# Anomaly-Aware Network Traffic Estimation via Outlier-Robust Tensor Completion

Qianqian Wang, Lei Chen, *Member, IEEE*, Qin Wang, *Member, IEEE*, Hongbo Zhu, *Member, IEEE*, and Xianbin Wang

***Abstract*—Accurately estimating network traffic from the partial measurements plays a crucial role in network management. However, the potential anomaly existing in real networks usually makes this goal difficult to achieve. Existing network traffic estimation methods generally impute network traffic independent of anomaly detection, which incurs significant performance degradation with network anomaly. To address this issue in the realistic network scenario, we propose a novel anomaly-aware network traffic estimation method to recover network traffic data concurrently with network anomaly detection. Specifically, by exploiting the inherent spatio-temporal characteristics, we first formulate the network traffic estimation as a low-rank tensor completion problem. Then, an outlier-robust tensor completion (OrTC) model is constructed by introducing both $L_{2,1}$-norm regularization and $L_F$-norm regularization, which can not only well fit the intrinsic low-rank property of real traffic data, but also is robust against both the dense noise and the sparse anomaly. Furthermore, an effective optimization algorithm OrTC-AM is designed to solve the non-convex and non-smooth OrTC model based on the popular alternating minimization method. Finally, the extensive experiments performed on the public dataset demonstrate that our proposed OrTC-AM method outperforms the previously widely used network traffic estimation methods.**

***Index Terms*—Network traffic estimation, network anomaly detection, low-rank tensor factorization, alternating minimization.**

Manuscript received March 11, 2020; revised July 16, 2020; accepted September 10, 2020. Date of publication September 21, 2020; date of current version December 9, 2020. This work was supported by the National Natural Science Foundation of China under Grant 61871446 and 61872190, the Key Software Engineering Discipline Construction Project of Jiangsu Province under Grant 010520200104 and the Discovery Program of the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN-2018-06254. The associate editor coordinating the review of this article and approving it for publication was M. Mellia. *(Corresponding author: Hongbo Zhu.)*

Qianqian Wang is with the School of Software Engineering, Jinling Institute of Technology, Nanjing 211169, China, also with the Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing 210023, China, and also with the Department of Electrical and Computer Engineering, Western University, London, ON N6A 5B9, Canada (e-mail: 2016010202@njupt.edu.cn).

Lei Chen is with the Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: chenlei@njupt.edu.cn).

Qin Wang and Hongbo Zhu are with the Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: wangqin@njupt.edu.cn; zhuhb@njupt.edu.cn).

Xianbin Wang is with the Department of Electrical and Computer Engineering, Western University, London, ON N6A 5B9, Canada (e-mail: xianbin.wang@uwo.ca).

Digital Object Identifier 10.1109/TNSM.2020.3024932

## I. Introduction

NETWORK traffic data provides a flow-time map, which characterizes the volumes of the traffic between origin and destination (OD) pairs in the networks. Accurately generating such a map will be the key enabler for effective routing and congestion control, risk analysis, security assurance, and proactive network failure prevention [1]. Therefore, obtaining complete and accurate traffic data is of the highest priority for network management.

However, due to the massive number of connected devices involved, it is impractical to exhaustively probe all OD pairs in large-scale Internet of Things (IoT) networks [2], [3]. An effective alternative strategy for network monitoring and management is achieved by just taking partial measurements from the full traffic data [4]. Moreover, the collected useful and accurate data could be even less because OD flows potentially undergo damages [5], willfully or accidentally, by diverse network conditions. Consequently, the observed real traffic measurements typically experience data missing and data pollution [6]. When analyzing such realistic measurements, there are two main technical challenges that need to solve for acquiring complete and accurate traffic data for large-scale IoT networks. One challenge is to impute the missing traffic data from the partial measurements, which can be referred to as the traffic data completion problem. The other one is to reconstruct the traffic data from the corrupted traffic measurements, e.g., the outlier data, which can be referred to as the traffic data recovery problem. Often these two problems exist simultaneously in practical scenarios.

Many efforts have been dedicated to addressing such two challenges in network monitoring and management. By utilizing the low-rank property of the traffic data, several studies formulated the traffic data estimation as a compressive sensing-based problem [7], [8] or a matrix completion-based problem [9], [10]. Besides these techniques, tensor-based approaches [11]–[14], as the matrix expansion in more than two-dimensional space, have been proven as a more effective tool to impute the missing traffic data. However, these completion approaches did not explicitly account for the outlier measurements and could incur significant performance degradation under serious corruption. On the other hand, to alleviate the impact of the anomaly (outlier) on traffic data recovery, some researchers proposed to utilize the prior assumption that the traffic anomaly follows Laplacian distribution [15], [16]. Therefore, researches can employ the $L_0$-norm or its convex approximation $L_1$-norm of tensor to fit the traffic anomaly, such as flash crowds, denial-of-service attacks, and spreading of worms, which usually affects a small fraction of data points with a relatively large perturbation. However, in most realistic scenarios, the traffic data is interfered with by the sparse gross anomaly and the dense slight noise simultaneously. The noise, which is usually caused by environmental interference, commonly exists in the entire traffic data with a relatively small value. Moreover, $L_1$-norm regularization based on the assumption of the Laplacian distribution can effectively detect the uniformly and randomly distributed sparse anomaly, but usually degrades the performance under the condition of the non-uniform distributed anomaly. In the practical network, the sparse anomaly is more likely to follow the structural distribution, e.g., the denial-of-service attacks usually exist in a certain part of nodes.

To overcome the abovementioned issues, this article presents a novel anomaly-aware network traffic estimation approach to recover traffic data and detect network anomaly simultaneously, which takes full advantage of the potential relationship between the two tasks to support each other for better performance. Specifically, by exploiting the inherent spatio-temporal correlation characteristics of the traffic data, we model the network traffic estimation problem as a low-rank tensor completion problem. In order to deal with the challenge of data missing, we propose a weighted CANDECOMP/PARAFAC (CP) tensor factorization (CP-wALS) algorithm based on the alternating least squares (ALS) method, which is an effective approach for large-scale tensor completion. Then, an outlier-robust tensor completion (OrTC) model is constructed by introducing $L_{2,1}$-norm regularization instead of $L_1$-norm to smooth the traffic anomaly. Such an idea is inspired by recent anomaly modeling methods designed for matrix-based data estimation problems [17], [18], which has been verified to be effective in anomaly detection. In addition, to improve the robustness of the proposed OrTC model, we further employ $L_F$-norm to fit the dense traffic noise, which has been widely used in traffic data estimation problems [19]. In this way, the proposed OrTC model can not only well fit the intrinsic low-rank property of traffic data, but also is robust against both the sparse anomaly and the dense noise. Furthermore, we employ the alternating minimization (AM) method [20] to design an effective algorithm OrTC-AM to solve the non-convex and non-smooth OrTC model. Finally, experimental results performed on the public dataset reveal that, compared with the classic tensor completion methods and the state-of-the-art traffic estimation methods, our proposed OrTC-AM method achieves the best performance from the perspective of the traffic data recovery and network anomaly detection.

The primary contributions of this work can be summarized as follows:

- A novel outlier-robust tensor completion (OrTC) model is proposed for estimating traffic data concurrently with network anomaly detection, which takes full advantage of the potential relationship between the two tasks to achieve better performance. The $L_{2,1}$-norm regularization and $L_F$-norm regularization are employed to fit sparse anomaly and dense noise, respectively, enabling our proposed OrTC model to adaptively handle a wider range of anomaly structures beyond existing methods.
- Based on the popular alternating minimization (AM) method, an effective algorithm OrTC-AM is designed to solve the non-convex and non-smooth OrTC model, which can not only well recover traffic data but also detect anomaly accurately. Furthermore, to deal with the challenge of large-scale traffic data estimation in practice, a weighted CP tensor factorization (CP-wALS) algorithm is proposed to address the tensor completion problem in the OrTC model.
- Using real traffic trace data, we compare our proposed algorithm with the state-of-art tensor-based completion and recovery algorithms from the perspective of traffic data recovery and network anomaly detection. Extensive experiments verify the superior estimation performance of our proposed OrTC-AM algorithm in the coexistence of network anomaly and noise.

The rest of this article is organized as follows. In Section II, we introduce the current research advances about traffic data estimation methods and tensor-based methods. Section III describes the notations used in this article. Section IV constructs the OrTC model and presents its problem formulation. In Section V, an optimization algorithm OrTC-AM is designed successively to solve the proposed OrTC model. In Section VI, a series of simulation experiments are conducted to evaluate the performance of OrTC-AM on the real traffic dataset. Finally, the conclusions are drawn in Section VII.

## II. Related Work

In this section, we review the related work on the network traffic data estimation and identify the differences in our work from the existing researches.

### A. Traffic Data Completion and Anomaly Detection

To capture spatio-temporal features in the traffic data, SRMF [21] proposed the first spatio-temporal model of traffic matrices (TMs). It found low-rank approximations to TMs, and recovered the missing data with the spatio-temporal operation and local interpolation. Furthermore, to alleviate the impact of noise on network traffic estimation, a set of studies have been made to extend the original matrix completion (MC) model to the noise-tolerant MC models. For instance, some researchers assumed that the sampled matrix was corrupted by pure Gaussian noise [22], or by pure outlier noise [23], [24]. By employing the Mixture of Gaussian distribution to model the complex noise, Xiao *et al.* [17] presented a novel noise-immune MC model to estimate network traffic. However, a two-dimension matrix is limited in capturing comprehensive correlations hidden in the traffic data, and thus the recovery performance of matrix-based methods always significantly decreases when the missing ratio is high.

To overcome the shortcomings of matrix-based traffic data estimation methods, some researchers exploited the traffic hidden higher-dimensional characteristics to improve the quality of the missing data recovery. Inspired by SRMF, Zhou *et al.* [12] proposed a novel spatio-temporal tensor completion method to recover the missing entries in tensors of traffic data. They utilized spatio-temporal properties information to regularize the tensor decomposition procedure, resulting in a unified framework for traffic tensor completion. Xie *et al.* [14] further modeled the Internet traffic data to exploit well the hidden structures (temporal stability, spatial correlation feature, and traffic periodic pattern) of the traffic data, and formulated the traffic data estimation problem as a low-rank tensor completion problem. Furthermore, Mardani *et al.* [19] completed the missing data from the sequentially acquired traffic data by considering the impact of the Gaussian noise. TensorDet was designed in [15] to focus on robust tensor recovery problems, which was a simple and effective way for faster low-rank tensor factorization and more accurate sparse anomaly detection. In [16], the authors presented a new proposal for an online subspace tracking of the Hankelized time-structured traffic tensor, which can complete the missing traffic data and detect the network anomaly simultaneously. However, the prior assumption of noise distribution as Gaussian distribution or Laplacian distribution probably leads to a significant performance degradation since the traffic data is not merely interfered with by a specific noise in most realistic scenarios.

