Optimal Volume Anomaly Detection and Isolation in Large-Scale IP Networks using Coarse-Grained Measurements

P. Casas[a,c], S. Vaton[a], L. Fillatre[b], I. Nikiforov[b]

[a] *Télécom Bretagne, Brest, France*
[b] *Université de Technolgie de Troyes, Troyes, France*
[c] *Universidad de la República, Montevideo, Uruguay*

Abstract

Recent studies from major network technology vendors forecast the advent of the Exabyte era, a massive increase in network traffic driven by high-definition video and high-speed access technology penetration. One of the most formidable difficulties that this forthcoming scenario poses for the Internet is congestion problems due to traffic volume anomalies at the core network. In the light of this challenging near future, we develop in this work different network-wide anomaly detection and isolation algorithms to deal with volume anomalies in large-scale network traffic flows, using coarse-grained measurements as a practical constraint. These algorithms present well-established optimality properties in terms of false alarm and miss detection rate, or in terms of detection/isolation delay and false detection/isolation rate, a feature absent in previous works. This represents a paramount advantage with respect to current in-house methods, as it allows to generalize results independently of particular evaluations. The detection and isolation algorithms are based on a novel linear, parsimonious, and non-data driven spatial model for a large-scale network traffic matrix. This model allows detecting and isolating anomalies in the Origin-Destination traffic flows from aggregated measurements, reducing the overhead and avoiding the challenges of direct flow measurement. Our proposals are analyzed and validated using real traffic and network topologies from three different large-scale IP backbone networks.

Key words: Network Monitoring and Traffic Analysis, Traffic Matrix, Network Traffic Modeling, Optimal Volume Anomaly Detection and Isolation.

1. Introduction

After a brief mid-decade slowdown, IP traffic will nearly double every two years in the near future. The overall IP traffic is expected to grow from 6.6 exabytes per month in 2007 to nearly 29 exabytes per month by 2011 (1 exabyte = 10^{18} bytes), more than quadrupling in less than a half decade [1]. Simultaneously, the evolution of access technologies and the development of optical access networks (Fiber To The Home technology) will dramatically increase the bandwidth for end-users, imposing serious and unforeseen problems at the core network, so far assumed infinitely provisioned. One of the most difficult challenges for network operators will be to correctly manage the large and unexpected congestion problems at the core network caused by volume anomalies. These observations are part of the key findings provided by Cisco's global IP traffic forecast 2006-2011 [1, 2].

Volume anomalies represent large and sudden link load changes due to strong variations in traffic flows. These variations arise from unexpected events such as flash crowds, network equipment failures, network attacks, and external routing modifications and traffic shifts. Large-scale monitoring systems are currently deployed in ISP (Internet Service Providers) and large enterprise networks to fight back against these unexpected events. In this work we focus on two central aspects of traffic monitoring for volume anomaly detection: (i) the rapid and accurate detection of volume anomalies and (ii) the isolation of the origins of the detected anomalies.

The first issue corresponds to the anomaly detection field, a difficult and extensively studied problem. Anomaly detection in data networks consists of identifying patterns that deviate from normal traffic behavior. Different types of network anomalies can be detected depending on the monitored data. We particularly focus on device-level data. Device monitoring consists in using the widely spread Simple Network Management Protocol (SNMP) to periodically collect management device readings, known as Management Information Base (MIB) variables. Every network device has a set of MIB variables that are specific to its functionality, like memory usage, CPU load, and interface bandwidth usage among others. SNMP is unique in that it is supported by basically every device in an IP network. SNMP is the most basic means of data collection

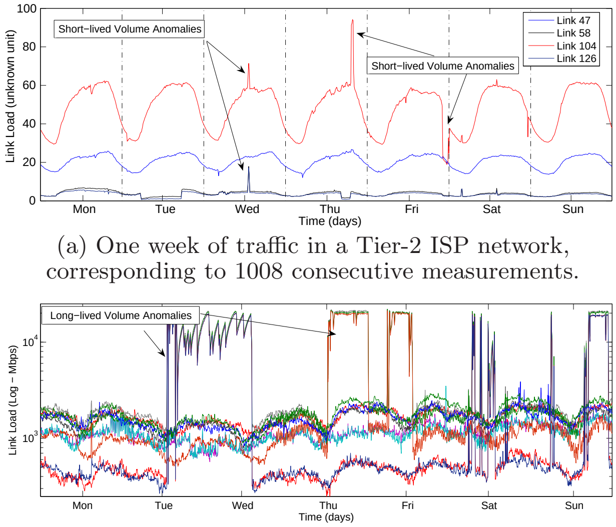
Email addresses: pedro.casas@telecom-bretagne.eu (P. Casas), sandrine.vaton@telecom-bretagne.eu (S. Vaton), lionel.fillatre@utt.fr (L. Fillatre), igor.nikiforov@utt.fr (I. Nikiforov)

for traffic analysis and provides the most coarse-grained information. At the same time it is the technique that causes the least measuring overhead, and thus represents an appealing choice for large-scale monitoring. However, it also has practical limitations, like missing data due to the use of the unreliable UDP transport protocol to export readings, or lack of readings synchronization in large-scale networks.

In this work we focus on network-wide volume anomaly detection, analyzing network traffic at the Origin-Destination (OD) flow level. An OD flow represents the total aggregated traffic flows transmitted between an ingress and an egress node or PoP (Point of Presence) in a network. A network-wide view of OD flows within a network is typically described by a Traffic Matrix (TM); a TM represents the total volume of traffic transmitted between every pair of ingress and egress points of a network. In practice, the term "volume of traffic" refers to the cumulative number of bytes between two consecutive measurements. The TM is a volume representation of OD flows traffic, and thus the types of anomalies we can expect to detect from its analysis are volume anomalies. Figure 1 depicts the occurrence of short-lived (a couple of hours at most) and long-lived volume anomalies in 1(a) four monitored links from a commercial international Tier-2 network and 1(b) several links from the Abilene network, an Internet2 backbone network in the US. As each OD flow typically spans multiple network links, a volume anomaly in one single OD flow is simultaneously visible on several links.

The algorithms that we develop in this work make use of standard SNMP per-link byte counts to detect volume anomalies in the TM. Link byte counts represent the accumulated number of bytes that cross through the link between two consecutive readings. From now on, we shall use the term "SNMP measurements" as a reference to this link data. The use of this aggregated coarse-grained data allows to conceive light and easy-to-deploy anomaly detection/isolation algorithms. However, it poses a challenging problem: the number of links in a network is generally much smaller than the number of OD flows, and thus the TM is not directly observable from link measurements.

The second issue that we address is the isolation of the origins of a detected anomaly. The isolation of an anomaly consists in inferring the exact location of the problem from a set of observed anomaly indications. This represents another critical task in network monitoring, given that a correct isolation may represent the difference between a successful or a failed countermeasure. In this work we assume that traffic anomalies are exogenous unexpected events (flash crowds, external routing modifications, external network attacks) that *significantly* modify the volume of one or multiple OD flows within the monitored network. For this reason, the isolation of the anomaly consists in finding the OD flows that suffer such a variation, referred from now on as the *anomalous* OD flows.



(a) One week of traffic in a Tier-2 ISP network, corresponding to 1008 consecutive measurements.

