A DDoS Attack Detection by Group-Sparse Plus Low-Rank Temporally-Smooth Decomposition for Extended OD Flow Matrix

Masao Yamagishi, Masaya Endo, and Isao Yamada
Tokyo Institute of Technology, Japan
E-mail: {myamagi, endo, isao}@sp.ce.titech.ac.jp

***Abstract*—We propose to use group-sparsities of the extended OD (Origin-Destination) flow matrix, expressing time transitions of traffic matrices, for effective detection and flexible characterization of diverse network traffic anomalies. We also formulate an approximate decomposition problem of an extended OD flow matrix into a group-sparse and a low-rank temporally-smooth matrix. Moreover, we present a characterization of typical DDoS (Distributed Denial of Service) attacks with a certain sparsity of aggregated traffic volumes per unit time for each group of OD flows to the same destination. The proposed characterization is combined with the proposed decomposition problem to establish a DDoS attack detection scheme. A numerical experiment demonstrates effectiveness of the proposed DDoS attack detection scheme.**

I. Introduction

Cyber attacks including DDoS (Distributed Denial of Service) attack are immense threats of the internet and detecting their anomalous traffic is a crucial task to maintain safe and secure societies. Although this detection problem has been investigated in a variety of ways, most existing studies need packet details or introduction of a new protocol [1], which may cause disadvantages: collecting packet details not only degrades the network performance but also causes privacy issues; introducing a new protocol needs much cost.

On the other hand, anomaly detection methods based on information gathering time-transition of traffic volume per *OD flow* over the network are free from these disadvantages. The OD flow f_{ij} is a set consisting of all traffics which enter the network at the origin node i, and exit the network at the destination node j $(i \neq j)$.[1] To represent the traffic volume $z_{ij}^{(k)}$ of the OD flow f_{ij} in certain time $k(k = 1, \cdots, T)$, we utilize *the traffic matrix* $\boldsymbol{Z}^{(k)} := [z_{ij}^{(k)}]$ as used often in the field of network tomography [2]. The time transition of the traffic matrix can be expressed as the matrix $\boldsymbol{Z}_\star := [\ \mathrm{vec}(\boldsymbol{Z}^{(1)}) \cdots \mathrm{vec}(\boldsymbol{Z}^{(\mathrm{T})})\ ] \in \mathbb{R}^{\mathrm{F}\times\mathrm{T}}$ which is called as *the extended OD flow matrix* in this paper, where the operator $\mathrm{vec}(\cdot)$ converts a matrix to a column vector expression. The traffic matrix and extended OD flow matrix are widely used for traffic-engineering [3]–[7]. This is because, unlike the OD flow f_{ij}, they can be inferred through a certain linear model[2] and contain no packet detail. This nature makes anomaly detection methods based on the extended OD flow matrix free from the above disadvantages.

Another benefit to use of the extended OD flow matrix is that its detailed features were known empirically through analyses of actual measurements of the internet. [10]–[12] reported that the extended OD flow matrix $\boldsymbol{Z}_\star$ can be modeled well as *the sum of a smooth diurnal part* $\boldsymbol{Z}_\star^{\mathrm{sd}} \in \mathbb{R}^{F\times T}$ *and a fluctuation part* $\boldsymbol{Z}_\star^{\mathrm{flu}} \in \mathbb{R}^{F\times T}$, i.e.,

$$\boldsymbol{Z}_\star = \boldsymbol{Z}_\star^{\mathrm{sd}} + \boldsymbol{Z}_\star^{\mathrm{flu}}. \tag{1}$$

Meanwhile, Lakhina et al. [13] reported that the extended OD flow matrix $\boldsymbol{Z}_\star$ can be modeled well as the sum of a low-rank matrix $\boldsymbol{H}_\star \in \mathbb{R}^{F\times T}$ consisting of non-anomalous traffic volume, a matrix $\boldsymbol{A}_\star \in \mathbb{R}^{F\times T}$ consisting of anomalous traffic volumes, and a matrix $\boldsymbol{N}_\star \in \mathbb{R}^{F\times T}$ consists of noise-like fluctuation traffic volumes, i.e.,

$$\boldsymbol{Z}_\star = \boldsymbol{H}_\star + \boldsymbol{A}_\star + \boldsymbol{N}_\star. \tag{2}$$

We refer, in this paper, to $\boldsymbol{H}_\star$, $\boldsymbol{A}_\star$, and $\boldsymbol{N}_\star$, respectively, as normal component, anomalous component, and noise component.

The model in (2) is adopted widely in the network tomography [7], [14]–[16] and plays a key role in anomaly detection. Focusing on the low-rankness of $\boldsymbol{H}_\star$, [17] proposed an anomaly detection by Principal Component Analysis (PCA) [18]. Using the low-rankness of $\boldsymbol{H}_\star$ and the sparsity of $\boldsymbol{A}_\star$ simultaneously, [19] proposed an anomaly detection by the so-called Robust Principal Component Analysis (RPCA) [20]–[22] which is an approximate decomposition from a matrix to the sum of a sparse matrix and a low-rank matrix (see Sec. II-C). Although [19] assumed complete information of $\boldsymbol{Z}_\star$, [6] presented RPCA-based anomaly detection from superpositions of $\boldsymbol{Z}_\star$ (see Sec. II-D).