### B. Other Tensor-Based Solutions

Tensor-based methods have been successfully utilized in various fields, such as computer vision [25], deep neural networks [26], and road traffic [27]. Liu *et al.* [28] first defined the nuclear norm of a tensor and translated tensor completion into a convex optimization problem. In [29], the authors proposed a tensor robust principal component analysis (TRPCA) method to recover the low-rank and sparse outlier components from the noisy data. In addition, Yokota *et al.* introduced tensor total variation in [30] to solve the noisy tensor completion problems based on tensor nuclear norm. Besides using the nuclear norm, some tensor factorization approaches were proposed for the tensor completion problem, such as CP decomposition [31] and Tucker decomposition [32]. For example, Tan *et al.* [33] proposed a tensor decomposition based imputation method to estimate the missing value in transportation traffic based on Tucker decomposition. Although the above solutions can complete missing data or detect anomaly effectively, they fail to be fully applied to anomaly-aware traffic data estimation in this work since they merely cope with one traffic data problem.

As an extension of [30], Yokota and Hontani [34] further solved the noise inequality constrained convex optimization problem by introducing Gaussian and Laplacian distributions respectively. Besides, a generative CP tensor completion framework was constructed in [35] in the presence of noise and outliers. Both [34] and [35] can impute the missing data and detect the anomaly simultaneously; however, due to the characteristics of the network traffic data, these solutions can not be directly applied to this work because of the following reasons. First, as explicated in Section I, $L_1$-norm regularization is not applicable to smooth the sparse network anomaly that usually follows a structural distribution. Second, the computation complexity of these two methods is relatively high, and thus not scalable for the large-scale network traffic data. Specifically, the singular value decomposition involved in nuclear norm based tensor completion [34] is of high time complexity, and the Khatri-Rao product involved in [35] for updating factor matrices is of high space complexity.

Based on the aforementioned analysis, we propose a novel outlier-robust tensor completion (OrTC) model for anomaly-aware traffic data estimation in this article by introducing both $L_{2,1}$-norm regularization and $L_F$-norm regularization. The proposed OrTC model can not only well fit the intrinsic low-rank property of traffic data, but also detect the sparse anomaly accurately, especially for the sparse structural anomaly. Furthermore, to deal with the challenge of large-scale traffic data estimation, a weighted CP tensor factorization (CP-wALS) algorithm is developed to address the tensor completion problem in the OrTC model.

## III. Preliminaries

The notation used in this article is described as follows. Scalars are represented with lowercase letters $(a, b, \ldots)$, vectors are expressed as boldface lowercase $(\mathbf{a}, \mathbf{b}, \ldots)$, and matrices are denoted by boldface capitals $(\mathbf{A}, \mathbf{B}, \ldots)$. The elements of a vector/matrix are given by the symbolic name of the vector/matrix with indexes in subscript. For example, the $i$-th entry of a vector $a$ is denoted by $a_i$, and element $(i, j)$ of a matrix $\mathbf{A}$ is expressed as $a_{ij}$.

Next, we introduce some basic concepts related to the tensor used in this article.

*Definition 1:* A tensor is a higher-order generalization of a vector (first-order tensor) and a matrix (second-order tensor). An N-way or $N^{th}$-order tensor is expressed as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, where N is the order of $\mathcal{X}$, also called way or mode. In this article, we focus on the 3-way tensor, i.e., $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, and element $(i_1, i_2, i_3)$ of $\mathcal{X}$ is denoted by $x_{i_1 i_2 i_3}$, $i_n \in \{1, 2, \ldots, I_n\}$ with $1 \le n \le 3$.

*Definition 2:* Slices are two-dimensional sub-arrays, defined by fixing all indexes but two. A 3-way tensor $\mathcal{X}$ has horizontal, lateral and frontal slices, which are denoted by $\mathbf{X}_{i_1::}$, $\mathbf{X}_{:i_2:}$ and $\mathbf{X}_{::i_3}$, respectively. In this article, we represent the frontal slice $\mathbf{X}_{::i_3}$ as $\mathbf{X}_{i_3}$.

*Definition 3:* The idea of CP factorization is to express a tensor as the sum of a finite number of rank one tensors. A 3-way tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ can be expressed as

$$\mathcal{X} = [[\mathbf{U}, \mathbf{V}, \mathbf{T}]] = \sum_{r=1}^{R} u_r \circ v_r \circ t_r \tag{1}$$

with an entry calculated by

$$x_{i_1 i_2 i_3} = \sum_{r=1}^{R} u_{i_1 r} v_{i_2 r} t_{k i_3 r}, \tag{2}$$

where $R > 0$, $u_{i_1 r}$, $v_{i_2 r}$, $t_{i_3 r}$ are the $i_1$-th, $i_2$-th, and $i_3$-th entry of vectors $u_r \in \mathbb{R}^{I_1}$, $v_r \in \mathbb{R}^{I_2}$, and $t_r \in \mathbb{R}^{I_3}$, respectively. The operator $\circ$ is the outer product of two vectors. By collecting the vectors in the rank one components, we have tensor factor matrices $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_R] \in \mathbb{R}^{I_1 \times R}$, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_R] \in \mathbb{R}^{I_2 \times R}$, and $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_R] \in \mathbb{R}^{I_3 \times R}$.

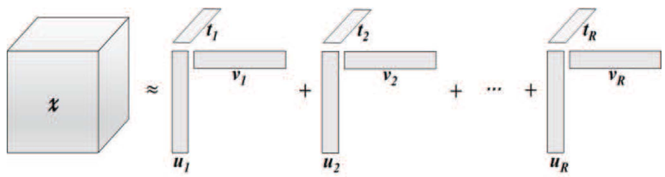



Fig. 1. CP factorization of a 3-way tensor.

In this article, we design the traffic data estimation method based on the CP factorization illustrated in Fig. 1.

*Definition 4:* Similar to the $L_F$-norm of a matrix, $L_F$-norm $\|\mathcal{X}\|_F$ for a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is defined as

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i_i=1}^{I_1} \sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} x_{i_1 i_2 i_3}^2}. \tag{3}$$

*Definition 5:* Tensor $L_{2,1}$-norm for an arbitrary tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, given by $\|\mathcal{X}\|_{2,1}$, is defined as (without loss of generality, we assume that outliers are distributed along the $3^{rd}$ dimension of a tensor)

$$\|\mathcal{X}\|_{2,1} = \sum_{i_3=1}^{I_3} \|\mathbf{X}_{::i_3}\|_F = \sum_{i_3=1}^{I_3} \sqrt{\sum_{i_i=1}^{I_1} \sum_{i_2=1}^{I_2} x_{i_1 i_2}^2}. \tag{4}$$

## IV. System Model and Problem Formulation

In this section, we first present our tensor-based traffic data estimation model and then formulate its problem.

### A. Traffic Estimation Model

As the empirical study on the real traffic data conducted by Xie *et al.* [14], the traffic data holds the features of temporal stability, spatial correlation, and periodicity pattern. To fully exploit such hidden features, we model the traffic data as a 3-way tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, where there are $I_1$ days to consider with each day having $I_2$ time intervals, and $I_3$ corresponds to $N \times N$ OD pairs.

Due to the sample-based traffic monitoring and the unavoidable data missing resulted from severe communication conditions, only partial measurements could be collected. By utilizing the prior low-rank characteristic of traffic data, we can easily employ the existing tensor completion models to recover the full traffic data. However, as stated in Section I, the practical sampled tensor $\mathcal{X}$ often suffers from the corruption caused by the complex environmental interference, malicious attacks, etc. Therefore, these corruptions, brought not only by the sparse gross anomaly but also by the dense slight noise, significantly degrade the estimation accuracy. To improve the precision of the traffic data reconstruction and the anomaly detection, we propose a novel anomaly-aware traffic estimation method by considering that the collected tensor $\mathcal{X}$ consists of the true underlying low-rank tensor $\hat{\mathcal{X}}$, the sparse anomaly tensor $\mathcal{A}$, and the non-sparse noise tensor $\mathcal{N}$, given by

$$\mathbf{P}_\Omega(\mathcal{X}) = \mathbf{P}_\Omega\Big(\hat{\mathcal{X}} + \mathcal{A} + \mathcal{N}\Big), \tag{5}$$

where $\Omega$ is the indices set of the observed traffic data, $\mathbf{P}_\Omega$ denotes the projection operator such that the $(i_1, i_2, i_3)$-th element of $\mathcal{X}$ equals to $x_{i_1 i_2 i_3}$, and zero if otherwise, i.e., there is no observed traffic data between a particular pair of nodes in a given time slot.

In order to avoid confusion, we summarize the related notations and their semantic meanings in Table I.

TABLE I
COMMONLY USED NOTATIONS

| Notation | Description |
|---|---|
| $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ | The corrupted tensor |
| $\hat{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ | The underlying true tensor |
| $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ | The anomaly tensor |
| $\mathcal{N} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ | The noise tensor |
| $\mathcal{W} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ | The indices tensor |
| $\mathbf{A}_{i_3} \in \mathbb{R}^{I_1 \times I_2}$ | The $i_3$-th frontal slice of tensor $\mathcal{A}$ |
| $\mathbf{U} \in \mathbb{R}^{I_1 \times R}$ | The $1^{st}$ way factor matrix of CP form |
| $\mathbf{V} \in \mathbb{R}^{I_2 \times R}$ | The $2^{nd}$ way factor matrix of CP form |
| $\mathbf{T} \in \mathbb{R}^{I_3 \times R}$ | The $3^{rd}$ way factor matrix of CP form |

### B. Problem Formulation

Our primary goal is to design an effective network traffic estimation method to reconstruct the true underlying traffic data and detect the network anomaly simultaneously, where the traffic data reconstruction includes imputing the missing measurements as well as denoising the inaccurate ones.

In order to deal with the challenge of traffic data completion, we employ the CP tensor factorization to solve tensor completion problems, which is an effective approach for large-scale traffic data [16], [19]. CP tensor factorization decomposes a tensor as the sum of *R* rank-one tensors. Therefore, $\hat{\mathcal{X}}$, denoting the intrinsic low-dimensionality tensor to be recovered, can be represented in the CP decomposition form as *Definition 3*, and the rank *R* satisfying $R \ll \min\{I_1, I_2, I_3\}$ is the constraint of the low-rank property. Furthermore, different from the existing $\mathrm{L}_1$-norm based anomaly detection methods, we use tensor $\mathrm{L}_{2,1}$-norm to characterize the sparsity property of the network anomaly. Although $\mathrm{L}_1$-norm regularization can effectively detect the uniformly and randomly distributed sparse anomaly, it generally degrades the performance under the condition of the non-uniform distributed anomaly. Practically, the sparse network anomaly usually tends to be the structural distribution rather than the uniform distribution. For instance, the denial-of-service attacks frequently exist in a certain part of OD pairs. Hence, we adopt $\mathrm{L}_{2,1}$-norm to smooth slice-wise anomaly, which has already been verified to be effective in matrix-based data estimation problems [17], [18]. Without loss of generality, we assume the frontal-slice sparsity structural characteristic since the anomaly always exists in $3^{rd}$ dimension, i.e., a certain part of OD pairs. Note that outliers can also be distributed along the other dimensions, and the proposed model and analysis can be applied directly. Finally, the dense noise with a relatively small value generally follows a Gaussian distribution, and thus we introduce $\mathrm{L}_F$-norm to fit the Gaussian noise.