(b) One week of traffic in the Abilene network, corresponding to 2016 consecutive measurements.

Figure 1: Network volume anomalies in large-scale IP networks. Each measurement corresponds to the cumulative number of bytes between two consecutive SNMP readings.

1.1. Related Work

The anomaly detection literature treats the detection of general anomalous traffic behaviors [14, 25, 26, 28, 31, 33] as well as specific kinds of network and traffic anomalies. A basic list includes flash crowd events [12, 13], network failures [9, 10, 11, 15, 30], network attacks [16, 18, 19, 32], and large traffic shifts [20] among others. The majority of these works operate on individual and independent time series, analyzing traffic at a particular network link, particular device readings or particular packet characteristics with classical forecasting and outliers analysis methods. For example, [15] uses exponential smoothing EWMA and Holt-Winters forecasting techniques to detect anomalous behaviors in router readings. [14] analyses frequency characteristics of flow traffic and SNMP measurements using a wavelets based filtering approach, exposing anomalies as sharp variations in the filtered data variance. [17] builds compact summaries of flow traffic data using the notion of *sketch*, applying then the same forecasting techniques used in previous works (ARIMA, Holt-Winters, etc.) on top of such summaries to detect significant forecast errors. [16] uses spectral analysis techniques over TCP flows for DoS (Denial of Service) detection, using traffic traces from a single network link. [20] uses BGP (Border Gateway Protocol) and SNMP data streams to detect large traffic shifts, using EWMA, seasonal analysis and Holt-Winters over single time series to filter periodic and trend components, detecting anomalies as impulse functions. [12] characterizes flash crowds in Web servers and provides a network aware clustering approach to distinguish these events from DoS attacks, proposing an adaptive CDN (Content Delivery Network) architecture to fight back against these extreme events. [30] represents one of the first papers that uses multiple time series for anomaly detection, synthesizing information from multiple MIB variables at a single router to improve results. Contrary to these works, we treat the anomaly detection problem from a network-wide perspective, exploiting spatial correlations across the time series of traffic from all the links of a network.

Network-wide anomaly detection has also been treated in different works [25, 26, 27, 28, 31, 33]. The methods proposed in [26, 27, 28] make use of rich flow and packet data to detect anomalies, but this data that can be too costly to collect and to process [3]. [25] detects and classifies anomalies by jointly analyzing the distribution of OD flows and traffic features (IP addresses and ports). The authors use the technique of Principal Components Analysis (PCA) and the subspace method previously introduced in the field of fault diagnosis for chemical engineering processes [21, 22] to analyze the ensemble of OD flows and the corresponding traffic features in a network. [27] uses the idea of sketch proposed in [17] and the PCA approach to identify anomalous traffic flows. [28] proposes a recursive method to detect anomalies in multivariate time series, which is validated using the number of packets and the number of individual IP flows aggregated in a TM. On the contrary, our methods make use of easy to collect coarse-grained SNMP link data to detect and isolate volume anomalies in OD flows.

The use of SNMP measurements to detect volume anomalies in OD flows has been considered in [25, 31, 33], but none of these works has provided a complete and reliable solution to the problem. [31] uses a Kalman-filtering approach to track the evolution of OD flows from SNMP measurements, detecting anomalies as large prediction errors. The method requires a long training phase where direct anomaly-free OD flow measurements are used to calibrate the underlying model. As we have recently shown [8], the assumed model has a particular structure that may require several periodical recalibrations to provide reliable results, which makes the method too costly to implement from a practical point of view. Besides, the paper does not tackle the anomaly isolation problem. Only [25, 33] treat the problem of both anomaly detection and isolation in OD flows from SNMP measurements. The authors of [25] use the PCA approach and the subspace method proposed in [21, 22] to separate SNMP measurements into a normal subspace and an anomalous subspace, where anomalies are detected. The use of the PCA technique and the subspace method has probably become the most famous approach for network-wide anomaly detection in recent years. However, the approach is a pure data-driven in-house method, and recent works [28, 29] have shown categorical evidence about its serious shortcomings for anomaly detection and isolation in data networks. Finally, our approach falls into the same category as [33], where anomalies are inferred from aggregated data by combining network tomography and anomaly detection techniques. [33] uses similar methods applied in previous works to detect volume anomalies in OD flows: Fourier and Wavelet analysis, ARIMA modeling and PCA decomposition. The isolation of anomalies is performed with different heuristics which are not evaluated from a complexity perspective and that might be too time-consuming for on-line application; in fact, all evaluations performed in [33] are conducted off-line over individual datasets spanning one week of traffic each. Unlike that work, we provide detection and isolation algorithms that can be applied in an on-line fashion with solid theoretical support on their optimality properties.

1.2. Contributions of the Paper

Despite the large literature in the field, we can see that to date there is no single approach to correctly detect and isolate traffic anomalies in the TM from SNMP measurements in an on-line fashion. A reliable implementation of such approach would be highly beneficial for network operators, providing a light and easy to deploy first-line monitoring tool for on-line anomaly detection and isolation.

In this paper we present a complete approach that meets these criteria with solid optimality properties in terms of false alarm and miss detection rate, or in terms of detection/isolation delay and false detection/isolation rate, a feature absent in previous works. Optimality support is fundamental in the conception of general algorithms, not tied to any particular network and more important, independent of particular evaluations in particular network and traffic scenarios. In-house methods may work rather well in certain scenarios, but without a principled and generalizable support they can be easily rebutted.

We begin by introducing a new parsimonious, linear and parametric model for the anomaly-free TM. This model presents important advantages: (i) it uses exclusively easily-available coarse-grained SNMP measurements, simplifying practical issues; (ii) it is non data-driven and as we will show through evaluation with real data, it is stable in time, making it possible to design reliable anomaly detection methods on top of it; (iii) it is easy to calibrate and needs a very small amount of anomaly-free data to provide solid results; (iv) using this parsimonious model we can remove the anomaly-free traffic from the anomaly detection problem, thus treating the detection and isolation of volume anomalies as a sequential change detection/isolation problem with a nuisance parameter. This problem has been previously studied with some significant results [40, 41, 42]. In our particular case, this allows to design optimal algorithms for volume anomaly detection and isolation, using the principles of the decision theory.

Based on this traffic model, we propose two different optimal algorithms for volume anomaly detection and isolation. The first algorithm is designed for optimal detection, maximizing the correct detection for a bounded false alarm rate. The second algorithm permits to simultaneously detect and isolate a particular anomalous OD flow within the TM, minimizing the maximum mean detection/isolation delay for given bounds in the false isolation and false alarm rates. Since a few anomaly-free SNMP measurements are sufficient to obtain a reliable model for the anomaly-free TM, we claim that the proposed methods are well adapted to dynamic routing scenarios and non-stationary traffic, but this case is out of the scope of the current study.

To provide strong evidence on the effectiveness of our methods, all the proposed algorithms are validated using real traffic data from three different backbone networks: the Internet2 Abilene backbone network, the European GEANT academic network, and a commercial international Tier-2 network. Additionally, we compare our algorithms against well-known works in the field, showing that similar or even better performance can be achieved with thorough theoretical foundation. This work represents a continuation of our previous works on traffic modeling and volume anomaly detection [4].

