Comput., Appl. Technol.*, Oct. 2011, pp. 116–121.

[28] A. Pavlidis, G. Sotiropoulos, K. Giotis, D. Kalogeras, and V. Maglaris, "NFV-compliant traffic monitoring and anomaly detection based on dispersed vantage points in shared network infrastructures," in *Proc. 4th IEEE Conf. Netw. Softwarization Workshops (NetSoft)*, Jun. 2018, pp. 197–201.

[29] M. Kourtis, G. Xilouris, G. Gardikis, and I. Koutras, "Statistical-based anomaly detection for NFV services," in *Proc. IEEE Conf. Netw. Function Virtualization Softw. Defined Netw. (NFV-SDN)*, Nov. 2016, pp. 161–166.

[30] A. Blaise, S. Wong, and A. H. Aghvami, "Virtual network function service chaining anomaly detection," in *Proc. 25th Int. Conf. Telecommun. (ICT)*, 2018, pp. 411–415.

[31] L. S. R. Sampaio, P. H. A. Faustini, A. S. Silva, L. Z. Granville, and A. Schaeffer-Filho, "Using NFV and reinforcement learning for anomalies detection and mitigation in SDN," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, 2018, pp. 432–437.

[32] M. Alberto *et al.*, "BRITE: Universal topology generation from a user's perspective," in *Proc. 9th IEEE Int. Symp. Modelind, Anal. Simulation Comput. Telecommun. Syst.*, 2001, pp. 346–356.

[33] K. Xie *et al.*, "Recover corrupted data in sensor networks: A matrix completion solution," *IEEE Trans. Mobile Comput.*, vol. 16, no. 5, pp. 1434–1448, May 2017.

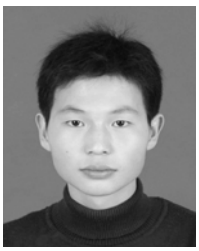
[34] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[35] W. Austin, D. Anderson, and J. Ghosh, "Fully supervised non-negative matrix factorization for feature extraction," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, Jul. 2018, pp. 5772–5775.

[36] L. Li *et al.*, "A first-principles approach to understanding the Internet's router-level topology," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun.*, New York, NY, USA, 2004, pp. 3–14.

JING CHEN is currently pursuing the master's degree with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China. She was with the School of Computer Science, Nanjing University of Posts and Telecommunications, from 2013 to 2017. Her main research interests include computer networks, network security, and software engineering.

MING CHEN was born in Nanjing, China, in 1956. He received the B.S. degree in communications engineering and the M.S. degree in information system from the University of Information Engineering, Zhengzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree in information system from the Institute of Communication Engineering, Nanjing, in 1991. He is currently a Professor and a Doctoral Supervisor with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing. His research interests include network architecture, UAV networks, network measurement, and future networks.



XIANGLIN WEI received the bachelor's degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2007, and the Ph.D. degree from the University of Science and Technology, Nanjing, in 2012. He is currently a Researcher with the Nanjing Telecommunication Technology Research Institute, Nanjing. His research interests include mobile edge computing, wireless network optimization, and the Internet of Things. He has served as an editorial member of many international journals and as a TPC Member of a number of international conferences. He has also organized a few special issues for many reputed journals.

BING CHEN received the B.S. and M.S. degrees from the Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, Jiangsu, China, in 1992 and 1995, respectively, and the Ph.D. degree from the College of Information Science and Technology, NUAA, where he is currently a Professor. His main research interests are computer networks and embedded systems.