Based on the preceding analysis, the network traffic estimation problem in the coexistence of the network noise

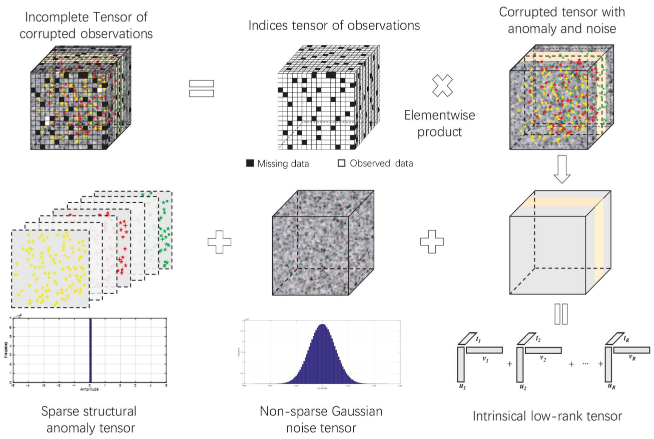



Fig. 2. Illustration of OrTC model.

and potential anomaly can be formulated as the following outlier-robust tensor completion (OrTC) model, i.e.,

$$\begin{aligned} \min_{\hat{\mathcal{X}},\mathcal{A},\mathcal{N}} \quad & \lambda\|\mathcal{A}\|_{2,1} + \frac{1}{2}\|\mathcal{N}\|_F^2 \\ s.t. \quad & \mathbf{P}_\Omega(\mathcal{X}) = \mathbf{P}_\Omega\left(\hat{\mathcal{X}} + \mathcal{A} + \mathcal{N}\right) \\ & \hat{\mathcal{X}} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{t}_r, R \ll \min\{I_1, I_2, I_3\} \end{aligned} \tag{6}$$

where $\lambda$ is a tunable parameter to balance the anomaly term with the noise term. The key of the OrTC model is to recover the true underlying traffic data and detect the network anomaly through minimizing the $\mathrm{L}_{2,1}$-norm of the anomaly tensor $\mathcal{A}$ and $\mathrm{L}_F$-norm of the noise tensor $\mathcal{N}$. The constraint in (6) ensures that the observed traffic tensor consists of the true underlying low-rank tensor $\hat{\mathcal{X}}$, the sparse anomaly tensor $\mathcal{A}$, and the non-sparse noise tensor $\mathcal{N}$, where the true underlying low-rank tensor $\hat{\mathcal{X}}$ is in a low-rank CP decomposition form. Therefore, we can illustrate the OrTC model as Fig. 2.

## V. Optimization Algorithm for OrTC Model

In this section, we propose an optimization approach to solve the OrTC model with the purpose of estimating the underlying traffic data concurrently with network anomaly detection.

To effectively solve the problem in (6), we can convert it to an unconstrained form according to the penalty function method [36], given by

$$\begin{aligned} \min_{\hat{\mathcal{X}},\mathcal{A},\mathcal{N}} \quad & \lambda\|\mathcal{A}\|_{2,1} + \frac{1}{2}\|\mathcal{N}\|_F^2 + \frac{\mu}{2}\left\|\mathbf{P}_\Omega\left(\mathcal{X} - \hat{\mathcal{X}} - \mathcal{A} - \mathcal{N}\right)\right\|_F^2 \\ s.t. \quad & \hat{\mathcal{X}} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{t}_r, \end{aligned} \tag{7}$$

where $\mathcal{X}, \mathcal{A}, \mathcal{N} \in \mathbb{R}^{I_1\times I_2\times I_3}$ denote the collected traffic tensor, structural anomaly tensor, and noise tensor, respectively. $\hat{\mathcal{X}}$ is the objective tensor to be recovered, which can be expressed in the CP decomposition form as a constraint. $\mu$ is a tunable parameter that balances the penalty function with the anomaly and noise terms.

We define the objective function as

$$\begin{aligned} \mathcal{L}\left(\hat{\mathcal{X}}, \mathcal{A}, \mathcal{N}\right) = \lambda\|\mathcal{A}\|_{2,1} + \frac{1}{2}\|\mathcal{N}\|_F^2 \\ + \frac{\mu}{2}\left\|\mathbf{P}_\Omega\left(\mathcal{X} - \hat{\mathcal{X}} - \mathcal{A} - \mathcal{N}\right)\right\|_F^2, \end{aligned} \tag{8}$$

and the problem in (7) can be rewritten as

$$\left(\hat{\mathcal{X}}^*, \mathcal{A}^*, \mathcal{N}^*\right) = \arg\min_{\hat{\mathcal{X}},\mathcal{A},\mathcal{N}} \mathcal{L}\left(\hat{\mathcal{X}}, \mathcal{A}, \mathcal{N}\right). \tag{9}$$

Since the OrTC model is non-convex and non-smooth, we employ the popular alternating minimization (AM) method to solve (9) iteratively as

$$\begin{cases} \mathcal{A}^{k+1} = \arg\min_{\mathcal{A}} \mathcal{L}\left(\hat{\mathcal{X}}^k, \mathcal{A}, \mathcal{N}^k\right) \\ \mathcal{N}^{k+1} = \arg\min_{\mathcal{N}} \mathcal{L}\left(\hat{\mathcal{X}}^k, \mathcal{A}^{k+1}, \mathcal{N}\right) \\ \hat{\mathcal{X}}^{k+1} = \arg\min_{\hat{\mathcal{X}}} \mathcal{L}\left(\hat{\mathcal{X}}, \mathcal{A}^{k+1}, \mathcal{N}^{k+1}\right), \end{cases} \tag{10}$$

where our proposed OrTC-AM method updates each tensor in a proximal-linear updating way while fixing the remaining tensors at their last updated value. Specifically, during each iteration $k$, the tensors are updated as follows.

### A. Subproblem 1: Update Anomaly Tensor $\mathcal{A}$

In this subsection, we fix $\hat{\mathcal{X}}$ and $\mathcal{N}$ and solve the subproblem 1 formulated in (10), i.e., $\mathcal{A}^{k+1} = \arg\min_{\mathcal{A}} \mathcal{L}(\hat{\mathcal{X}}^k, \mathcal{A}, \mathcal{N}^k)$.

*Theorem 1:* Each frontal slice $\mathbf{A}_{i_3}$ of tensor $\mathcal{A}$ can be calculated by

$$\mathbf{A}_{i_3}{}^* = \max\left\{1 - \frac{\lambda}{\mu\|\mathbf{P}_\Omega(\mathbf{O}_{i_3})\|_F}, 0\right\}\mathbf{P}_\Omega(\mathbf{O}_{i_3}), \tag{11}$$

where $\mathbf{O}_{i_3}$ is the $i_3$-th frontal slice of tensor $\mathcal{O}$ that is defined as $\mathcal{O} = \mathcal{X} - \hat{\mathcal{X}} - \mathcal{N}$.

*Proof:* By fixing $\hat{\mathcal{X}}$ and $\mathcal{N}$, we can update $\mathcal{A}$ by

$$\mathcal{A}^* = \arg\min_{\mathcal{A}} \lambda\|\mathcal{A}\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{P}_\Omega\left(\mathcal{X} - \hat{\mathcal{X}} - \mathcal{A} - \mathcal{N}\right)\right\|_F^2. \tag{12}$$

We define $\mathcal{O} = \mathcal{X} - \hat{\mathcal{X}} - \mathcal{N}$. Since the outliers have the frontal-slice sparsity structural characteristic, we unfold $\mathcal{A}$ along the $3^{rd}$ order, and then (12) can be expressed as

$$\mathcal{A}^* = \arg\min_{\mathcal{A}} \sum_{i_3=1}^{I_3}\left(\lambda\|\mathbf{A}_{i_3}\|_F + \frac{\mu}{2}\|\mathbf{P}_\Omega(\mathbf{O}_{i_3} - \mathbf{A}_{i_3})\|_F^2\right), \tag{13}$$

where the matrix $\mathbf{A}_{i_3} \in \mathbb{R}^{I_1\times I_2}$ denotes the $i_3$-th frontal slice of the tensor $\mathcal{A}$, i.e., $\mathbf{A}_{::i_3}$. The tensor of $I_1 \times I_2 \times I_3$ can be divided into $I_3$ frontal slices which can be optimized separately as

$$\mathbf{A}_{i_3}^* = \arg\min_{\mathbf{A}_{i_3}} \lambda\|\mathbf{A}_{i_3}\|_F + \frac{\mu}{2}\|\mathbf{P}_\Omega(\mathbf{O}_{i_3} - \mathbf{A}_{i_3})\|_F^2. \tag{14}$$

We define $f(\mathbf{A}_{i_3}) = \lambda\|\mathbf{A}_{i_3}\|_F + \mu\|\mathbf{P}_\Omega(\mathbf{O}_{i_3} - \mathbf{A}_{i_3})\|_F^2/2$. According to the definition of the subdifferential, the subdifferential of function $f(\mathbf{A}_{i_3})$ at $\mathbf{A}_{i_30}$ is

$$\partial f(\mathbf{A}_{i_30}) = \begin{cases} \mu\mathbf{P}_\Omega(\mathbf{A}_{i_30} - \mathbf{O}_{i_3}) + \frac{\lambda}{\|\mathbf{A}_{i_30}\|_F}\mathbf{A}_{i_30} & \mathbf{A}_{i_30} \neq 0 \\ \mu\mathbf{P}_\Omega(\mathbf{A}_{i_30} - \mathbf{O}_{i_3}) + \mathbf{Y} & \mathbf{A}_{i_30} = 0 \end{cases}. \tag{15}$$

where $\mathbf{Y}$ satisfies $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2}$ and $\|\mathbf{Y}\|_F \leq \lambda$.