However, the existing RPCA-based detection methods [6], [19] are not necessarily appropriate to detect malicious traffic volumes caused by DDoS attacks. RPCA-based anomaly

[1] Denote the number of OD flows by F.

[2] The linear model (5) shows that entries of extended OD flow matrix are superposition of the traffic volume carried over links (link matrix) and routing information (routing matrix). The link matrix and routing matrix are available in practice by Simple Management Protocol (SNMP) [8] and Open Shortest Path First (OSPF) protocol [9].

detections adopt a principle to uniformly characterize non-anomalous traffic volumes by the low-rankness as well as diverse anomalous ones by the sparsity. In other words, RPCA-based anomaly detections do not care sufficiently about the causes of diverse anomalies. For example, in the typical DDoS attack, a large number of unspecified nodes send numerous packets to target nodes, which causes *group-sparse* malicious traffic volumes in the extended OD flow matrix rather than sparse ones (Fig. 3(a) below). Moreover, if the DDoS attack occurs in a periodic fashion (we call this kind of attacks as *periodic DDoS attack*), it may cause low-rank (but possibly nonsmooth) malicious traffic volumes indistinguishable from the normal component of low-rank. Hence the existing RPCA-based detection methods may fail to detect group-sparse or low-rank traffic volumes caused by the DDoS attacks.

In this paper, to resolve the above inappropriateness in use of the RPCA-based anomaly detections for DDoS attacks, we propose a use of group-sparsity of $\boldsymbol{A}_\star$ and temporally smoothness of $\boldsymbol{H}_\star$. Use of the group-sparsity in place of the sparsity realizes a flexible characterization of anomalous traffic volumes caused by cyber attacks including the typical DDoS attack. Use of the temporally smoothness realizes a more accurate characterization of the non-anomalous traffic volume and is helpful to distinguish the non-anomalous traffic volume from the anomalous one caused by the periodic DDoS attack. The use of the temporally smoothness is supported by two aspects: $\boldsymbol{H}_\star$ is expected to be temporally smooth by a simple comparison of the two models (2) and (1), i.e., $(\boldsymbol{H}_\star, \boldsymbol{A}_\star + \boldsymbol{N}_\star) \approx (\boldsymbol{Z}_\star^{\mathrm{sd}}, \boldsymbol{Z}_\star^{\mathrm{flu}})$; the periodic DDoS attack tends to cause non-smooth malicious traffic volumes. To embody the use of the group-sparsity and the temporally smoothness, we propose an approximate decomposition problem of an extended OD flow matrix into a group-sparse matrix plus a low-rank temporal-smooth matrix (see (20)) and show its solution can be efficiently approximated well by convex problem solvers (see Sec. III-A). We also present that the anomalous component due to the DDoS attacks is characterized as group-sparsity of OD flows grouped according to the destinations. This grouping is employed in the proposed decomposition problem to present a DDoS attack detection (see Sec. III-B). Finally, numerical experiment evaluates the utility of proposed DDoS attack detection (see Sec. IV).

II. Preliminaries

A. Notation

Denote respectively as $\mathbb{R}, \mathbb{R}_+$, and $\mathbb{N}$ the set of all real numbers, nonnegative real numbers, and natural numbers. Bold letters express a vector and a matrix. Denote the identity matrix by $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$. We denote the transpose of a vector or matrix by $(\cdot)^\top$, the rank of a matrix by $\mathrm{rank}(\cdot)$, and the trace of a square matrix by $\mathrm{tr}(\cdot)$. For $\boldsymbol{x} := (x_1, x_2 \ldots, x_n)^\top \in \mathbb{R}^n$, ℓ_p-norm ($p \geq 1$) is defined by $\|\boldsymbol{x}\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$. For $\boldsymbol{X} := [x_{i,j}] \in \mathbb{R}^{m\times n}$, define $\|\boldsymbol{X}_0\| := |\{(i,j) \mid x_{i,j} \neq 0\}|$ (ℓ_0 pseudo-norm), $\|\boldsymbol{X}\|_1 := \sum_{i,j} |x_{i,j}|$ (ℓ_1-norm), $\|\boldsymbol{X}\|_F := \sqrt{\mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^\top)}$ (Frobenius norm), and $\|\boldsymbol{X}\|_* := \sum_i \sigma_i(\boldsymbol{X})$ (Nuclear norm), where $\sigma_i(\boldsymbol{X})$ denotes the i-th largest singular value of $\boldsymbol{X}$.

B. Observation Model of Extended OD Matrix

We model IP network by a directed graph $G(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ and $\mathcal{L}$ denote, respectively, a set of nodes and a set of edges. Each node $n \in \mathcal{N}$ represents a network element that generates, receives and/or relays network traffic, e.g., an end-host, an Ethernet switch, an IP router, or a point of presence. Each edge $l \in \mathcal{L}$ represents a physical link between two nodes. For simple notation, we denote $N = |\mathcal{N}|$ and $L = |\mathcal{L}|$. We model the set of traffics traverse from one node to another by an OD (Origin-Destination) flow. Let $\mathcal{F}$ be the set of the OD flows for every different OD pairs. The number of OD flows, say $F := |\mathcal{F}| = N(N-1)$, far exceeds the number of physical links, i.e., $F \gg L$. Per OD-flow, single path routing is considered where in each flow packets traverse a unique path to reach the destination. Then the routing matrix $\boldsymbol{R} := [r_{l,f}] \in \{0,1\}^{L\times F}$ is defined by

$$r_{l,f} := \begin{cases} 1 & \text{if flow } f \text{ traverses link } l, \\ 0 & \text{otherwise,} \end{cases} \quad (l,f) \in \mathcal{L}\times\mathcal{F}. \tag{3}$$