The remainder of this paper is organized as follows. In Section 2 we present a linear parsimonious model to describe the anomaly-free OD flows traffic. Section 3 presents an optimal volume anomaly detection algorithm that maximizes the power of the test for a given false alarm rate. Section 4 presents a recursive algorithm for simultaneously detecting and isolating volume anomalies in single OD flows, minimizing the maximum mean detection/isolation delay for bounded false alarm and false isolation rates. In Section 5 we present an in-depth validation and evaluation of the traffic model and the detection/isolation algorithms, comparing their performance against well-known algorithms previously proposed. Section 6 discusses complexity and implementation issues of the proposed algorithms. Finally, Section 7 concludes this work.

2. Linear Parsimonious Traffic Matrix Modeling

The first and maybe the most critical step in anomaly detection is to conceive an accurate and stable traffic model for what constitutes an anomaly-free behavior. In this work we intend to detect volume anomalies in a backbone TM from SNMP measurements, thus we develop a traffic model for the anomaly-free behavior of the OD flows within a large-scale IP network. Throughout the paper, the vector $X_t = \{x_t(1), .., x_t(m)\}^T$ represents the value of the TM at time t, where $x_t(k)$ stands for the traffic volume of each OD flow $k = 1..m$ at measurement time t. Similarly, the SNMP measurements vector $Y_t = \{y_t(1), .., y_t(r)\}^T$ represents the links traffic volume at time t, where $y_t(i)$ represents the total traffic volume in link $i = 1..r$ at measurement time t. The TM X_t and the links traffic Y_t are related through the routing matrix R:

$$Y_t = R\, X_t \tag{1}$$

where R_{ij} is equal to 1 if OD flow j traverses link i and 0 otherwise. Note that we have intentionally omitted the subscript t in the routing matrix R; in this work we assume that R is constant in time. In Section 6 we discuss the implications of this choice.

Monitoring the behavior of X_t based on R and Y_t data represents a poorly posed problem, because the number of unknown OD flows is much larger than the number of links, $m >> r$. To solve this problem, we propose a spatial, linear, and low-dimensionality representation of X_t in the absence of volume anomalies. The basic idea of this model is that the traffic flows X_t, sorted from smallest to largest traffic volume can be decomposed at every time t over a known family of q basis functions (columns of the matrix S), $S = \{\mathbf{s}(1), \mathbf{s}(2), \ldots, \mathbf{s}(q)\}$, with the great virtue that $q << m$, even several orders of magnitude smaller (in the evaluation we show that $q < 10$ even for a network with more than $m > 1000$ OD flows). Therefore, we assume that X_t can be expressed as:

$$X_t = S\boldsymbol{\mu}_t + \boldsymbol{\xi}_t \tag{2}$$

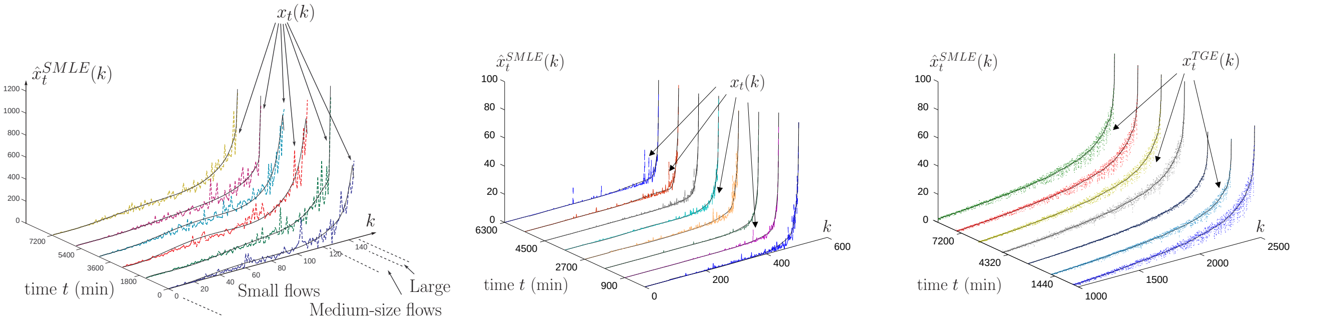
where $\boldsymbol{\xi}_t$ is a white Gaussian noise with covariance matrix $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$ that models the natural variability of the OD flows together with the modeling errors. The vector $\boldsymbol{\mu}_t = \{\mu_t(1) \ldots \mu_t(q)\}^T$ is an unknown time-varying vector which describes the OD flows intensity distribution with respect to the set of vectors $\mathbf{s}(i)$. We found in [8] that the order of OD flows sorted from smallest to largest traffic volume remains reasonably stable in time for several days in various large-scale networks, different not only in the topology but also in the nature of traffic. Figure 2 shows the OD flows traffic for (a) the Abilene network, (b) the GEANT network, and (c) a commercial Tier-2 ISP network, sorted from smallest to largest traffic volume, for different times t.

The sorted volumes of OD flows can be approximated by a non-decreasing function with a certain smoothness. The curve obtained by interpolating this function is parameterized by using a polynomial splines approximation. Given the shape of this curve, a cubic splines approximation is used. The spline basis is finally designed to approximate the sorted volume of OD flows by using m points uniformly chosen in the interval $[1; m]$. The vectors $\mathbf{s}(i)$ in S form the set of basis vectors that describe the spatial distribution of the traffic. From now on, we shall refer to this Spline-Based model as the SB model.

To illustrate the structure of the matrix S, let us consider the polynomial splines of degree $p = 3$ with $p - 1$ continuous derivatives and two integer knots k_1 and k_2 such that $1 < k_1 < k_2 < m$. A natural cubic spline $c(x)$ with the two knots k_1 and k_2 has the form:

$$\begin{aligned} c(x) \;&=\; \mu(1) + \mu(2)\, x + \mu(3)\, x^2 + \mu(4)\, x^3 \\ &\quad + \mu(5)\, (x - k_1)_+^3 + \mu(6)\, (x - k_2)_+^3 \end{aligned}$$

where x belongs to a real interval $[a; b]$ containing $[1; m]$, i.e. $[1; m] \subseteq [a; b]$, the reals $\mu(i)$ are the spline coefficients a $(x)_+ = \max\{0, x\}$. The interested reader can find additional information on splines representations in [34]. Then, the sampled vector $\mathbf{c} = (c(k))_{1 \leq k \leq m}$ verifies $\mathbf{c} = V\boldsymbol{\mu}$ where the matrix V is given by:



(a) Abilene Network (b) GEANT Network (c) Tier-2 ISP Network

Figure 2: Approximation of real OD flows (dashed lines) by the spline-based (SB) model (full lines) in 3 operational networks. $x_t(k)$ is the real volume of OD flow k. $\hat{x}_t^{SMLE}(k)$ stands for the estimated OD flow k using the SB model, defined in equation (6). $x_t^{TGE}(k)$ is the estimated OD flow k using the tomogravity estimation method, introduced in [5].

$$V = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2^2 & 2^3 & 0 & 0 \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \vdots & \vdots & \vdots & 0 & 0 \\ 1 & \vdots & \vdots & \vdots & 1 & 0 \\ 1 & \vdots & \vdots & \vdots & 2^3 & 0 \\ 1 & \vdots & \vdots & \vdots & \vdots & 1 \\ 1 & \vdots & \vdots & \vdots & \vdots & 2^3 \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & m & m^2 & m^3 & (m-k_1)^3 & (m-k_2)^3 \end{pmatrix}$$

The matrix S is obtained from V by permuting the rows according to the OD flows sorting order: the i-th row of S is the j-th row of V, provided that the OD flow i becomes the j-th OD flow after sorting from smallest to largest OD traffic volume.