If and only if $0 \in \partial f(\mathbf{A}_{i_30})$, $\mathbf{A}_{i_30}$ is the global minimum of the function $f(\mathbf{A}_{i_3})$, which can be expressed as

$$\begin{cases} \mu\mathbf{P}_\Omega(\mathbf{A}_{i_30} - \mathbf{O}_{i_3}) + \frac{\lambda}{\|\mathbf{A}_{i_30}\|_F}\mathbf{A}_{i_30} = 0 & \mathbf{A}_{i_30} \neq 0 \\ \mu\mathbf{P}_\Omega(\mathbf{A}_{i_30} - \mathbf{O}_{i_3}) + \mathbf{Y} = 0 \text{ and } \|\mathbf{Y}\|_F \leq \lambda & \mathbf{A}_{i_30} = 0. \end{cases} \tag{16}$$

Then, we can derive out

1) when $\|\mathbf{P}_\Omega(\mathbf{O}_{i_3})\|_F \leq \lambda/\mu$, $\mathbf{A}_{i_30} = 0$;
2) when $\|\mathbf{P}_\Omega(\mathbf{O}_{i_3})\|_F > \lambda/\mu$, $\mathbf{A}_{i_30} \neq 0$, we can get

$$\mu\mathbf{P}_\Omega(\mathbf{A}_{i_30} - \mathbf{O}_{i_3}) + \frac{\lambda}{\|\mathbf{A}_{i_30}\|_F}\mathbf{A}_{i_30} = 0. \tag{17}$$

Since (17) is non-smooth, we need to discuss the solution of problem in (17) under different conditions. When $(i_1, i_2, i_3) \notin \Omega$, $(a_{i_30})_{i_1i_2} = 0$; otherwise, $(a_{i_30})_{i_1i_2} = \mu o_{i_1i_2}/(\mu + \lambda/\|\mathbf{A}_{i_30}\|_F)$, i.e.,

$$\mathbf{A}_{i_30} = \frac{\mu\mathbf{P}_\Omega(\mathbf{O}_{i_3})}{\mu + \lambda/\|\mathbf{A}_{i_30}\|_F}. \tag{18}$$

Furthermore, we conduct $\|\mathbf{A}_{i_30}\|_F$ by

$$\|\mathbf{A}_{i_30}\|_F = \left\|\frac{\mu\mathbf{P}_\Omega(\mathbf{O}_{i_3})}{\mu + \lambda/\|\mathbf{A}_{i_30}\|_F}\right\|_F = \frac{\mu\|\mathbf{P}_\Omega(\mathbf{O}_{i_3})\|_F}{\mu + \lambda/\|\mathbf{A}_{i_30}\|_F}, \tag{19}$$

and the reduced representation of $\|\mathbf{A}_{i_30}\|_F$ is

$$\|\mathbf{A}_{i_30}\|_F = \|\mathbf{P}_\Omega(\mathbf{O}_{i_3})\|_F - \frac{\lambda}{\mu}. \tag{20}$$

Substituting (20) into (18), we can update each frontal slice $\mathbf{A}_{i_3}$ of tensor $\mathcal{A}$ as

$$\mathbf{A}_{i_3} = \left(1 - \frac{\lambda}{\mu\|\mathbf{P}_\Omega(\mathbf{O}_{i_3})\|_F}\right)\mathbf{P}_\Omega(\mathbf{O}_{i_3}). \tag{21}$$

Jointly considering the condition 1) and 2), we can deduce the solution of subproblem 1 as (11). This finishes the proof of Theorem 1. ■

Next, we will conduct subproblem 2 to update the noise tensor $\mathcal{N}$.

### B. Subproblem 2 - Update Noise Tensor $\mathcal{N}$

In this subsection, we fix $\hat{\mathcal{X}}$ and $\mathcal{A}$ and solve the subproblem 2 formulated in (10), i.e., $\mathcal{N}^{k+1} = \arg\min_{\mathcal{N}} \mathcal{L}(\hat{\mathcal{X}}^k, \mathcal{A}^{k+1}, \mathcal{N})$.

*Theorem 2:* The noise tensor $\mathcal{N}$ can be updated by

$$\mathcal{N}^* = \frac{\mu}{1+\mu}\mathbf{P}_\Omega(\mathcal{P}), \tag{22}$$

where we define $\mathcal{P} = \mathcal{X} - \hat{\mathcal{X}} - \mathcal{A}$.

*Proof:* By fixing $\hat{\mathcal{X}}$ and $\mathcal{A}$, we can update $\mathcal{N}$ by

$$\mathcal{N}^* = \arg\min_{\mathcal{N}} \frac{1}{2}\|\mathcal{N}\|_F^2 + \frac{\mu}{2}\left\|\mathbf{P}_\Omega\left(\mathcal{X} - \hat{\mathcal{X}} - \mathcal{A} - \mathcal{N}\right)\right\|_F^2. \tag{23}$$

Denoting $f(\mathcal{N})$ as inner objective to be minimized, the estimate $\mathcal{N}$ is obtained in a closed form by setting $\partial f(\mathcal{N})/\partial\mathcal{N} = 0$. Similar to problem in (17), we need to discuss the optimal solution of $\mathcal{N}$ under different conditions. When $(i_1, i_2, i_3) \notin \Omega$, $n_{i_1i_2i_3} = 0$; otherwise, $n_{i_1i_2i_3} = \mu p_{i_1i_2i_3}/(1+\mu)$, where the element $p_{i_1i_2i_3}$ is the $(i_1, i_2, i_3)$-th of the tensor $\mathcal{P} = \mathcal{X} - \hat{\mathcal{X}} - \mathcal{A}$. Consequently, we can obtain the updated $\mathcal{N}^*$ as given in (22).

This finishes the proof of Theorem 2. Finally, we will conduct subproblem 3 to recover low-rank tensor $\hat{\mathcal{X}}$. ■

### C. Subproblem 3 - Update Low-Rank Tensor $\hat{\mathcal{X}}$

In this subsection, we fix $\mathcal{A}$ and $\mathcal{N}$ and solve the subproblem 3 formulated in (10), i.e., $\hat{\mathcal{X}}^{k+1} = \arg\min_{\hat{\mathcal{X}}} \mathcal{L}(\hat{\mathcal{X}}, \mathcal{A}^{k+1}, \mathcal{N}^{k+1})$.

*Theorem 3:* The intrinsical low-rank tensor $\hat{\mathcal{X}}$ can be recovered by $\hat{\mathcal{X}} = [[\mathbf{U}, \mathbf{V}, \mathbf{T}]]$, and each row of the factor matrices $\mathbf{U}$ can be updated as

$$\begin{aligned}\mathbf{u}_{i_1}^* &= \mathbf{q}_{(1)i_1}\mathbf{W}_{(1)}^{i_1}(\mathbf{T}, \mathbf{V}) \\ &\quad \times \left(\mathbf{W}_{(1)}^{i_1}(\mathbf{T}, \mathbf{V})^T\mathbf{W}_{(1)}^{i_1}(\mathbf{T}, \mathbf{V})\right)^\dagger,\end{aligned} \tag{24}$$

where $\mathbf{W}_{(1)}^{i_1}(\mathbf{T}, \mathbf{V}) = \operatorname{diag}(\mathbf{w}_{(1)i_1})\mathbf{T} \odot \mathbf{V}$. The operation $\dagger$ and $\odot$ indicate the Moore-Penrose pseudoinverse and the Khatri-Rao product, respectively. $\mathbf{q}_{(1)i_1}$ is the $i_1$-th row of the mode-1 unfolding matrix of the tensor to be completed, which is defined as $\mathcal{Q} = \mathcal{X} - \mathcal{A} - \mathcal{N}$. Similarly, $\mathbf{w}_{(1)i_1}$ is the $i_1$-th row of the mode-1 unfolding matrix of the indices tensor $\mathcal{W}$. The element $w_{i_1i_2i_3}$ of $\mathcal{W} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ satisfies

$$w_{i_1i_2i_3} = \begin{cases} 1 & \text{if } (i_1i_2i_3) \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

In the same way, each row of the factor matrices $\mathbf{V}$ and $\mathbf{T}$ can be calculated by

$$\begin{aligned}\mathbf{v}_{i_2}^* &= \mathbf{q}_{(2)i_2}\mathbf{W}_{(2)}^{i_2}(\mathbf{T}, \mathbf{U}) \\ &\quad \times \left(\mathbf{W}_{(2)}^{i_2}(\mathbf{T}, \mathbf{U})^T\mathbf{W}_{(2)}^{i_2}(\mathbf{T}, \mathbf{U})\right)^\dagger,\end{aligned} \tag{26}$$

where $\mathbf{W}_{(2)}^{i_2}(\mathbf{T}, \mathbf{U}) = \operatorname{diag}(\mathbf{w}_{(2)i_2})\mathbf{T} \odot \mathbf{U}$. $\mathbf{q}_{(2)i_2}$ and $\mathbf{w}_{(2)i_2}$ denote the $i_2$-th row of the mode-2 unfolding matrix of $\mathcal{Q}$ and $\mathcal{W}$, respectively.

$$\begin{aligned}\mathbf{t}_{i_3}^* &= \mathbf{q}_{(3)i_3}\mathbf{W}_{(3)}^{i_3}(\mathbf{V}, \mathbf{U}) \\ &\quad \times \left(\mathbf{W}_{(3)}^{i_3}(\mathbf{V}, \mathbf{U})^T\mathbf{W}_{(3)}^{i_3}(\mathbf{V}, \mathbf{U})\right)^\dagger,\end{aligned} \tag{27}$$

where $\mathbf{W}_{(3)}^{i_3}(\mathbf{V}, \mathbf{U}) = \operatorname{diag}(\mathbf{w}_{(3)i_3})\mathbf{V} \odot \mathbf{U}$. $\mathbf{q}_{(3)i_3}$ and $\mathbf{w}_{(3)i_3}$ denote the $i_3$-th row of the mode-3 unfolding matrix of $\mathcal{Q}$ and $\mathcal{W}$, respectively.

*Proof:* By fixing $\mathcal{A}$ and $\mathcal{N}$, we can update $\hat{\mathcal{X}}$ by

$$\hat{\mathcal{X}}^* = \arg\min_{\hat{\mathcal{X}}} \ \frac{1}{2}\left\|\mathbf{P}_\Omega\left(\mathcal{X} - \hat{\mathcal{X}} - \mathcal{A} - \mathcal{N}\right)\right\|_F^2$$
$$s.t. \ \ \hat{\mathcal{X}} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{t}_r. \quad (28)$$

The subproblem 3 can be perceived as a tensor completion problem. In this article, we employ the CP tensor factorization to solve tensor completion problem, which is an effective approach for large-scale data. Then, we can optimize the factor matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{T}$ to solve the problem as

$$(\mathbf{U}^*, \mathbf{V}^*, \mathbf{T}^*) = \arg\min_{\mathbf{U},\mathbf{V},\mathbf{T}} \frac{1}{2}\left\|\mathbf{P}_\Omega\left(\mathcal{Q} - \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{t}_r\right)\right\|_F^2$$
$$= \arg\min_{\mathbf{U},\mathbf{V},\mathbf{T}} \frac{1}{2}\|\mathcal{W} * (\mathcal{Q} - [[\mathbf{U}, \mathbf{V}, \mathbf{T}]]\|_F^2, \quad (29)$$

where Hadamard product $*$ is the element-wise product.