We suppose that $\boldsymbol{R}$ is fixed and given. Let $z_{f,t}^\star$ denote the unknown integrated traffic volume of flow $f \in \mathcal{F}$ through time $[t, t+1)$. Similarly, let $y_{l,t}$ denote the integrated traffic volume of link $l \in \mathcal{L}$ through time $[t, t+1)$. Since the link traffic volumes arise from the superposition of the flow traffic volumes, $y_{l,t}$ can be represented by

$$y_{l,t} = \sum_{f\in\mathcal{F}} r_{l,f} z_{f,t}^\star + \tilde{\varepsilon}_{l,t}, \ t \in \mathcal{T} := \{1, \ldots, T\}, \tag{4}$$

where the additive noise $\tilde{\varepsilon}_{l,t} \in \mathbb{R}$ models packet loss on link $l \in \mathcal{L}$ and observation errors. By introducing the link matrix $\boldsymbol{Y} := [y_{l,k}] \in \mathbb{R}_+^{L\times T}$, the extended OD flow matrix $\boldsymbol{Z}_\star := [z_{f,t}^\star] \in \mathbb{R}_+^{F\times T}$, and the additive noise matrix $\tilde{\boldsymbol{\varepsilon}} := [\tilde{\varepsilon}_{l,t}] \in \mathbb{R}^{L\times T}$, we restate (4) as

$$\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Z}_\star + \tilde{\boldsymbol{\varepsilon}}. \tag{5}$$

We refer to (5) as the observation model of the extended OD flow matrix. In addition, combining (5) and Lakhina's model (2) yields the observation model

$$\boldsymbol{Y} = \boldsymbol{R}(\boldsymbol{H}_\star + \boldsymbol{A}_\star) + \boldsymbol{\varepsilon} \tag{6}$$

of the sum of the normal component and the anomalous component, where $\boldsymbol{\varepsilon} := \boldsymbol{R}\boldsymbol{N}_\star + \tilde{\boldsymbol{\varepsilon}}$. One of the main goals of anomaly detection is to estimate $(\boldsymbol{H}_\star, \boldsymbol{A}_\star)$ from the link matrix $\boldsymbol{Y}$ and the routing matrix $\boldsymbol{R}$. Note that the link matrix and the routing matrix are often available in practical situations[2] while entire information of all OD flows is hard to obtain.

Remark 1 (On use of partial flow traffic volumes): *In the real IP network, we can observe traffic volume of partial OD flows.*[3] *These observable OD flow traffic volumes can be*

[3]Routers that support Netflow [23] obtain flow traffic volumes.

utilized by replacing the observation models (5) *and* (6) *with*

$$\boldsymbol{Y}' = \boldsymbol{R}'\boldsymbol{Z}_\star + \tilde{\varepsilon}', \tag{7}$$

$$\boldsymbol{Y}' = \boldsymbol{R}'(\boldsymbol{H}_\star + \boldsymbol{A}_\star) + \boldsymbol{\varepsilon}', \tag{8}$$

where the routing matrix $\boldsymbol{R}$ *is replaced by the model matrix*

$$\boldsymbol{R}' := \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{I}^R \end{bmatrix} \in \{0,1\}^{(L+F_O)\times F} \tag{9}$$

(F_O is the number of the observable OD flows) with a diagonal matrix $\boldsymbol{I}^R \in \{0,1\}^{F_O\times F}$ *to extract traffic volumes of the observable OD flows, and, accordingly, the observation* $\boldsymbol{Y}$ *and the additive noises* $\tilde{\varepsilon}, \varepsilon$ *are also replaced with* $\boldsymbol{Y}'$*,* $\tilde{\varepsilon}'$*, and* ε'*. Since both models* (6) *and* (8) *are of the same form of the linear model, discussions in Sec. II-C, Sec. II-D, and Sec. III can be applied to both models while we focus mainly on the model* (6) *for simplicity.*

C. Robust Principal Component Analysis

The classical PCA problem [18] fits a low-rank matrix $\boldsymbol{L} \in \mathbb{R}^{m_1\times m_2}$ to a given high-dimensional data matrix $\boldsymbol{M} \in \mathbb{R}^{m_1\times m_2}$ while trying to minimize the norm of the error $\boldsymbol{M} - \boldsymbol{L} =: \boldsymbol{S} \in \mathbb{R}^{m_1\times m_2}$ for arbitrarily fixed $r \in \mathbb{N}$:

$$\underset{(\boldsymbol{L},\boldsymbol{S})}{\text{minimize}}\ \|\boldsymbol{S}\|_{\mathrm{F}},\ \text{s.t. } \operatorname{rank}(\boldsymbol{L}) \le r,\ \boldsymbol{M} = \boldsymbol{L} + \boldsymbol{S}. \tag{10}$$