It should be clear to the reader that the SB model cannot be generalized to all network topologies and scenarios, but that it holds for networks with large traffic aggregation. In the evaluation we show that this model provides accurate results for different network topologies and traffic scenarios, including a commercial network, the GEANT academic network, and Abilene, a network topology/traffic that is usually used as benchmark regarding TM studies. The dashed lines in figure 2 depict the value of each sorted OD flow $x_t(k), k = 1 \dots m$, the full lines represent the splines approximation of the sorted flows. In order to appreciate the time stability of this approximation, the curves are plotted for various consecutive days. From (1) and (2), we can express links traffic as a function of $\boldsymbol{\mu}_t$:

$$Y_t = G\boldsymbol{\mu}_t + \boldsymbol{\upsilon}_t, \tag{3}$$

where $G = RS$ and $\boldsymbol{\upsilon}_t \sim \mathcal{N}(0, \Phi)$, with $\Phi = R\Sigma R^T$. It is assumed that G is a full column rank matrix. In fact, since the number of columns in G is very small, the product RS and its rank can be computed very fast. To simplify notation and computations, we introduce the standardized measurements vector Z_t:

$$Z_t = \Phi^{-\frac{1}{2}} Y_t = H\boldsymbol{\mu}_t + \boldsymbol{\delta}_t, \tag{4}$$

where $H = \Phi^{-\frac{1}{2}} G$, $\boldsymbol{\delta}_t \sim \mathcal{N}(0, I_r)$ and I_r is the $r \times r$ identity matrix. The purpose of this transformation is simply to reduce a given noise covariance matrix to the identity one.

If the covariance matrix Σ is unknown some additional experiments should be done. The solution consists in computing an empirical covariance matrix $\widehat{\Sigma}$ from a few measurements; in Section 5.2 we show that using just 1 hour of SNMP measurements is enough to provide proper results. Some very basic results on the estimation of $\widehat{\Sigma}$ can be found in [35].

In this work we use this low-dimensionality model to filter the contribution of the anomaly-free traffic into the SNMP measurements, producing residuals sensitive to volume anomalies. As we explain in the following section, we treat the detection of volume anomalies as a statistical decision problem with a nuisance parameter, represented by the anomaly-free traffic. This allows to infer anomalies in the TM directly from aggregated data, without the preliminary TM estimation step. This approach clearly improves the accuracy and reduces the detection delay, because it does not drag possible errors from previous steps. Nevertheless and in order to validate the SB traffic model, we will use it to infer a TM from SNMP measurements in the validation Section 5.2.

The TM can be easily inferred from SNMP measurements using equation (4). We particularly use a maximum likelihood estimation approach to compute an estimated traffic matrix. The maximum likelihood estimate presents well established statistical properties [35]. Since the traffic linear model (4) is Gaussian, the maximum likelihood estimate of $\boldsymbol{\mu}_t$, namely $\hat{\boldsymbol{\mu}}_t^{MLE}$ corresponds to the least mean squares estimation:

$$\hat{\boldsymbol{\mu}}_t^{MLE} = (H^T H)^{-1} H^T Z_t \tag{5}$$

This finally leads to the estimate of the traffic matrix X_t, which we will refer as the Spline-based Maximum Likelihood Estimate (SMLE) $\hat{X}_t^{SMLE}$, defined by:

$$\hat{X}_t^{SMLE} = S\,\hat{\boldsymbol{\mu}}_t^{ML} = \left(S(H^T H)^{-1} H^T\,\Phi^{-\frac{1}{2}}\right) Y_t \qquad (6)$$

3. Optimal Volume Anomaly Detection

The goal of the proposed detection algorithm is to detect the presence of an additive anomaly $\boldsymbol{\varphi}$ in one or more OD flows of the traffic matrix X_t from the SNMP measurements vector Y_t, with the highest probability of detection for a given upper bounded probability of false alarm. The detection of this anomalous variation can be treated as a hypothesis testing problem, considering two possible traffic situations or hypotheses: the null hypothesis $\mathcal{H}_0$, where OD flows are anomaly-free, and the alternative hypothesis $\mathcal{H}_1$, where OD flows present an anomaly and thus traffic is no longer characterized by our anomaly-free-traffic model (3). For every new SNMP measurement, the method has to choose between $\mathcal{H}_0$ and $\mathcal{H}_1$ with the "best detection performance". We shall explain below what do we mean by best detection performance.

In order to continuously adapt the decision thresholds of the method, the previously introduced anomaly-free-traffic model is slightly modified, explicitly considering the temporal variation of the covariance matrix Σ. The Gaussian noise $\boldsymbol{\xi}_t$ is now assumed to have a covariance matrix $\gamma_t^2\Sigma$; the matrix $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$ is assumed to be known and stable in time. The scalar γ_t is unknown and serves to model the mean level of OD flows volume variance.

Considering equation (4), the previous hypothesis testing problem can be formulated as follows:

$$\mathcal{H}_0 \;=\; \{Z \sim \mathcal{N}(\boldsymbol{\varphi} + H\boldsymbol{\mu}, \gamma_t^2\, I_r);\; \boldsymbol{\varphi} = 0,\; \boldsymbol{\mu}{\in}\mathbb{R}^q\} \quad (7)$$

$$\mathcal{H}_1 \;=\; \{Z \sim \mathcal{N}(\boldsymbol{\varphi} + H\boldsymbol{\mu}, \gamma_t^2\, I_r);\; \boldsymbol{\varphi} {\neq} 0,\; \boldsymbol{\mu}{\in}\mathbb{R}^q\} \quad (8)$$

where $\boldsymbol{\varphi}$ represents an anomaly. Note that we have intentionally removed the time index t from Z and $\boldsymbol{\mu}$, explicitly stating that the test is applied for a single measurements vector $Z = Z_t$ at a certain time t. In the anomaly detection problem, the modeled anomaly-free traffic $\boldsymbol{\mu}$ is considered as a nuisance parameter since (i) it is completely unknown, (ii) it is not necessary for the detection and (iii) it could possibly mask the anomalies. In order to remove the nuisance parameter from the detection problem, the standardized measurements vector Z is projected onto the left null space of H, using the projection matrix $P_H^{\perp} = I_r - H\left(H^T H\right)^{-1} H^T$. Briefly speaking, we remove the "interference" of $\boldsymbol{\mu}$ from the problem. For this reason it is possible to chose between $\mathcal{H}_0$ and $\mathcal{H}_1$, provided that the projection of the anomaly $\boldsymbol{\varphi}$ onto the left null space of H is nonzero. For example, suppose that a volume anomaly of size θ occurs in OD flows j and k; then it is easy to see that $\boldsymbol{\varphi} = \theta\,\Phi^{-\frac{1}{2}}\,\mathbf{r}$, where $\mathbf{r}$ stands for the sum of the normalized columns $\mathbf{r}_j$ and $\mathbf{r}_k$ of the routing matrix R.