Because of the effectiveness and implementation convenience of ALS algorithm [37], we design a weighted CP tensor factorization algorithm based on ALS, i.e., CP-wALS, to update the factor matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{T}$. The CP-wALS algorithm fixes all but one matrix at a time, and continues to repeat the entire procedure until some convergence criterion is satisfied. For example, we fix $\mathbf{V}$ and $\mathbf{T}$ to solve for $\mathbf{U}$, the problem reduces to a linear least-squares problem, and the problem in (29) can be rewritten in matrix form as

$$\mathbf{U}^* = \arg\min_{\mathbf{U}} \frac{1}{2}\left\|\mathbf{W}_{(1)} * \left(\mathbf{Q}_{(1)} - \mathbf{U}(\mathbf{T} \odot \mathbf{V})^T\right)\right\|_F^2$$
$$= \arg\min_{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_{I_1}} \frac{1}{2}\sum_{i_1=1}^{I_1} \left\|\mathbf{w}_{(1)i_1} * \left(\mathbf{q}_{(1)i_1} - \mathbf{u}_{i_1}(\mathbf{T} \odot \mathbf{V})^T\right)\right\|_F^2, \quad (30)$$

where the matrix $\mathbf{Q}_{(1)}$ is the mode-1 matricization of the tensor $\mathcal{Q}$ and likewise for $\mathbf{W}_{(1)}$. The vector $\mathbf{q}_{(1)i_1}$ is the $i_1$-th row of the matrix $\mathbf{Q}_{(1)}$, and likewise for $\mathbf{w}_{(1)i_1}$ and $\mathbf{u}_{i_1}$.

In this way, we can optimize the factor matrix $\mathbf{U}$ separately by each row $\mathbf{u}_{i_1}$, which can be deduced as

$$\begin{aligned} f(\mathbf{u}_{i_1}) &= \left\|\mathbf{w}_{(1)i_1} * \left(\mathbf{q}_{(1)i_1} - \mathbf{u}_{i_1}(\mathbf{T} \odot \mathbf{V})^T\right)\right\|_F^2 \\ &= \left\|\left(\mathbf{q}_{(1)i_1} - \mathbf{u}_{i_1}(\mathbf{T} \odot \mathbf{V})^T\right)\mathrm{diag}\left(\mathbf{w}_{(1)i_1}\right)\right\|_F^2 \\ &= \mathrm{tr}\left(\mathbf{q}_{(1)i_1}\mathrm{diag}\left(\mathbf{w}_{(1)i_1}\right)\mathbf{q}_{(1)i_1}^T\right) \\ &\quad - \mathrm{tr}\left(\mathbf{q}_{(1)i_1}\mathrm{diag}\left(\mathbf{w}_{(1)i_1}\right)(\mathbf{T} \odot \mathbf{V})\mathbf{u}_{i_1}^T\right) \\ &\quad - \mathrm{tr}\left(\mathbf{u}_{i_1}(\mathbf{T} \odot \mathbf{V})^T\mathrm{diag}\left(\mathbf{w}_{(1)i_1}\right)\mathbf{q}_{(1)i_1}^T\right) \\ &\quad + \mathrm{tr}\left(\mathbf{u}_{i_1}(\mathbf{T} \odot \mathbf{V})^T\mathrm{diag}\left(\mathbf{w}_{(1)i_1}\right)(\mathbf{T} \odot \mathbf{V})\mathbf{u}_{i_1}^T\right), \end{aligned} \quad (31)$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator.

According to trace derivative rules, the optimal solution of $\mathbf{u}_{i_1}$ is then given by setting $\partial f(\mathbf{u}_{i_1})/\partial \mathbf{u}_{i_1} = 0$, i.e.,

$$\begin{aligned} 0 = &-\mathbf{q}_{(1)i_1}\mathrm{diag}\left(\mathbf{w}_{(1)i_1}\right)(\mathbf{T} \odot \mathbf{V}) \\ &- \left((\mathbf{T} \odot \mathbf{V})^T\mathrm{diag}\left(\mathbf{w}_{(1)i_1}\right)\mathbf{q}_{(1)i_1}^T\right)^T \\ &+ \left((\mathbf{T} \odot \mathbf{V})^T\mathrm{diag}\left(\mathbf{w}_{(1)i_1}\right)(\mathbf{T} \odot \mathbf{V})\mathbf{u}_{i_1}^T\right)^T \\ &+ \mathbf{u}_{i_1}(\mathbf{T} \odot \mathbf{V})^T\mathrm{diag}\left(\mathbf{w}_{(1)i_1}\right)(\mathbf{T} \odot \mathbf{V}). \end{aligned} \quad (32)$$

Thus we have the solution of $\mathbf{u}_{i_1}^*$ as (24). Note that $\mathbf{W}_{(1)}^{i_1}(\mathbf{T}, \mathbf{V})^T\mathbf{W}_{(1)}^{i_1}(\mathbf{T}, \mathbf{V}) \in \mathbb{R}^{R\times R}$, and thus we need only calculate the Moore-Penrose pseudoinverse of an $R \times R$ matrix, which significantly decrease the computation complexity since $R \ll \min\{I_1, I_2, I_3\}$. In the same way, the low-rank property also contributes to reducing the computation complexity of other tensor calculations involved in (24), such as the Khatri-Rao product. Therefore, the CP-wALS algorithm can be scalable for the large-scale traffic data completion provided that the data is of low-rank property.

Similarly, we can update each row of the factor matrices $\mathbf{V}$ and $\mathbf{T}$ as (26) and (27), respectively. Then, the intrinsical low-rank tensor $\hat{\mathcal{X}}$ can be recovered by $\hat{\mathcal{X}} = [[\mathbf{U}, \mathbf{V}, \mathbf{T}]]$, and the tensor $\mathcal{Q}$ in next iteration $n+1$ can be updated as

$$\mathcal{Q}^{n+1} = \mathbf{P}_\Omega(\mathcal{Q}^n) + \mathbf{P}_{\overline{\Omega}}\left(\hat{\mathcal{X}}^n\right), \quad (33)$$

until the convergence criterion is satisfied. Here, $\overline{\Omega}$ denotes the complement set of $\Omega$, i.e., the index set of missing entries, and $\mathbf{P}_{\overline{\Omega}}(\hat{\mathcal{X}}^n)$ is the complement set of missing entry imputed by $\hat{\mathcal{X}}$. This finishes the proof of Theorem 3. ■

Putting the aforementioned analysis all together, the proposed algorithm can be summarized as in Algorithm 1.

Based on the proposed OrTC model, we can obtain the underlying low-rank traffic data $\hat{\mathcal{X}}$ and further detect the abnormal OD pairs at a given time slot according to anomaly tensor $\mathcal{A}$. Next, we discuss the time complexity of Algorithm 1. As shown in Algorithm 1, in the outer loop, the OrTC-AM algorithm first separates the anomaly and the noise from the sampled traffic data based on the tensor completion result of previous data; then, it imputes the missing traffic data by solving the tensor completion problem in the inner loop. At each iteration *k*, updating $\mathcal{A}^k$ and $\mathcal{N}^k$ both spend $\mathcal{O}(I_1I_2I_3)$. In the inner loop, we iteratively solve the tensor completion problem by employing the ALS method. At each iteration *n*, updating the $1^{st}$ way factor matrix $\mathbf{U}^n$ needs $\mathcal{O}(R^2I_1I_2I_3 + R^3I_1)$, and likewise for $\mathbf{V}^n$ and $\mathbf{T}^n$. Hence, solving the problem in (28) at each iteration *k* needs $\mathcal{O}(t_{ALS}(R^2I_1I_2I_3 + R^3(I_1 + I_2 + I_3)))$, where $t_{ALS}$ is the iteration count of the ALS method required to converge. Because of the low-rank characteristic of the underlying traffic data $\hat{\mathcal{X}}$, the rank *R* satisfies $R \ll \min\{I_1, I_2, I_3\}$. Therefore, the time complexity of Algorithm 1 is $\mathcal{O}(t_{AM}(t_{ALS}(R^2I_1I_2I_3))))$, where $t_{AM}$ denotes the iteration count that the outer loop takes for convergence. In addition, the space complexity of Algorithm 1 is $\mathcal{O}(I_1I_2I_3)$.

## VI. Performance Evaluation

In this section, we conduct extensive experiments to validate the effectiveness of our proposed OrTC-AM algorithm using the public Abilene dataset [38], which contains a time series of OD pairs traffic collected from the Internet2 backbone network.

TABLE II
COMPARISONS OF THE COMPETING METHODS

| Method | Missing Data Completion | Anomaly Detection | Gaussian-noise-tolerance |
|---|---|---|---|
| DSTC [14] | minimize the sum of Tucker rank | / | / |
| Online-SGD [19] | CP-based tensor completion | / | $L_F$-norm |
| TRPCA [29] | / | $L_1$-norm | / |
| TensorDet [15] | / | $L_0$-norm | / |
| BRCP-TC [35] | CP-based tensor completion | $L_1$-norm | $L_F$-norm |
| OrTC-AM | CP-based tensor completion | $L_{2,1}$-norm | $L_F$-norm |

**Algorithm 1** OrTC-AM Algorithm

**Input:** The collected traffic data tensor $\mathcal{X}$, the set consisting of the indices of sampled entries $\Omega$, the parameters $\lambda$, $\mu$, and a small threshold $\epsilon$.

**Output:** The optimal objective traffic data tensor $\hat{\mathcal{X}}$, and the structural anomaly tensor $\mathcal{A}$.

Initialize $\mathcal{X}^0 = \mathcal{X}$, $\hat{\mathcal{X}}^0 = 0$ $\mathcal{N}^0 = 0$, and $\mathcal{A}^0 = 0$;
$k = 0$;
**repeat**
  **Evaluate** $\mathcal{A}^{k+1}$**:**
  update $\mathcal{O} = \mathcal{X}^k - \hat{\mathcal{X}}^k - \mathcal{N}^k$;
  **for** $i_3 = 1, \dots, I_3$ **do**
    update each frontal slice $\mathbf{A}_{i_3}$ by (11).
  **end for**
  **Evaluate** $\mathcal{N}^{k+1}$**:**
  update $\mathcal{P} = \mathcal{X}^k - \hat{\mathcal{X}}^k - \mathcal{A}^{(k+1)}$;
  update $\mathcal{N}^{k+1}$ by (22).
  **Evaluate** $\hat{\mathcal{X}}^{k+1}$**:**
  update $\mathcal{Q}^0 = \mathcal{X}^k - \mathcal{A}^{k+1} - \mathcal{N}^{k+1}$;
  Randomly initialize $\mathbf{U}^0$ $\mathbf{V}^0$, and $\mathbf{T}^0$;
  $n = 0$;
  **repeat**
    **for** $i_1 = 1, \dots, I_1$ **do**
      update each row $\mathbf{u}_{i_1}^{(n+1)}$ by (24);
    **end for**
    **for** $i_2 = 1, \dots, I_2$ **do**
      update each row $\mathbf{v}_{i_2}^{n+1}$ by (26);
    **end for**
    **for** $i_3 = 1, \dots, I_3$ **do**
      update each row $\mathbf{t}_{i_3}^{n+1}$ by (27);
    **end for**
    update $\hat{\mathcal{X}}^{n+1} = [[\mathbf{U}^{n+1}, \mathbf{V}^{n+1}, \mathbf{T}^{n+1}]]$;
    update $\mathcal{Q}^{n+1}$ by (33);
    $n = n + 1$.
  **until** $\left\|\hat{\mathcal{X}}^{n+1} - \hat{\mathcal{X}}^n\right\|_F^2 / \left\|\hat{\mathcal{X}}^n\right\|_F^2 \leq \epsilon$
  update $\hat{\mathcal{X}}^{k+1} = \hat{\mathcal{X}}^{n+1}$.
  update $\mathcal{X}^{k+1} = \mathbf{P}_\Omega(\mathcal{X}) + \mathbf{P}_{\overline{\Omega}}\left(\mathcal{A}^{k+1} + \mathcal{N}^{k+1} + \hat{\mathcal{X}}^{k+1}\right)$;
  set $k = k + 1$.
**until** $\left\|\hat{\mathcal{X}}^{k+1} - \hat{\mathcal{X}}^k\right\|_F^2 / \left\|\hat{\mathcal{X}}^k\right\|_F^2 \leq \epsilon$