By the Schmidt-Eckart-Young Theorem [24], the optimal low-rank matrix $\widehat{\boldsymbol{L}} \in \mathbb{R}^{m_1\times m_2}$ in the minimizer $(\widehat{\boldsymbol{L}}, \boldsymbol{M} - \widehat{\boldsymbol{L}}) \in \mathbb{R}^{m_1\times m_2}\times\mathbb{R}^{m_1\times m_2}$ of problem (10) is given by truncating the singular value decomposition (SVD) of $\boldsymbol{M}$, retaining only the contribution of its r largest singular values.[4] If $\boldsymbol{M} = \boldsymbol{L}_0 + \boldsymbol{S}_0$ originates from a low-rank matrix $\boldsymbol{L}_0$ ($\operatorname{rank}(\boldsymbol{L}_0) \le r$) and $\boldsymbol{S}_0$ contains Gaussian noise (entries of $\boldsymbol{S}_0$ distributed as $\mathcal{N}(0,\sigma^2)$ with σ^2 small), then the truncated SVD will recover a matrix $\widehat{\boldsymbol{L}} \approx \boldsymbol{L}_0$. However, if $\boldsymbol{S}_0$ contains very large entries (outliers) then the truncated SVD will return a matrix that is largely deviating from $\boldsymbol{L}_0$, even if $\boldsymbol{S}_0$ only affects a small fraction of the entries of $\boldsymbol{L}_0$. In other words, if $\boldsymbol{S}_0$ models a sparse error, i.e., it represents a small number of largely corrupted entries of $\boldsymbol{L}_0$, the truncated SVD will fail to recover $\boldsymbol{L}_0$ (in most cases).

On the other hand, [20], [21], and [22] showed independently from each other that under certain assumptions (on $\boldsymbol{L}_0$ and $\boldsymbol{S}_0$) one can exactly recover the low-rank matrix $\boldsymbol{L}_0$ from $\boldsymbol{M} = \boldsymbol{L}_0 + \boldsymbol{S}_0$, where $\boldsymbol{S}_0$ is a sparse matrix, by solving the following convex optimization problem:

$$\underset{(\boldsymbol{L},\boldsymbol{S})}{\text{minimize}}\ \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{S}\|_1,\ \text{s.t. } \boldsymbol{M} = \boldsymbol{L} + \boldsymbol{S}, \tag{11}$$

where $\lambda \in \mathbb{R}$ is a trade-off parameter between the rank of $\boldsymbol{L}$ and the sparsity of $\boldsymbol{S}$. We will refer to problem (11) as Principal Component Pursuit (PCP) or as Sparse Plus Low-Rank decomposition. Problem (11) is a convex relaxation of

$$\underset{(\boldsymbol{L},\boldsymbol{S})}{\text{minimize}}\ \operatorname{rank}(\boldsymbol{L}) + \lambda\|\boldsymbol{S}\|_0,\ \text{s.t. } \boldsymbol{M} = \boldsymbol{L} + \boldsymbol{S}. \tag{12}$$

D. Network Anomaly Detection via PCP

Under the assumption that $\boldsymbol{H}_\star$ is low-rank and $\boldsymbol{A}_\star$ is sparse, by following the idea of the PCP (11), the authors in [7] presented to estimate $(\boldsymbol{H}_\star, \boldsymbol{A}_\star)$ from the observation $\boldsymbol{Y}$ in (6) by a solution $(\widehat{\boldsymbol{H}}, \widehat{\boldsymbol{A}})$ of *the PCP for anomaly detection*:

$$\underset{(\boldsymbol{H},\boldsymbol{A})}{\text{minimize}}\ \|\boldsymbol{H}\|_* + \lambda\|\boldsymbol{A}\|_1 \text{ s.t. } \boldsymbol{Y} = \boldsymbol{R}(\boldsymbol{H} + \boldsymbol{A}). \tag{13}$$

Since $\|\cdot\|_*$ is a convex relaxation of $\operatorname{rank}(\cdot)$, minimization of the objective function in problem (13) enhances the low-rankness of $\widehat{\boldsymbol{H}}$, which is consistent with the low-rankness of $\boldsymbol{H}_\star$.

However, problem (13) adopts a principle to uniformly characterize anomalous traffic volumes by the sparsity and does not care sufficiently the causes of diverse anomalies. Hence, the anomaly detection by $\widehat{\boldsymbol{A}}$ cannot be expected its high accuracy in a wide variety of circumstances. We will see that the anomaly detection by $\widehat{\boldsymbol{A}}$ fails to identify the malicious flow traffic volumes due to the DDoS attacks in Figure 3(b), which demonstrates that the PCP for anomaly detection is not a general-purpose tool.