The quality of a statistical test is defined by the false alarm rate and the power function. The above mentioned testing problem is difficult because (i) $\mathcal{H}_0$ and $\mathcal{H}_1$ are composite hypotheses and (ii) there is an unknown nuisance parameter $\boldsymbol{\mu}$. A composite hypothesis refers to a statistical hypothesis that does not completely specify the probability distribution of the test statistic, i.e. it does not reduce to a single point into the probability space. There is no general way to test between composite hypotheses with a nuisance parameter.

Let K_α be the class of tests with an upper bounded maximum false alarm probability, $K_\alpha = \{\phi : \sup_{\boldsymbol{\mu}} \Pr_{\boldsymbol{\varphi}=0,\boldsymbol{\mu}}(\phi(Z) = H_1) \leqslant \alpha\}$, $0 < \alpha < 1$. The probability $\Pr_{\boldsymbol{\varphi}=0,\boldsymbol{\mu}}$ stands for the measurements vector Z being generated by the distribution $\mathcal{N}(H\boldsymbol{\mu}, \gamma_t^2\, I_r)$, and α is the prescribed upper bound for the probability of false alarm. The power function or hit rate is defined by the probability of correct detection $\beta_\phi(\boldsymbol{\varphi}, \boldsymbol{\mu}) = \Pr_{\boldsymbol{\varphi}\neq 0,\boldsymbol{\mu}}(\phi(Z) = \mathcal{H}_1)$. A priori, the power function depends on the parameter $\boldsymbol{\varphi}$ as well as on the nuisance parameter $\boldsymbol{\mu}$, which is highly undesirable.

In this work we use the statistical test $\phi^* : \mathbb{R}^r \mapsto \{\mathcal{H}_0, \mathcal{H}_1\}$ of [38, 39], inspired by the fundamental paper of Wald [37]. To solve this problem, Wald [37] proposes a test $\phi^*(\cdot) \in K_\alpha$, which has uniformly best constant power (UBCP) in the class K_α over a certain family of surfaces S. The adaptation of Wald's theory to the problem with nuisance parameters in the case of problem (7) - (8) has been done in [38, 39] by using the theory of invariant tests. Here, the family of surfaces of constant power $S = \{S_c : c \geq 0\}$ is defined by $S_c = \{\boldsymbol{\varphi} : \left\|P_H^{\perp}\boldsymbol{\varphi}\right\|^2 = c^2\}$. The UBCP invariant test realizes the best possible constant power $\beta_{\phi^*}(\boldsymbol{\varphi}, \boldsymbol{\mu}) = \beta_{\phi^*}(\boldsymbol{\varphi}', \boldsymbol{\mu}), \forall \boldsymbol{\varphi}, \boldsymbol{\varphi}' \in S_c$ and $\beta_{\phi^*}(\boldsymbol{\varphi}, \boldsymbol{\mu}) \geqslant \beta_\phi(\boldsymbol{\varphi}, \boldsymbol{\mu})$ over the tests with a given false alarm rate $\phi \in K_\alpha$. Finally, the threshold λ_α is chosen to satisfy the false alarm rate α, $\Pr_{\boldsymbol{\varphi}=0,\boldsymbol{\mu}}(\Lambda(Z) \geqslant \lambda_\alpha) = \alpha$. Hence, the test $\phi^*(\cdot)$ decides between $\mathcal{H}_0$ and $\mathcal{H}_1$ with the best detection probability for a bounded false alarm rate, which represents the major advantage of our approach.

The test is designed as follows, where $\|\cdot\|$ represents the Euclidean norm:

$$\phi^*(Z) = \begin{cases} \mathcal{H}_0 & \text{if } \; \Lambda(Z) = \left\|P_H^{\perp} Z\right\|^2 / \gamma_t^2 < \lambda_\alpha \\ \mathcal{H}_1 & \text{else} \end{cases} \qquad (9)$$

As we will show in Section 5, this strong theoretical support also has a major impact in practice, providing results that largely outperform previous proposals. The Optimal Spline-Based Detection method developed in this section will be referred as the OSBD method in the rest of the paper.

4. Optimal Sequential Volume Anomaly Detection and Isolation

In this section we introduce an optimal volume anomaly detection algorithm that has also the ability of isolating the anomaly, i.e. finding which is the particular OD flow responsible for the abnormal links traffic variation. We consider the same simplifying hypothesis as in [25], considering only "local" anomalies, namely anomalies in a single OD flow at one time. Different from Section 3, we now seek to detect and isolate an additional anomalous volume θ in one single OD flow k. This traduces into an additive change $\boldsymbol{\theta} = \theta\, \mathbf{r}_k$ in the SNMP measurements vector Y_t.

Instead of maximizing the probability of anomaly detection for a bounded false alarm probability, we design an algorithm that minimizes the maximum mean detection/isolation delay for an upper bounded probability of false isolation and a lower bounded mean time between consecutive false alarms, a usual measure of the false alarm rate. The mean detection/isolation delay is another crucial design criterion; indeed, the faster the detection and isolation, the faster the resolution of the problem.

The problem of detecting and isolating a volume anomaly that occurs at an unknown time t_0 is a particular case of a classical change detection/isolation problem, where the objective is to compute an alarm time T at which a change of type $\nu \in \{1, 2, \ldots, m\}$ in the probability distribution of a random sequence of measurements is detected. The alarm time T corresponds to the time when an anomaly in OD flow ν is detected and isolated. Before going into the details of the particular algorithm, let us formally define the optimality minimax criterion that we use in the design. The optimality criterion consists of minimizing the maximum mean delay for detection/isolation, given by:

$$\overline{\mathbb{E}}(T) = \sup_{t_0 \geq 1, 1 \leq k \leq m} \mathbb{E}^k_{t_0}(T - t_0 | T \geq t_0), \tag{10}$$

where $\mathbb{E}^k_{t_0}(T - t_0 | T \geq t_0)$ denotes the conditional expectation of $T - t_0$ when the event $\{T \geq t_0\}$ is true and the k-th change type occurs at time t_0, subject to the following constraints: (i) a lower bound for the mean time between two false alarms:

$$\mathbb{E}_0(T) \geq \upsilon \tag{11}$$

where υ is a prescribed lower bound and $\mathbb{E}_0(\cdot)$ denotes the expectation when all the measurements have the same probability density function f_0, corresponding to the anomaly-free traffic; (ii) an upper bound for the maximum probability of false isolation:

$$\max_{1 \leq k \leq m} \max_{1 \leq j \neq k \leq m} \sup_{t_0 \geq 1} \mathrm{Pr}^k_{t_0}(\nu = j | T \geq t_0) \leq \eta \tag{12}$$

where $\mathrm{Pr}^k_{t_0}(\nu = j | T \geq t_0)$ corresponds to the probability that the decision is j whereas the true change type is $k \neq j$. In brief, we require that the maximum mean detection/isolation delay given by (10) should be *as small as possible* subject to performance bounds on the mean time between consecutive false alarms and the maximum probability of false isolation.