### A. Experimental Data

Considering that the collected raw traffic tensor may be the aggregation of clean traffic component and noise component, we employ the tensor BCPF algorithm [39] to discern the nominal "ground truth" traffic data $\mathcal{X}_T$ from the raw traffic data. Similar to reference [15], for more efficient data processing, data normalization is often applied in the data preprocessing step to scale the variables or features of data, and the normalized values are within the range [0,1] as

$$x_{i_1 i_2 i_3} = \frac{x_{i_1 i_2 i_3} - \min_{i_1, i_2, i_3}\{x_{i_1 i_2 i_3}\}}{\max_{i_1, i_2, i_3}\{x_{i_1 i_2 i_3}\} - \min_{i_1, i_2, i_3}\{x_{i_1 i_2 i_3}\}}. \quad (34)$$

Furthermore, to simulate the true network environments, we set up different noise scenarios as follows:

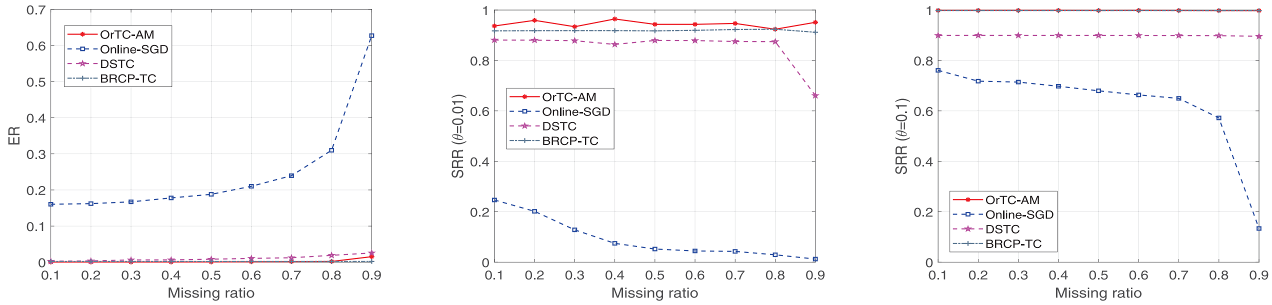
- Gaussian noise $\mathcal{N}_G$: 100% of the entries are contaminated with $\mathcal{N}(0, 0.01)$;
- Structural anomaly $\mathcal{A}_S$: the structural anomaly is set to be randomly generated within the range of [0, 5] and with a $3^{rd}$ slice-wise pollution ratio 1%;
- Random anomaly $\mathcal{A}_R$: 1% of the entries are corrupted with uniformly distributed noise between [0, 5].

Then, these noise and anomaly are injected into the "ground truth" traffic data $\mathcal{X}_T$. Accordingly, the incomplete traffic data $\mathbf{P}_\Omega(\mathcal{X})$ can be obtained by sampling uniformly at random from the aggregate tensor $\mathcal{X}_T + \mathcal{N}_G + \mathcal{A}_S(\mathcal{A}_R)$ with a missing ratio $p_m$.

### B. Experimental Setting

Our proposed OrTC-AM method involves 3 hyper-parameters $\lambda$, $\mu$, and the rank *R*. It is important to adjust these parameters to proper values. We use a 5-fold cross-validation procedure to determine the optimal values of parameter $\lambda$ and *R*. Specifically, we randomly partition the whole data from $\mathbf{P}_\Omega(\mathcal{X})$ into 5 roughly equivalent subsets, and then select each subset as the testing data and the remaining subsets as the training data. This process is independently repeated for 10 times to avoid the bias introduced by randomly partitioning the data. We use the following measurement to assess the reconstruction performance of the missing entries $\|\mathbf{P}_{\overline{\Omega}}(\mathcal{X}_T - \hat{\mathcal{X}})\|_F^2 / \|\mathbf{P}_{\overline{\Omega}}(\mathcal{X}_T)\|_F^2$. Finally, the value leading to the best performance is used to construct the optimal OrTC model and then applied to the performance evaluation for the data recovery and anomaly detection. In addition, similar to the implementation of the HaLTRC algorithm [28], we do not fix the value of parameter $\mu$, but set its initial value as $L_F$-norm of sampling tensor, and then makes it incrementally iterate at the rate of 1.25, which can help convergence.

We compare our proposed OrTC-AM algorithm with five different competing methods: 1) low-rank tensor completion method: DSTC algorithm [14]; 2) $L_F$-norm-based low-rank tensor completion method: Online-SGD algorithm [19]; 3) $L_1$-norm-based low-rank anomaly detection method: TRPCA algorithm [29]; 4) low-rank Tucker-based anomaly detection method: TensorDet algorithm [15]; 5) $L_F$-norm and $L_1$-norm-based low-rank tensor completion and anomaly detection method: BRCP-TC algorithm [35]. All the involved parameters in these competing methods were optimized by using the same 5-fold cross-validation procedure as in our proposed OrTC-AM algorithm. Table II summarizes the five competing methods and our proposed OrTC-AM method with the characteristics of missing data completion, anomaly detection, and Gaussian-noise-tolerant.




(a) ER (b) SRR ($\theta = 0.01$) (c) SRR ($\theta = 0.1$)

Fig. 3. Scenario 1: Traffic data without noise and anomaly.

### *C. Performance Metric*

The OrTC-AM algorithm aims to recover the true underlying traffic data as well as position the network anomaly. We denote the recovered low-rank traffic data as $\hat{\mathcal{X}}$, and the detected anomaly as $\mathcal{A}$. We use the following four metrics [14], [15] to evaluate the performance of the proposed OrTC-AM algorithm.

*Definition 6 (Error Ratio, ER):* A metric for measuring the recovery error of entries in the tensor after the interpolation, which can be calculated as

$$\mathrm{ER} = \left\| \left( \hat{\mathcal{X}} - \mathcal{X}_T \right) \right\|_F^2 / \|(\mathcal{X}_T)\|_F^2 . \tag{35}$$

*Definition 7 (Successful Recovery Ratio, SRR):* A metric for measuring the successful recovery of entries in the tensor after the interpolation, which can be calculated as:

$$\mathrm{SRR} = \sum_{i_1, i_2, i_3} \rho_{i_1 i_2 i_3} / (I_1 \times I_2 \times I_3),$$
$$\text{where} \quad \rho_{i_1 i_2 i_3} = \begin{cases} 1 & \text{if } \left| \frac{\hat{x}_{i_1 i_2 i_3} - x_{T\, i_1 i_2 i_3}}{x_{T\, i_1 i_2 i_3}} \right| \le \theta \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

where $\hat{x}_{i_1 i_2 i_3}$ and $x_{T\, i_1 i_2 i_3}$ denote the recovered data and the raw data at $(i_1, i_2, i_3)$-th element of the corresponding tensor, respectively. The parameter $\theta$ is the accuracy ratio of successful recovery.

*Definition 8 (True Positive Ratio, TPR):* A metric for measuring the proportion of anomaly that is correctly identified, which can be calculated as:

$$\mathrm{TPR} = \sum_{i_1, i_2, i_3} \alpha_{i_1 i_2 i_3} / (\mathrm{length}(\mathcal{A}_{\mathrm{T}}(:))),$$
$$\text{where} \quad \alpha_{i_1 i_2 i_3} = \begin{cases} 1 & \text{if } (i_1, i_2, i_3) \in \mathcal{A} \cap \mathcal{A}_T \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

where $\mathcal{A}_T$ represents the true anomaly, i.e., structural anomaly $\mathcal{A}_S$ or random anomaly $\mathcal{A}_R$. The function $\mathrm{length}(\cdot)$ returns the total number of true anomaly.

*Definition 9 (False Positive Ratio, FPR):* A metric for measuring the proportion of non-anomaly that is wrongly identified as the anomaly, which can be calculated as:

$$\mathrm{FPR} = \sum_{i_1, i_2, i_3} \beta_{i_1 i_2 i_3} / \left( \mathrm{length} \left( \overline{\mathcal{A}_{\mathrm{T}}}(:) \right) \right),$$
$$\text{where} \quad \beta_{i_1 i_2 i_3} = \begin{cases} 1 & \text{if } (i_1, i_2, i_3) \in \mathcal{A} \cap \overline{\mathcal{A}_T} \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

where the non-anomaly $\overline{\mathcal{A}_T}$ is the complement set of $\mathcal{A}_T$.

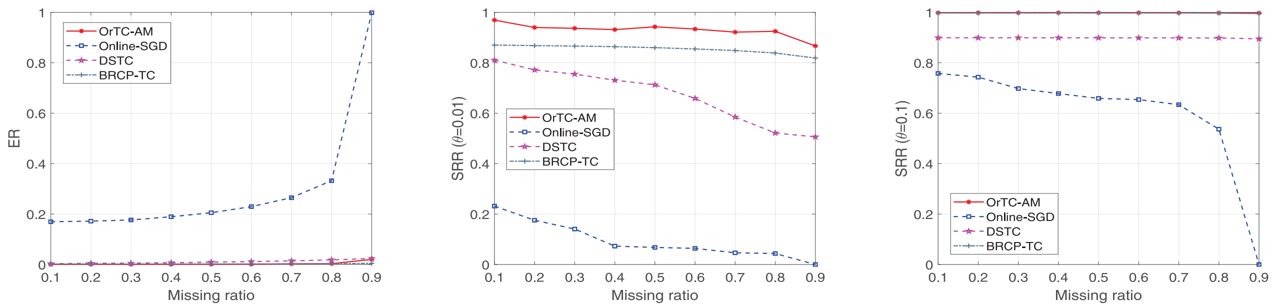
### *D. Experiment Results*

The objective of our proposed OrTC-AM algorithm is to estimate the intrinsical traffic data and detect the network anomaly simultaneously. Therefore, we compare the proposed algorithm with others from the following two aspects: traffic data recovery and anomaly detection.

*1) Traffic Data Recovery:* In order to investigate the traffic data recovery performance of our proposed algorithm, we compare OrTC-AM with three tensor completion methods, i.e., Online-SGD, DSTC, and BRCP-TC, under three different application scenarios.