III. A GENERALIZED PCP: GROUP-SPARSE PLUS LOW-RANK TEMPORALLY-SMOOTH DECOMPOSITION FOR ANOMALY DETECTION

A. Generalized PCP for Anomaly Detection

First, we design a cost function of which suppression enhances the group-sparsity of the estimation of $\boldsymbol{A}_\star$. Let $\mathcal{F}_j \subset \mathcal{F}$ ($j = 1,\dots,J$) be nonempty and let $\mathcal{T}_k \subset \mathcal{T}$ ($k = 1,\dots,K$). Then $\mathcal{F}_j \times \mathcal{T}_k$ forms a group w.r.t. OD flows and times. We introduce $\mathfrak{G} := \{\mathcal{F}_j \times \mathcal{T}_k \mid j = 1,\dots,J, k = 1,\dots,K\}$ to characterize anomalous traffic volumes. (note: $\mathfrak{G}$ forms possibly overlapping groups). For $\mathcal{G} \in \mathfrak{G}$, define an operator

$$\pi_{\mathcal{G}} \colon \mathbb{R}^{F\times T} \to \mathbb{R}^{|\mathcal{G}|} \colon \boldsymbol{X}(:= [x_{f,t}]) \mapsto (x_{i,j})_{(i,j)\in\mathcal{G}} \tag{14}$$

to extract entries of an extended OD flow matrix belonging to the group $\mathcal{G}$. Then suppressing

$$\mathbb{R}^{F\times T} \to \mathbb{R}_+ \colon \boldsymbol{A} \to \sum_{\mathcal{G}\in\mathfrak{G}} \|\pi_{\mathcal{G}}(\boldsymbol{A})\|, \tag{15}$$

enhances the sparseness of the vector $(\|\pi_{\mathcal{G}}(\boldsymbol{A})\|)_{\mathcal{G}\in\mathfrak{G}} \in \mathbb{R}_+^{|\mathfrak{G}|}$ because the function in (15) is nothing but the ℓ_1-norm of the vector $(\|\pi_{\mathcal{G}}(\boldsymbol{A})\|)_{\mathcal{G}\in\mathfrak{G}}$. In other words, suppressing the function in (15) enhances the group-sparsity of $\boldsymbol{A}$ with the grouping $\mathfrak{G}$.

Second, we design a cost function of which suppression enhances the temporal smoothness of the estimation of $\boldsymbol{H}_\star$. Our cost function is nothing but the high-order total-variation

[4] The optimal low-rank matrix is obtained as

$$\widehat{\boldsymbol{L}} = \boldsymbol{U}\operatorname{diag}(\sigma_1,\dots,\sigma_r,0,0,0)\boldsymbol{V}^\top,$$

where we denote that the SVD of $\boldsymbol{M}$ by $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ with orthonormal matrices $\boldsymbol{U} \in \mathbb{R}^{m_1\times m_1}$, $\boldsymbol{V} \in \mathbb{R}^{m_2\times m_2}$ and with the diagonal matrix $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1,\sigma_2,\dots,\sigma_{\operatorname{rank}(\boldsymbol{M})},0,\dots,0)$ consisting of singular values of $\boldsymbol{M}$ in nonincreasing order $\sigma_1 \ge \sigma_2 \ge \dots \ge \sigma_{\operatorname{rank}(\boldsymbol{M})} > 0$.

[25] where the variation is computed in time direction. Define $\boldsymbol{D}_\tau = [(d_\tau)_{i,j}] \in \{-1,0,1\}^{\tau\times(\tau-1)}$

$$(d_\tau)_{i,j} := \begin{cases} -1 & \text{if } i = j \\ 1 & \text{if } i+1 = j \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

($\forall \tau \in \mathbb{N}\colon \tau \geq 2$).[5] Then the matrix $\boldsymbol{Z}\boldsymbol{D}_T \in \mathbb{R}^{F\times(T-1)}$ comprises the differences of temporally adjacent entries of $\boldsymbol{Z} \in \mathbb{R}^{F\times T}$. Furthermore, we recursively define the matrix for computing the k-th ($k \in \mathbb{N}$) order variations

$$\boldsymbol{D}_T^{(1)} := \boldsymbol{D}_T \in \{-1,0,1\}^{T\times(T-1)} \tag{17}$$

$$\boldsymbol{D}_T^{(k)} := \boldsymbol{D}_T^{(k-1)}\boldsymbol{D}_{T-(k-1)} \in \mathbb{R}^{T\times(T-k)}. \tag{18}$$

Then suppressing the k-th order total variation

$$\mathbb{R}^{F\times T} \to \mathbb{R}\colon \boldsymbol{H} \mapsto \|\boldsymbol{H}\boldsymbol{D}_T^{(k)}\|_1, \tag{19}$$

i.e., the absolute sum of k-th order variation of $\boldsymbol{H}$, enhances the smoothness of $\boldsymbol{H}$.

Finally, using (15) and (19) and extending the idea of the PCP for anomaly detection (13), we propose *Group-Sparse Plus Low-Rank Temporally-Smooth Decomposition for anomaly detection* as *a generalized PCP for anomaly detection*:

for $\boldsymbol{Y} \in \mathbb{R}^{L\times T}, \boldsymbol{R} \in \{0,1\}^{L\times F}$, nonempty closed convex sets $C_H, C_A \subset \mathbb{R}^{F\times T}$, and $\lambda_1 \geq 0, \lambda_2 \geq 0$,

$$\begin{aligned} \underset{(\boldsymbol{H},\boldsymbol{A})}{\text{minimize}}\ & \|\boldsymbol{H}\|_* + \lambda_1\|\boldsymbol{H}\boldsymbol{D}_T^{(k)}\|_1 + \lambda_2 \sum_{\mathcal{G}\in\mathfrak{G}} \|\pi_{\mathcal{G}}(\boldsymbol{A})\|_2 \\ \text{s.t.}\ & \boldsymbol{Y} = \boldsymbol{R}(\boldsymbol{H}+\boldsymbol{A}) \\ & \boldsymbol{H} \in C_H,\ \boldsymbol{A} \in C_A. \end{aligned} \tag{20}$$

Note that C_H and C_A represent prior knowledge on $\boldsymbol{H}$ and $\boldsymbol{A}$. A typical choice is $C_H = C_A = \mathbb{R}_+^{F\times T}$ to represents nonnegativity of traffic volumes.