In order to design an algorithm that verifies this minimax criterion, we shall treat the detection/isolation of a volume anomaly that occurs at an unknown time t_0 as a sequential hypothesis testing problem, where the null hypothesis $\mathcal{H}_0 = \{\text{OD flows are anomaly-free}\}$ ($t_0 = +\infty$) is tested against m alternatives $\mathcal{H}^k_{t_0} = \{$the k-th OD flow presents an anomalous additional amount of traffic θ from time $t_0\}$, $k = 1, \ldots, m$. Sequential approaches are used to minimize the number of observations needed to decide among the hypotheses. The sequential hypothesis testing problem can be written as:

$$\begin{aligned} \mathcal{H}_0 \quad &: \quad Z_t \sim \mathcal{N}(H\, \boldsymbol{\mu}_t, \gamma_t^2\, I_r)\,, \quad t = 1, 2, .. \\ \mathcal{H}^k_{t_0} \quad &: \quad \begin{cases} Z_t \sim \mathcal{N}(H\, \boldsymbol{\mu}_t, \gamma_t^2\, I_r), \ t = 1, .., t_0 - 1, .. \\ Z_t \sim \mathcal{N}(H\, \boldsymbol{\mu}_t + \theta\, \Phi^{-\frac{1}{2}}\, \mathbf{r}_k, \gamma_t^2\, I_r), t = t_0, .. \end{cases} \end{aligned} \tag{13}$$

where Z_t is the standardized measurements vector. As we did before, we can remove the nuisance parameter $\boldsymbol{\mu}$ from the detection problem. In order to only keep the anomalies-sensitive part of Z_t, we compute the residual process $U_t = W Z_t$, using a linear transformation W into a set of $r - q$ linearly independent variables. The matrix W^T is the linear rejector that eliminates the anomaly-free traffic by projection onto the left null space of H, built from the first $r - q$ eigenvectors of $P^{\perp}_{H}$ corresponding to eigenvalue 1. The rejector verifies the following relations: $WH = 0$, $W^T W = P^{\perp}_{H}$ and $W W^T = I_{r-q}$. Hypotheses $\mathcal{H}^k_{t_0}$ can be thus simplified by filtering the anomaly-free traffic:

$$\mathcal{H}^k_{t_0} : \begin{cases} U_t \sim \mathcal{N}(0, \gamma_t^2\, I_{r-q})\,, \quad t = 1, .., t_0 - 1, .. \\ U_t \sim \mathcal{N}(\theta\, \mathbf{v}_k, \gamma_t^2\, I_{r-q})\,, \quad t = t_0, t_0 + 1, .. \end{cases}$$

where $\mathbf{v}_k = W\, \Phi^{-\frac{1}{2}}\, \mathbf{r}_k$ corresponds to the signature in the residuals of a change in OD flow k.

The recursive algorithm proposed in [43, 44] perfectly fits this detection/isolation problem, with one useful feature, that of minimizing the mean number of samples needed to detect a change and decide among the different change types with bounded false alarm and false isolation rates. This algorithm is asymptotically optimal, i.e. it asymptotically minimizes the maximum mean delay for detection/isolation $\overline{\mathbb{E}}(T)$ when both the false alarm and the false isolation rates go to 0 : $\max\{\upsilon^{-1}, \eta\} \to 0$. The output of the recursive detection/isolation algorithm is twofold: (i) the alarm or stopping time T_r, which corresponds to the instant when an alarm is raised, and (ii) a decision ν_r, which corresponds to the type of change that the algorithm decides for among the m possible change types:

$$
\begin{aligned}
T_r &= \min_{1\leqslant k\leqslant m}\{T_r(k)\}, \quad \nu_r = \arg\min_{1\leqslant k\leqslant m}\{T_r(k)\} \\
T_r(k) &= \inf\{t\geqslant 1 : s_t(k)\geqslant 0\}, \quad k=1\ldots m \\
s_t(k) &= \min_{0\leqslant j\neq k\leqslant m}[g_t(k,j)-h_{k,j}], \quad k=1\ldots m \quad (14)
\end{aligned}
$$

with $g_t(k,j) = g_t(k,0) - g_t(j,0)$. The recursive functions $g_t(k,0)$ are defined by

$$
\begin{aligned}
g_t(k,0) &= (g_{t-1}(k,0)+u_t(k,0))_+ \quad (15)\\
u_t(k,0) &= \log\frac{f_k(U_t)}{f_0(U_t)} \quad (16)
\end{aligned}
$$

where $g_0(k,0) = 0$ for every $1 \leqslant k \leqslant m$ and $g_t(0,0) = 0$ for all t, f_0 is the probability density function of residuals under anomaly-free behavior and f_k is the probability density function of residuals $U_{t_0}, U_{t_0+1}, ..$ after the k-th type of change. The thresholds $h_{k,j}$ are chosen by the following formula:

$$
h_{k,j} = \begin{cases} h_\text{d} & \text{if } 1\leqslant k\leqslant m \quad \text{and } j=0 \\ h_\text{i} & \text{if } 1\leqslant k,j\leqslant m \quad \text{and } j\neq k \end{cases} \quad (17)
$$

where h_d and h_i are the detection and isolation thresholds. Basically, the anomaly detection is performed by comparing the m recursive functions $g_t(k,0)$ against the detection threshold h_d, while the anomaly isolation is performed by comparing the difference between these m recursive functions with the isolation threshold h_i. The stopping time $T_r(k)$ is the first time when the alternative hypothesis $\mathcal{H}^k_{t_0}$ is chosen by the sequential test as the most likely hypothesis. The stopping time T_r corresponds to the earliest of all the times $T_r(k)$ with $1 \leqslant k \leqslant m$. The detected anomaly is declared in OD flow k if the earliest of all these times was $T_r(k)$.

The choice of the detection and isolation thresholds h_d and h_i is discussed in [43], with practical comments and simulation results about the effectiveness of such thresholds. In practice, the detection threshold h_d is fixed so as to achieve the desired false alarm rate. As it follows from [43], some statistical issues of the recursive algorithm can be solved by choosing $h_\text{d} \geq h_\text{i}$, and thus we will generally consider $h_\text{i} = h_\text{d}$. In other words, given the desired false alarm rate, we fix h_d and take the biggest value of h_i to minimize the false isolation rate.

A final remark about the computation of the probability density functions in (16): f_0 is nothing but a Gaussian density function of law $\mathcal{N}(0, \gamma_t^2\, I_{r-q})$. In the case of f_k, the amplitude of the anomaly θ is completely unknown, and we must assume a certain distribution for it in order to correctly define f_k. Given that we are dealing with volume anomalies, it is reasonable to assume that the amplitude θ is uniformly distributed between two defined bounds θ_1 and θ_2. In this case, it is easy to see that f_k is simply a Gaussian mixture density. The bounds are introduced just for technical reasons and they can be chosen arbitrarily when dealing with volume anomalies. However, it is possible to control the sensitivity of the algorithm to detect small traffic changes instead of volume anomalies, see [44] for additional details. The choice of the bounds has little impact as regards anomaly isolation, because the signature is based on the direction of the anomaly and not on its amplitude.

The optimal sequential volume anomaly detection and isolation algorithm presented in this section will be referred as the Sequential Spline-Based (SSB) method in the rest of the paper.

Network	nodes/links	ODFlows	data	sampling
Abilene	12 - 54	132	OD flows	5’
GEANT	23 - 74	506	OD flows	15’
Tier-2 ISP	50 - 168	2450	SNMP	10’

Table 1: Network Topologies and Datasets.