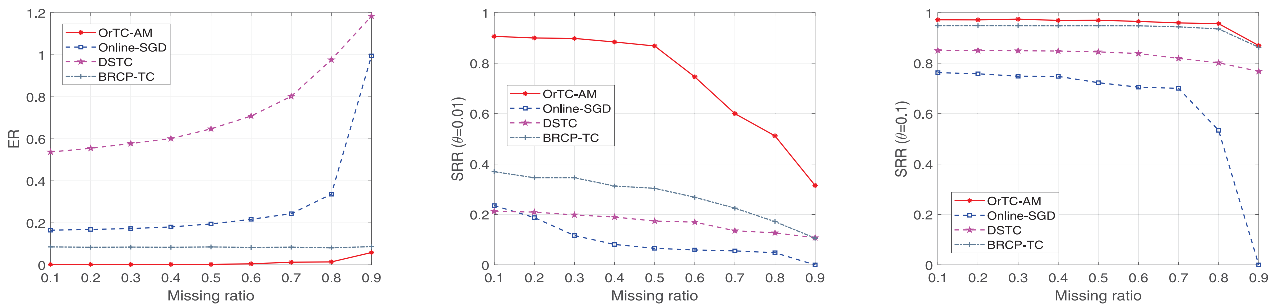
*Scenario 1 (Traffic Data Without Noise and Anomaly):* We assume that all collected traffic data is accurate in scenario 1. Fig. 3 shows the variation of the recovery error ratio and successful recovery ratio when the traffic data is sampled at different proportions. The horizontal axis of the three subgraphs represents the missing ratio, while the vertical axis in Fig. 3(a) represents ER, and the vertical axis in Fig. 3(b) and Fig. 3(c) represent SRR with $\theta = 0.01$ and $\theta = 0.1$, respectively. We can observe that ER is consistent with the missing ratio trend, and SRR generally varies inversely with the missing ratio. In this scenario, specifically, except that the BRCP-TC algorithm achieves slightly better ER values than our proposed OrTC-AM algorithm does when the missing ratio is more than 0.6, the OrTC-AM algorithm consistently outperforms all other competing methods (i.e., Online-SGD and DSTC). For the Online-SGD algorithm, it completes the missing data from the sequentially acquired traffic data, and thus its recovery accuracy is lower than the other completion methods with entire data. Furthermore, from Fig. 3(b) and Fig. 3(c), our proposed OrTC-AM algorithm obtains relatively better results compared with BRCP-TC. When we relax the accuracy ratio from 99% to 90%, these two algorithms can restore almost all data. This result indicates that the completion methods (i.e., OrTC-AM, DSTC, and BRCP-TC) can impute the missing data accurately under the ideal condition even when the missing ratio is high.

*Scenario 2 (Traffic Data Only With Gaussian Noise):* This set of experiments assume that the collected traffic data are corrupted by only Gaussian noise but not by network anomaly. After sampling the traffic data in different proportions, the experimental results are shown in Fig. 4. Similar to Fig. 3, OrTC-AM and BRCP-TC outperform all other competing methods from the aspect of ER and SRR. The SRR




(a) ER (b) SRR ($\theta = 0.01$) (c) SRR ($\theta = 0.1$)

Fig. 4. Scenario 2: Traffic data only with noise.




(a) ER (b) SRR ($\theta = 0.01$) (c) SRR ($\theta = 0.1$)

Fig. 5. Scenario 3: Traffic data with noise and structural anomaly.

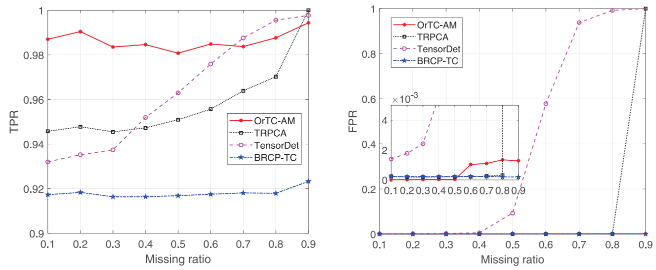
($\theta = 0.01$) performance of DSTC declines due to the impact of the dense Gaussian noise. This result demonstrates that the $L_F$-norm-based methods are Gaussian-noise-robust.

*Scenario 3 (Traffic Data With Gaussian Noise and Structural Anomaly):* This set of experiments assume that the collected traffic data are corrupted by both the Gaussian noise and network anomaly. The performance gaps among the four algorithms are illustrated in Fig. 5. We can observe that the proposed OrTC-AM method outperforms the other competing methods in both ER and SRR. More concretely, the ER of Online-SGD and DSTC is considerably high since they are not anomaly robustness. From Fig. 5(b), we observe that the reconstruction performance (99% accuracy ratio of each data) of all algorithms declines with the increase of missing ratio. Even though, OrTC-AM's SRR remains around 90% when the missing ratio is 50% and falls to 31.5% when there is only 10% sampled data. On the other hand, the successful recovery ratio of other algorithms is only 36.8% in the best case. Furthermore, when the accuracy ratio is relaxed to 90%, both of OrTC-AM and BRCP-TC can consistently maintain around 90% SRR when the missing ratio varies, which demonstrates that they can recover a large part of traffic data precisely. All in all, both of the $L_1$ and $L_{2,1}$ methods are robust to the network anomaly, and the $L_{2,1}$ method can result in a better performance for the structural anomaly.

*2) Anomaly Detection:* In order to investigate the anomaly detection performance of the proposed OrTC-AM method, we compare with three anomaly detection methods (i.e., TRPCA, TensorDet, and BRCP-TC) under two different application scenarios.

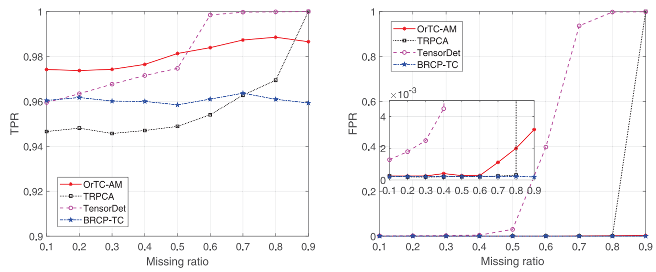
*Scenario 4 (Anomaly Detection With Gaussian Noise and Structural Anomaly):* This set of experiments assume that the collected traffic data are corrupted by both the Gaussian noise and the structural anomaly. Fig. 6 reports the TPR and FPR values of the TRPCA method, the TensorDet method, the BRCP-TC method, and our proposed OrTC-AM method when the traffic data is sampled at different proportions. Compared with TRPCA, TensorDet, and BRCP-TC, our proposed method achieves a lower FPR and a higher TPR when the missing ratio is less than 50%. Moreover, when the missing ratio reaches 0.5 or more, OrTC-AM can still maintain a high TPR (more than 98%, 6% more than BRCP-TC) and a tiny FPR (less than 0.2%, and just 0.1% more than BRCP-TC), demonstrating that OrTC-AM can accurately detect abnormal OD pairs at a given time slot even with a small sampling rate 10%. Comparatively, the performance of TensorDet and TRPCA declines significantly when the missing ratio is 0.9, e.g., the FPR rockets to 1, which means these two methods identifies all the data as outliers wrongly. These results prove that OrTC-AM is a robust anomaly detection algorithm that can fully utilize the structural information to detect the anomaly more accurately.

*Scenario 5 (Anomaly Detection With Gaussian Noise and Random Anomaly):* In this scenario, we compare four methods when the collected traffic data are corrupted by both the Gaussian noise and random anomaly, which means that the anomaly randomly distributes on all traffic data. Similar to scenario 4, TensorDet and TRPCA cannot detect the anomaly precisely when the missing ratio is high, and OrTC-AM consistently outperforms all other competing methods from the aspect of TPR, i.e., OrTC-AM can detect the true anomaly with more than 97.5% accuracy. Although BRCP-TC achieves



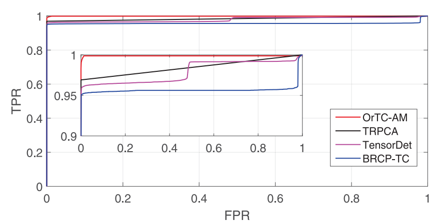

(a) True Positive Ratio (b) False Positive Ratio

Fig. 6. Scenario 4: Anomaly detection with noise and structural anomaly.

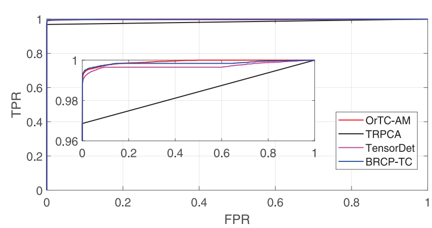



(a) True Positive Ratio (b) False Positive Ratio

Fig. 7. Scenario 5: Anomaly detection with noise and random anomaly.




(a) Structure Anomaly




(b) Random Anomaly

Fig. 8. ROC with structure anomaly and random anomaly.

slightly better FPR values than our proposed OrTC-AM method does when the missing ratio is more than 50%, OrTC-AM can continually keep the FPR less than 0.004 when the traffic data is sampled at different proportions. These results demonstrate that OrTC-AM is a robust anomaly detection technique that is not only applicable to the structural anomaly but also the random anomaly.

Furthermore, we use the Receiver Operating Characteristic (ROC) curve to evaluate the performance of anomaly detection in scenarios 4 and 5. The ROC curve plots the true positive rate against the false positive rate at various discrimination thresholds. In this case, we vary the anomaly thresholds by fixing the missing ratio as 0.5. From the ROC curves illustrated in Fig. 8, we can find that the OrTC-AM algorithm has a noticeable better detection performance than other algorithms in case of the structural anomaly. On the other hand, these four algorithms have similar performance to detect the random anomaly, and OrTC-AM is slightly better than other algorithms. Therefore, we can come to the same conclusion as Fig. 6 and Fig. 7, i.e., OrTC-AM is a robust anomaly detection technique, especially for the structural anomaly.

## VII. Conclusion

In this article, we aim to recover the full traffic data from partial traffic measurements. Considering the coexistence of potential network anomaly and noise in practical applications, we proposed a novel outlier-robust tensor completion (OrTC) model to accomplish this task. Based on the AM method, an effective algorithm OrTC-AM was designed to solve the non-convex and non-smooth OrTC model. Compared with existing network traffic estimation methods, the proposed OrTC-AM algorithm can not only complete missing traffic data and recover corrupted traffic data simultaneously, but also detect network anomaly accurately. Finally, the experimental results performed on the real dataset justified the advantages of our proposed OrTC-AM algorithm over the state-of-the-art competing methods. Future work will focus on extending our OrTC model to integrate more tensor completion techniques for improving the estimation performance on large-scale increasing network traffic data. Furthermore, to choose an optimal number of the rank $R$ is the key problem of CP tensor factorization. In this article, we utilized the cross-validation procedure to determine the optimal value, and we will explore automatic selection methods [39], [40] in further research.