Since problem (20) is convex optimization, its solutions can be approximated well by the so-called proximal splitting methods which are iterative algorithms including ADMM [26], forward-backward splitting [27], and primal-dual splitting [28]–[30]. Remark that (20) reverts the PCP for anomaly detection (13) when $\lambda_1 = 0$ and $\mathfrak{G} = \{\{f\} \times \{t\} \mid (f,t) \in \mathcal{F}\times\mathcal{T}\}$.

Remark 2 (On use of partial flow traffic volumes in the generalized PCP for anomaly detection (20)**):** *As mentioned in Remark 1, we can introduce the generalized PCP with the observation model* (8)*, as a counterpart of the generalized PCP* (20) *with the observation model* (6)*, by replacing* $(\boldsymbol{Y}, \boldsymbol{R})$ *in* (20) *with* $(\boldsymbol{Y}', \boldsymbol{R}')$ *in* (8)*.*

B. Application to DDoS Attack Detection

In case of DDoS attacks, traffics to target nodes increase abnormally as a large number of unspecified nodes send numerous packets to target nodes. Focusing this nature, we introduce groups of OD flow traffic volumes according to the destinations (and times), i.e.,

$$\mathfrak{G}_{\mathrm{DDoS}} := \{\mathcal{D}_j^{(t)} \subset \mathcal{F}\times\mathcal{T} \mid (j,t) \in \mathcal{N}\times\mathcal{T}\} \tag{21}$$

$$\text{where } \mathcal{D}_j^{(t)} := \{f \in \mathcal{F} \mid \mathrm{des}(f) = j\} \times \{t\}, \tag{22}$$

$$\mathrm{des}\colon \mathcal{F} \to \mathcal{N}\colon f \mapsto (\text{destination node of } f),$$

to estimate malicious traffic volumes as the solution of the generalized PCP for anomaly detection (20). Usually, since the number of the target nodes is relatively small, the vector $\left(\pi_{\mathcal{D}_j^{(t)}}(\boldsymbol{A}_\star)\right)_{(j,t)\in\mathcal{N}\times\mathcal{T}} \in \mathbb{R}^{NT}$ is expected to be sparse. In other words, $\boldsymbol{A}_\star$ is group-sparse with the groups $\mathfrak{G}_{\mathrm{DDoS}}$.

Using $\mathfrak{G}_{\mathrm{DDoS}}$ in the generalized PCP (20), we also present *a generalized PCP for DDoS attack detection*: for $\boldsymbol{Y} \in \mathbb{R}^{L\times T}, \boldsymbol{R} \in \{0,1\}^{L\times F}$, and $\lambda_1 \geq 0, \lambda_2 \geq 0$,

$$\begin{aligned} \underset{(\boldsymbol{H},\boldsymbol{A})}{\text{minimize}}\ & \|\boldsymbol{H}\|_* + \lambda_1\|\boldsymbol{H}\boldsymbol{D}_T^{(k)}\|_1 + \lambda_2 \sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{N}} \left\|\pi_{\mathcal{D}_j^{(t)}}(\boldsymbol{A})\right\|_2 \\ \text{s.t.}\ & \boldsymbol{Y} = \boldsymbol{R}(\boldsymbol{H}+\boldsymbol{A}) \\ & \boldsymbol{H} \in \mathbb{R}_+^{F\times T},\ \boldsymbol{A} \in \mathbb{R}_+^{F\times T}. \end{aligned} \tag{23}$$

IV. Numerical Examples

To evaluate the performance of the generalized PCP for DDoS attack detection (23), we conduct a numerical simulation on a toy example where the partial OD flows are available and where the DDoS attack and periodic DDoS attack occur.[6]

Consider the network in Fig. 1 with nodes $N = 15$ and links $L = 48$ (note: the number of flows is $F = 210$). We design routing matrix $\boldsymbol{R} \in \mathbb{R}^{L\times F}$ according to the shortest path routing algorithm. We suppose that the traffic volume of 20% of all the OD flows is available (i.e. $F_O = 42$), and hence the model matrix $\boldsymbol{R}' = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{I}^R \end{bmatrix} \in \{0,1\}^{(L+F_O)\times F}$ is defined as in (9). We sample the link and flow traffic volumes $T = 200$ times.

The measurements are generated by

$$\boldsymbol{Y}' = \left[\boldsymbol{R}'(\boldsymbol{H}_\star + \boldsymbol{A}_\star) + \boldsymbol{\varepsilon}'\right]_+, \tag{24}$$

where $[\cdot]_+$ represents the operation to replace negative entries by 0 and $\boldsymbol{\varepsilon}'$ is chosen from the Gaussian distribution with variance $\sigma^2 = 10^{-2}(\|\boldsymbol{R}'(\boldsymbol{H}_\star + \boldsymbol{A}_\star)\|_F^2/(LT + F_O T))$. Each

[5]
$$\boldsymbol{D}_\tau = \begin{bmatrix} -1 & 0 & \cdots & \cdots & 0 & 0 \\ 1 & -1 & \ddots & & 0 & 0 \\ 0 & 1 & \ddots & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & -1 & 0 \\ 0 & 0 & & \ddots & 1 & -1 \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \in \{-1,0,1\}^{\tau\times(\tau-1)}$$

[6] Further experiments involving actual traffic traces will be reported elsewhere.