5. Validation and Performance Evaluation

In this section we present the validation of the proposed traffic model and the evaluation of the anomaly detection/isolation algorithms using real and artificial measurements in different operational backbone networks. We first describe the datasets used in the evaluation; secondly, we validate the anomaly-free-traffic model and compare its ability to infer the TM from SNMP measurements against well-known methods in the field; then we compare the performance of the OSBD method against the well-known PCA approach [25]; finally, we evaluate the SSB method in different network scenarios and compare its performance against two celebrated algorithms, one based in a sequential implementation of the PCA approach and the other based on Kalman filtering techniques [31]. In all cases we show that the performance of our algorithms in practice is in agreement with the thorough theoretical foundation.

5.1. The Datasets

Data used for validation and evaluation consists of real traffic measurements from three operational networks: the Abilene network, an Internet2 backbone network at the US; the GEANT network, an European academic network; and a commercial Tier-2 ISP network. Table 1 presents the topology of each network. Abilene traffic data consists of 5’ sampled TMs collected via Netflow from the Abilene Observatory [45] and available at [46]. GEANT traffic data consists of 15’ sampled TMs, built from IGP and BGP routing information and Netflow data in [47], available at the TOTEM website [48]. The Tier-2 ISP network is a private commercial network and data is not public. Direct OD flow measurements are not available for this network. Instead, link traffic volumes are gathered every 10’ via SNMP. In order to validate our traffic model in this network topology, we compare our estimate against the well-known tomogravity estimate [5]. The tomogravity estimation method is a widely accepted method to estimate OD flow volumes from link traffic measurements, routing, and topology information.

In the following evaluations, we assume that traffic flows X_t are unknown and consider the SNMP measurements Y_t as the input known data. In order to verify the stability properties of the proposed model and algorithms, two sets of measurements are used for each network topology: the “learning” dataset, used for calibration purposes, and the “testing” dataset, used to evaluate the performance of the algorithms. We shall use T_{learn} and T_{test} as the sets of time indexes associated with measurements from the learning and testing datasets respectively.

5.2. Traffic Model Validation and Performance Evaluation

The spatial SB model presented in this work is the first parametric, linear, and parsimonious model for the TM proposed in the literature. For this reason, we provide substantial evidence of its relevance and applicability in the three presented networks. Let us begin by showing that the model is stable in time and that it permits to correctly infer OD flows volume from SNMP measurements, using the Abilene and the GEANT datasets.

The learning dataset is composed of one hour of SNMP measurements and it is used to construct the splines basis S. Given that the sampling rate in GEANT is smaller than the one used in Abilene, we interpolate intermediate measurements in the learning dataset of the former topology. The testing dataset is composed of 672 consecutive SNMP measurements. The learning dataset is measured one hour before the testing dataset. The SB model is computed for each network using each learning dataset, following these steps: (i) the tomogravity estimate (TGE) $\hat{x}_t{}^{TGE}(k)$ is computed for all OD flows k and all $t \in T_{\text{learn}}$; (ii) the mean flow values $\bar{x}\ ^{TGE}(k) = \frac{1}{\#(T_{\text{learn}})} \sum_{t \in T_{\text{learn}}} \hat{x}_t{}^{TGE}(k)$ are computed, where $\#\,(T_{\text{learn}})$ is the number of time indexes in the learning dataset; (iii) finally, the obtained mean values $\bar{x}\ ^{TGE}(k)$ are sorted in ascending order to obtain a rough estimate of the OD flows traffic volume. In both cases the SB model is designed with cubic B-splines ($p = 3$) and 2 knots, representing small, medium-size, and large OD flows, see figure 2. The use of cubic splines comes directly from the shape of the curve to approximate. We use the Matlab Splines Toolbox to design q splines $\mathbf{s}(i)$, $1 \leq i \leq q$. The choice of cubic splines and the number of knots results in a total of $q = (p+1)+2 = 6$ splines [34]. This clearly reflects the low-dimensionality of our anomaly-free-traffic model, as q is effectively much smaller than m for both network topologies. The mean value $\bar{x}\ ^{TGE}(k)$ of each OD flow is used to compute an estimate $\hat{\sigma}_k^2$ of σ_k^2, which leads to an estimate $\hat{\Phi}$ of Φ, quite efficient and sufficient in practice.

The obtained calibrated model is used to infer the OD flows volume from the SNMP measurements of the testing dataset, using the SMLE estimate defined in (6). To qualify the accuracy of the SMLE estimate and to test the performance of the short learning step, we compute the relative root mean squared error (RRMSE) for every time t in the testing dataset:

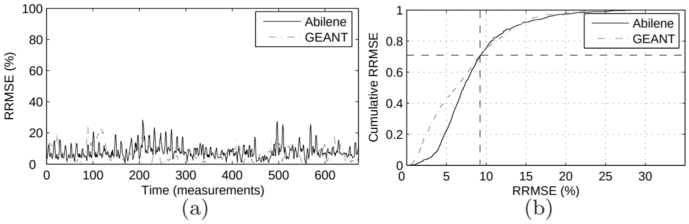


Figure 3: (a) RRMSE(t) and (b) Cumulative RRMSE(t) for 672 measurements in Abilene and GEANT

$$\text{RRMSE}(t) = \frac{\sqrt{\sum_{k=1}^{m} \left(x_t(k) - \hat{x}_t^{SMLE}(k)\right)^2}}{\sqrt{\sum_{k=1}^{m} x_t(k)^2}}\ , \quad \forall t \in T_{\text{test}} \quad (18)$$

where $x_t(k)$ is the true traffic volume of OD flow k at time t and $\hat{x}_t^{SMLE}(k)$ denotes the corresponding SMLE estimate. The RRMSE has been used in previous works [6, 7] as a summary of the relative estimation error for all m OD flows at every time t. Figure 3(a) presents the temporal evolution of the RRMSE for the 672 measurements in the testing datasets for Abilene and GEANT. In both cases, the relative error remains stable in time. This result confirms the hypothesis about a certain time-invariance which has been concluded from figure 2.

Figure 3(b) shows that more than 70% of the time estimation relative errors are below 10%. A deeper study of the RRMSE shows that in most cases, large RRMSE values correspond to large relative errors in the lowest-volume OD flows, which are well-known to be hard to estimate [5]. Note however that small OD flows have little impact as regards our problem of volume anomaly detection and are generally less important to estimate. The mean values of the RRMSE for the evaluation period are 8.14% for Abilene and 7.04% for GEANT. Methods proposed in the literature as “accurate” estimates present relative errors that vary between 5% and 15% [6, 7], thus we conclude that the obtained results are highly satisfactory.