## References


[1] Y. Zhang, M. Roughan, C. Lund, and D. L. Donoho, "Estimating point-to-point and point-to-multipoint traffic matrices: An information-theoretic approach," *IEEE/ACM Trans. Netw.*, vol. 13, no. 5, pp. 947–960, Oct. 2005.

[2] T. L. N. Nguyen and Y. Shin, "Matrix completion optimization for localization in wireless sensor networks for intelligent IoT," *Sensors*, vol. 16, no. 5, pp. 722–732, 2016.

[3] Y. Zhao, L. T. Yang, and J. Sun, "Privacy-preserving tensor-based multiple clusterings on cloud for industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2372–2381, Apr. 2019.

[4] G. Liang, N. Taft, and B. Yu, "A fast lightweight approach to origin-destination IP traffic estimation using partial measurements," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2634–2648, Jun. 2006.

[5] J. Jung, B. Krishnamurthy, and M. Rabinovich, "Flash crowds and denial of service attacks: Characterization and implications for CDNs and Web sites," in *Proc. 11th Int. Conf. World Wide Web (WWW)*, 2002, pp. 293–304.

[6] P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," in *Proc. 2nd ACM SIGCOMM Workshop Internet Meas. (IMW)*, 2002, pp. 71–82.

[7] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and Internet traffic matrices," in *Proc. ACM SIGCOMM Conf. Data Commun.*, 2009, pp. 267–278.

[8] Y.-C. Chen, L. Qiu, Y. Zhang, G. Xue, and Z. Hu, "Robust network compressive sensing," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2014, pp. 545–556.

[9] G. Gürsun and M. Crovella, "On traffic matrix completion in the Internet," in *Proc. Internet Meas. Conf. (IMC)*, 2012, pp. 399–412.

[10] A. Gunnar, M. Johansson, and T. Telkamp, "Traffic matrix estimation on a large IP backbone: A comparison on real data," in *Proc. 4th ACM SIGCOMM Conf. Internet Meas. (IMC)*, 2004, pp. 149–160.

[11] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Revision*, vol. 51, no. 3, pp. 455–500, 2009.

[12] H. Zhou, D. Zhang, K. Xie, and Y. Chen, "Spatio-temporal tensor completion for imputing missing Internet traffic data," in *Proc. IEEE 34th Int. Perform. Comput. Commun. Conf. (IPCCC)*, Nanjing, China, 2015, pp. 1–7.

[13] K. Lin, H. Zheng, X. Feng, and Z. Chen, "A novel spatial–temporal regularized tensor completion algorithm for traffic data imputation," in *Proc. 10th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Hangzhou, China, 2018, pp. 1–6.

[14] K. Xie *et al.*, "Accurate recovery of Internet traffic data: A sequential tensor completion approach," *IEEE/ACM Trans. Netw.*, vol. 26, no. 2, pp. 793–806, Apr. 2018.

[15] 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.

[16] H. Kasai, W. Kellerer, and M. Kleinsteuber, "Network volume anomaly detection and identification in large-scale networks based on online time-structured traffic tensor tracking," *IEEE Trans. Netw. Service Manag.*, vol. 13, no. 3, pp. 636–650, Sep. 2016.

[17] F. Xiao, L. Chen, H. Zhu, R. Hong, and R. Wang, "Anomaly-tolerant network traffic estimation via noise-immune temporal matrix completion model," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1192–1204, Jun. 2019.

[18] P. Xu, T. Cui, and L. Chen, "ANLoC: An anomaly-aware node localization algorithm for WSNs in complex environments," *Sensors*, vol. 19, no. 8, pp. 1912–1931, 2019.

[19] M. Mardani, G. Mateos, and G. B. Giannakis, "Subspace learning and imputation for streaming big data matrices and tensors," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2663–2677, May 2015.

[20] M. Hardt, "Understanding alternating minimization for matrix completion," in *Proc. IEEE 55th Annu. Symp. Found. Comput. Sci.*, Philadelphia, PA, USA, 2014, pp. 651–660.

[21] M. Roughan, Y. Zhang, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and Internet traffic matrices (extended version)," *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 662–676, Jun. 2012.

[22] M. Mardani, G. Mateos, and G. B. Giannakis, "Rank minimization for subspace tracking from incomplete data," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Vancouver, BC, Canada, 2013, pp. 5681–5685.

[23] K. Xie *et al.*, "On-line anomaly detection with high accuracy," *IEEE/ACM Trans. Netw.*, vol. 26, no. 3, pp. 1222–1235, Jun. 2018.

[24] M. Mardani and G. B. Giannakis, "Estimating traffic and anomaly maps via network tomography," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1533–1547, Jun. 2016.

[25] Y. Wu, H. Tan, Y. Li, J. Zhang, and X. Chen, "A fused CP factorization method for incomplete tensors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 751–764, Mar. 2019.

[26] B. D. Haeffele, and R. Vidal, "Global optimality in tensor factorization, deep learning, and beyond," 2015. [Online]. Available: https://arxiv.org/abs/1506.07540.

[27] Q. Zhao, C. Chen, R. Du, S. Bi, and B. Yang, "HaTTC: An urban traffic sensing method based on tensor completion technique," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, USA, 2014, pp. 175–180.

[28] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[29] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Apr. 2020.

[30] T. Yokota and H. Hontani, "Simultaneous visual data completion and denoising based on tensor rank and total variation minimization and its primal–dual splitting algorithm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 3843–3851.

[31] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an explanatory multi-modal factor analysis," in *Proc. UCLA Workshop*, vol. 16, 1970, pp. 1–84.

[32] L. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[33] H. Tan *et al.*, "A tensor-based method for missing traffic data completion," *Transp. Res. C*, vol. 28, pp. 15–27, Mar. 2013.

[34] T. Yokota and H. Hontani, "Simultaneous tensor completion and denoising by noise inequality constrained convex optimization," *IEEE Access*, vol. 7, pp. 15669–15682, 2019.

[35] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S. Amari, "Bayesian robust tensor factorization for incomplete multiway data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 736–748, Apr. 2016.

[36] C. Cuvelier, A. Segal, and A. Steenhoven, "The penalty function method," in *Finite Element Methods and Navier-Stokes Equations*. Dordrecht, The Netherlands: Kluwer, Jan. 1986, pp. 263–297.

[37] B. W. Bader *et al.* (Feb. 2015). *MATLAB Tensor Toolbox Version 2.6*. [Online]. Available: https://www.sandia.gov/tgkolda/TensorToolbox/index-2.6.html

[38] *The Abilene Observatory Data Collections*. Accessed: Jul. 20, 2004. [Online]. Available: http://abilene.internet2.edu/observatory/datacollections.html

[39] Q. Zhao, L. Zhang, and A. Cichocki, "Bayesian CP factorization of incomplete tensors with automatic rank determination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1751–1763, Sep. 2015.

[40] T. Yokota, Q. Zhao, and A. Cichocki, "Smooth PARAFAC decomposition for tensor completion," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5423–5436, Oct. 2016.

**Qianqian Wang** received the B.S. degree in computer science and technology and the M.S. degree in computer application technology from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2004 and 2007, respectively. She is pursuing the dual Ph.D. degree with Western University and the Nanjing University of Posts and Telecommunications.

She is currently a Lecturer with the School of Software Engineering, Jinling Institute of Technology, Nanjing. Her research interests include Internet of Things, traffic data analysis, and tensor completion.

**Lei Chen** (Member, IEEE) received the M.S. degree in computer software and theory and the Ph.D. degree in communication and information system from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2005 and 2014, respectively. He is currently a Professor with the School of Computer Science, Nanjing University of Posts and Telecommunications. He was a Visiting Researcher with the University of North Carolina, Chapel Hill, NC, USA, from June 2016 to June 2017. His research interests include machine learning, pattern recognition, and computer network.

**Qin Wang** (Member, IEEE) received the Ph.D. degree in communication and information system from the Nanjing University of Posts and Telecommunications (NJUPT), China, in 2016, where she is an Associate Professor. Prior to joining NJUPT, she was with the College of Engineering and Computing Sciences, New York Institute of Technology from February 2017 and August 2020. From July 2018 to June 2020, she was a Postdoctoral Research Fellow with the Electronic Science and Technology, NJUPT. From 2015 to 2016, she was a Visiting Scholar with the Department of Computer Science, San Diego State University, USA. Her research interests include multimedia communications, multimedia pricing, resource allocation in 5G, and Internet of Things. She serves as a Co-Editor of *IoT Journal* (Elsevier) and *International Journal of Grid and Utility Computing*, a Track Co-Chair of MIPR 2019, a TPC co-chair for many conferences, such as ICC 2018, ICNC 2018, MIPR 2019, and GLOBECOM 2019.

**Hongbo Zhu** (Member, IEEE) received the B.S. degree in communications engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982, and the Ph.D. degree in information and communications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 1996.

He is currently a Professor with the Nanjing University of Posts and Telecommunications. He is also the Head of the Coordination Innovative Center of Internet of Things (IoT) Technology and Application, which is the first governmental authorized Coordination Innovative Center of IoT, China. He has authored and coauthored over 200 technical papers published in various journals and conferences. He is also leading a big group and multiple funds on IoT and wireless communications with current focus on architecture and enabling technologies for IoT. His research interests include mobile communications, wireless communication theory, and electromagnetic compatibility. He also serves as a referee or an expert in multiple national organizations and committees.

**Xianbin Wang** received the Ph.D. degree in electrical and computer engineering from the National University of Singapore in 2001.

He is a Professor and a Tier 1 Canada Research Chair with Western University, Canada. Prior to joining Western, he was with Communications Research Centre Canada (CRC) as a Research Scientist/Senior Research Scientist from July 2002 to December 2007. From January 2001 to July 2002, he was a System Designer with STMicroelectronics. His current research interests include 5G and beyond, Internet-of-Things, communications security, machine learning, and intelligent communications. He has over 400 peer-reviewed journal and conference papers, in addition to 30 granted and pending patents and several standard contributions.

Prof. Wang has received many awards and recognitions, including the Canada Research Chair, CRC President's Excellence Award, the Canadian Federal Government Public Service Award, the Ontario Early Researcher Award and six IEEE Best Paper Awards. He was involved in many IEEE conferences, including GLOBECOM, ICC, VTC, PIMRC, WCNC, and CWIT, in different roles, such as a symposium chair, a tutorial instructor, a track chair, a session chair, and a TPC co-chair. He is currently serving as the Vice Chair of IEEE London Section and the Chair of ComSoc Signal Processing and Computing for Communications Technical Committee. He currently serves as an Editor/Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON BROADCASTING, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was also an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2011, and IEEE WIRELESS COMMUNICATIONS LETTERS from 2011 to 2016. He is a Fellow of Canadian Academy of Engineering, Engineering Institute of Canada, and an IEEE Distinguished Lecturer.