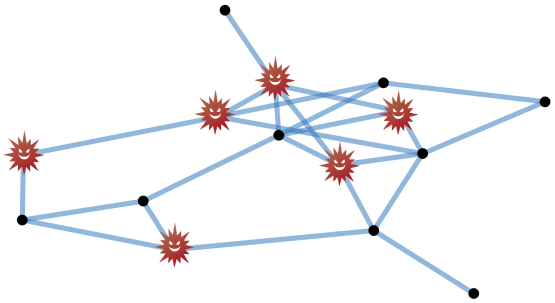

Fig. 1. Network topology: Dots and lines represent respectively nodes and links. Compromised nodes which take part in DDoS attacks are also described.

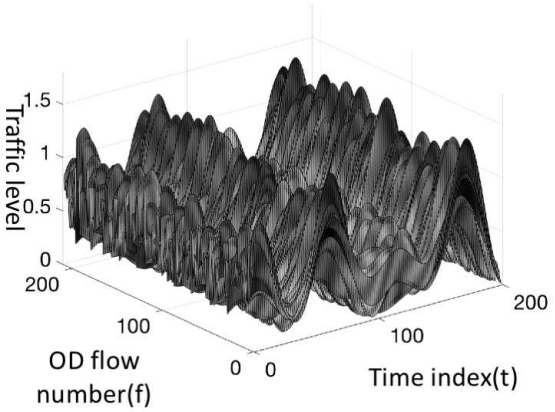

Fig. 2. Visualization of $\boldsymbol{H}_\star$

entry $h_{f,t}$ of $\boldsymbol{H}_\star = [h_{f,t}] \in \mathbb{R}^{F\times T}$ consists of the sum of $h_{f,t}^{(m)} (m = 1, \ldots, 5)$

$$h_{f,t} = \sum_{m=1}^{5} h_{f,t}^{(m)} \tag{25}$$

$$h_{f,t}^{(m)} = A_f^{(m)} \sin\left(2\pi B_f^{(m)} t - 2\pi C_f^{(m)}\right) + A_f^{(m)}, \tag{26}$$

where[7] $A_f^{(m)} \sim \mathcal{U}[0.05, 0.5], B_f^{(m)} \sim \mathcal{U}[0, \frac{7}{200}]$, and $C_f^{(m)} \sim \mathcal{U}[0, 1]$ (see Fig. 2 for visualization of $\boldsymbol{H}_\star$). The anomalous component $\boldsymbol{A}_\star$ is generated as follows. Six *compromised* nodes are chosen randomly in advance (see Fig. 1) and participate the typical DDoS attack as well as the periodic DDoS attack. Target nodes of the periodic DDoS attack are also randomly selected in advance. Simultaneously, all nodes are attacked randomly with a probability of 0.5% each time period. When node j is attacked at time t, each entry of $\pi_{\mathcal{D}_j^{(t)}}(\boldsymbol{A}_\star)$ (see (14) and (22)) is set to 5 if its corresponding flow is originated from a compromised node, and set to 0 otherwise. Fig. 3(a) depicts $\boldsymbol{A}_\star$ employed in this experiment.

We compare the PCP for anomaly detection (13) and the generalized PCP for anomaly detection (23). For the generalized PCP (23), $k(= 2)$nd-order total variation is utilized. The parameter λ in problem (13) and parameters (λ_1, λ_2) in problem (23) are chosen to achieve the best performance. The approximate solutions of the PCP (13) and the generalized PCP (23) are obtained by applying ADMM [26].

Figure 3(b) shows the approximate solution of the conventional PCP (13) and clearly demonstrates that the conventional method is inappropriate for the DDoS attack detection. Meanwhile, Fig. 3(c) shows that the approximate solution of the generalized PCP (23) succeeds in detecting DDoS attacks. The comparison between Fig. 3(b) and Fig. 3(c) clearly demonstrates utility of the generalized PCP (23) for DDoS attack detection.

[7]For $a, b \in \mathbb{R}$: $a < b$, the symbol $\mathcal{U}[a, b]$ denotes the continuous uniform distribution on $[a, b]$.

References

[1] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.

[2] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 365–377, 1996.

[3] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic matrix estimation: Existing techniques and new directions," *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 4, pp. 161–174, 2002.

[4] Q. Zhao, Z. Ge, J. Wang, and J. Xu, "Robust traffic matrix estimation with imperfect information: Making use of multiple data sources," *ACM SIGMETRICS Performance Evaluation Review*, vol. 34, no. 1, pp. 133–144, 2006.

[5] M. Roughan, Y. Zhang, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and internet traffic matrices," *IEEE/ACM Transactions on Networking*, vol. 20, no. 3, pp. 662–676, 2012.

[6] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 50–66, 2013.

[7] P. Tune and M. Roughan, "Internet traffic matrices: A primer," *Recent Advances in Networking*, vol. 1, pp. 1–56, 2013.

[8] S. William, "SNMP, SNMPv2, and CMIP: The practical guide to network management," 1993.

[9] A. Shaikh and A. G. Greenberg, "OSPF monitoring: Architecture, design, and deployment experience," in *USENIX Symposium on Networked Systems Design and Implementation*, vol. 1, 2004, pp. 57–70.

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

[11] A. Soule, A. Nucci, R. Cruz, E. Leonardi, and N. Taft, "How to identify and estimate the largest traffic matrix elements in a dynamic environment," *ACM SIGMETRICS Performance Evalution Review*, vol. 32, no. 1, pp. 73–84, 2004.

[12] A. Soule, K. Salamatian, and N. Taft, "Combining filtering and statistical methods for anomaly detection," in *ACM SIGCOMM Conference on Internet Measurement*, 2005, pp. 331–334.

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

[14] T. Kudo, T. Morita, T. Matsuda, and T. Takine, "PCA-based robust anomaly detection using periodic traffic behavior," in *IEEE International Conference on Communications Workshops*, 2013, pp. 1330–1334.

[15] Y. Zhang, Z. Ge, A. Greenberg, and M. Roughan, "Network anomography," in *ACM SIGCOMM Conference on Internet Measurement*, 2005, pp. 317–330.

[16] Y. Huang, N. Feamster, and R. Teixeira, "Practical issues with using network tomography for fault diagnosis," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 5, pp. 53–58, 2008.

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

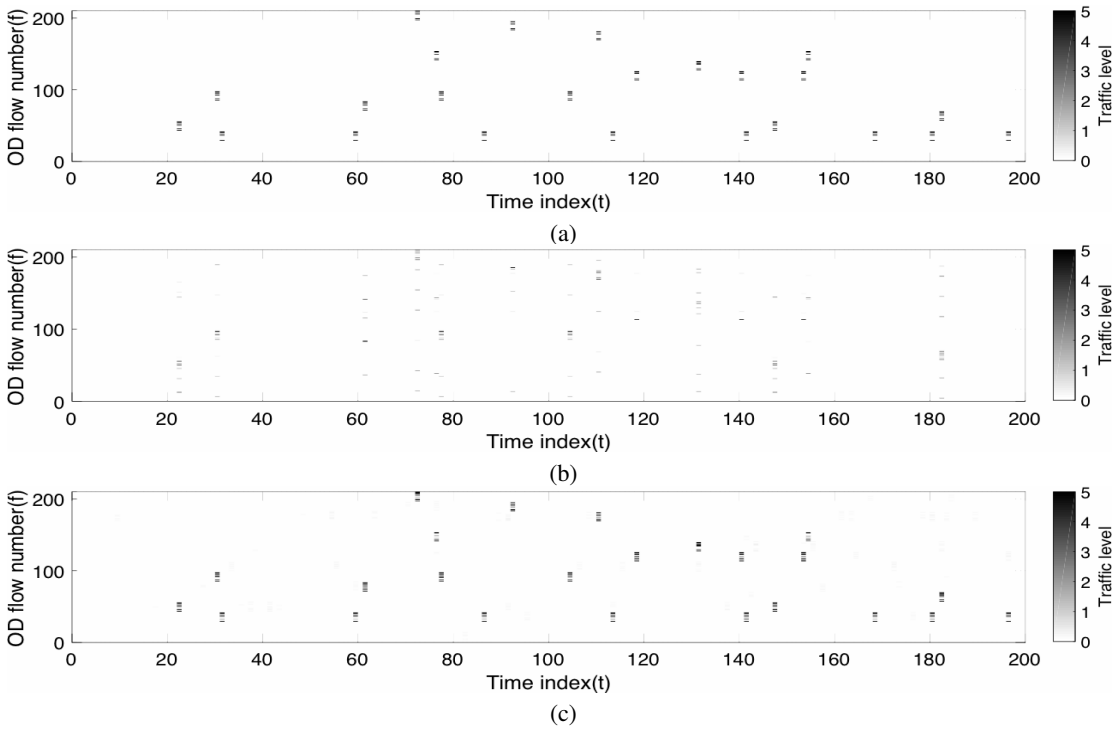

Fig. 3. (a) (True) $\boldsymbol{A}_\star$ imitating DDoS attack, (b) (Conventional) Detection with the PCP (13), (c) (Proposed) Detection with the generalized PCP (23).

[18] I. T. Jolliffe, "Principal component analysis and factor analysis," in *Principal Component Analysis*. Springer, 1986, pp. 115–128.

[19] A. Abdelkefi, Y. Jiang, W. Wang, A. Aslebo, and O. Kvittem, "Robust traffic anomaly detection with principal component pursuit," in *ACM CoNEXT Student Workshop*, 2010.

[20] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Sparse and low-rank matrix decompositions," in *IEEE Annual Allerton Conference on Communication, Control, and Computing*, 2009, pp. 962–967.

[21] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in Neural Information Processing Systems*, 2009, pp. 2080–2088.

[22] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2009.

[23] B. Claise, "Cisco systems netflow services export version 9," 2004.

[24] A. Ben-Israel and T. N. Greville, *Generalized inverses: Theory and applications*. Vol. 15, Springer Science & Business Media, 2003.

[25] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 22, no. 2, pp. 503–516, 2000.

[26] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[27] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[28] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[29] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, 2013.

[30] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.