In the validation of the model for the Tier-2 ISP network, we compare the value of the SMLE estimate $\hat{x}_t^{SMLE}(k)$ against the tomogravity estimate $\hat{x}_t^{TGE}(k)$, using the relative root mean squared difference (RRMSD) between both estimates:

$$\text{RRMSD}(t) = \frac{\sqrt{\sum_{k \in top\,TG\text{-}T_h} \left(\hat{x}_t^{TGE}(k) - \hat{x}_t^{SMLE}(k)\right)^2}}{\sqrt{\sum_{k \in top\,TG\text{-}T_h} \left(\hat{x}_t^{TGE}(k)\right)^2}}\ , \ \forall t \in T_{\text{test}} \quad (19)$$

Comparing all flows in (19) is not a reasonable approach. The tomogravity estimate provides quite accurate results for relatively high-volume flows, but poor for small flows

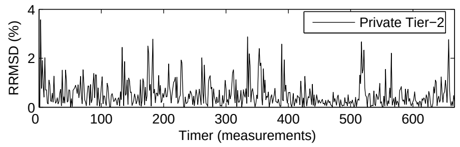
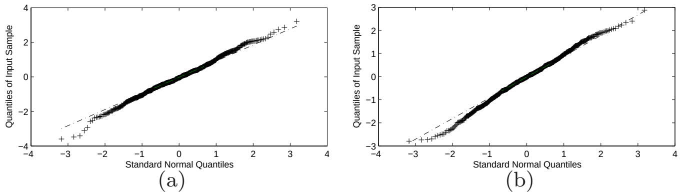


Figure 4: RRMSD(t) for 1500 flows in a Tier-2 ISP network



(a)

(b)

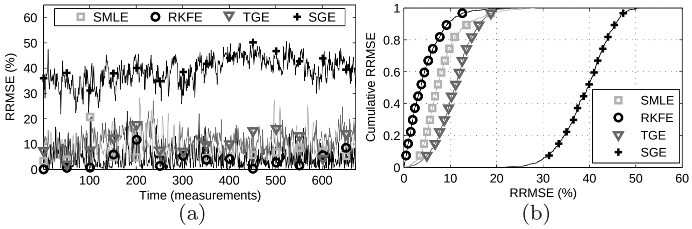
Figure 5: QQ-plots for 2 residual processes from (a) Abilene and (b) GEANT.

[5]; we define the $topTG$-T_h flows as those estimated flows by the tomogravity method that are stable in time and which mean value exceeds a threshold T_h. In this sense we only keep the most accurately estimated flows, removing the noisy or erratic estimates which seem to be wrongly estimated.

Figure 4 depicts the temporal evolution of the RRMSD between the TGE and SMLE estimates, for a Tier-2 ISP network. In this evaluation, we tune T_h such that 60% of the total flows are compared in the RRMSD index, which represents approximately 95% of the total traffic volume. The relative difference between TGE and SMLE is stable in time and has a mean value of 0.57%. This seems reasonable since the splines decomposition conducted in the training dataset is based on the TGE estimate. Based on our previous observations about the tomogravity estimate, we conclude that the SB model is also appropriate for this Tier-2 ISP network.

As a final validation of the SB model, we verify the Gaussian assumption for Abilene and GEANT, analyzing the residual processes U_t. Quantile-Quantile plots for two of these residual processes are plotted in figure 5, both for Abilene and GEANT. These residual processes clearly follow a Gaussian distribution. We also verify the Gaussian assumption by applying a Kolmogorov-Smirnov goodness-of-Fit hypothesis test to the residual processes. The acceptance rate of this test at the 5% level is 98.5% for Abilene and 97.7% for GEANT, which also confirms the Gaussian assumption.

An obvious question that arises when introducing a new TM model is how accurate this model is as regards the inference of a TM with respect to existing work in the literature. Figure 6 presents a comparative summary of the performance of the SMLE estimate in Abilene, considering three well-known TM inference methods: a Recursive Kalman Filter Estimate (RKFE), the Simple Gravity Estimate (SGE) [5], and the Tomo-Gravity Estimate (TGE) [5]. The RKFE method [8] corresponds to an enhanced extension of the recursive TM estimation method presented in [6]. This method uses a 24hs learning dataset composed of direct OD flow measurements for calibration purposes. The obtained mean values of the relative error are 8.14%, 4.48%, 11.15%, and 39.08% for the SMLE, RKFE, TGE, and SGE respectively. From figure 6(b) we can see that the SMLE and the RKFE produce estimation relative errors below 10% for approximately 75% and 92% of the TMs respectively, while this result drops to nearly 40% for the TGE, and to 0% for the SGE. The better performance achieved by the RKFE method has a clear explanation: the Kalman filter uses all previous SNMP measurements until time t to perform an estimated TM at time t, while the rest of the methods only use Y_t to produce an estimate $\hat{X}_t$. The performance gain of the SMLE method w.r.t. the TGE method may not be that important, but the SB model has a clear advantage: its parametric, linear, and parsimonious structure allows to define optimal algorithms. For example, the SMLE is asymptotically optimal, i.e. it is asymptotically unbiased and efficient, which is not the case for the TGE. Presented results evidence the accuracy of the proposed spatial model w.r.t. previous highly respected work.



(a)

(b)

Figure 6: (a) RRMSE(t) and (b) Cumulative RRMSE(t) for 672 measurements in Abilene, for the SMLE, the RKFE, the TGE, and the SGE.

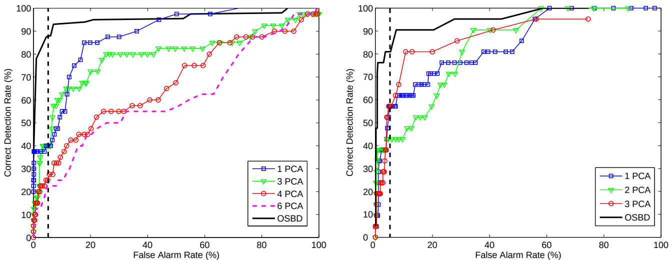
5.3. Numerical Evaluation of the Optimal Detection Algorithm

The performance of the OSBD method presented in Section 3 is compared against the performance obtained with the well-known PCA approach introduced in [25]. This method is chosen as benchmark given its relevance in the anomaly detection literature [23, 24, 25, 27, 29]. The PCA approach consists of a decomposition of the SNMP measurements into a principal components basis, separating traffic into a *normal subspace* that captures the anomaly-free traffic behavior, and an *anomalous subspace* that provides residuals sensitive to anomalies. This approach as presented in [25] is not designed to work on-line; instead, the analysis is performed off-line over a time window of n consecutive SNMP measurements vectors $\mathbf{Y}_{1..n} = \{Y_{t_1}, Y_{t_2}, .., Y_{t_n}\}^T$ ($n = 1008$ consecutive measurements in [25], which corresponds to one week of traffic). Each column in $\mathbf{Y}$ represents a time series of n samples of SNMP measurements for each network link. The normal subspace $\mathcal{S}$ corresponds to the space spanned by the first k principal components of $\mathbf{Y}_{1..n}$, namely $\mathbf{u}_{i=1..k}$, while the remaining $r-k$ components are used to build the anomalous subspace $\hat{\mathcal{S}}$. Given $\mathcal{S}$ and $\hat{\mathcal{S}}$, every SNMP measurements vector $Y \in \mathbf{Y}_{1..n}$ can be separated into the modeled traffic Y_{model} and the residual traffic Y_{residual} by simple projection onto $\mathcal{S}$ and $\hat{\mathcal{S}}$ respectively:

$$Y = \overbrace{\mathbf{PP}^T Y}^{Y_{\text{model}}} + \overbrace{\left(\mathbf{I}-\mathbf{PP}^T\right) Y}^{Y_{\text{residual}}} \tag{20}$$

where $\mathbf{P} \in \mathbb{R}^{r\times k}$ stands for the matrix with the first k principal components $\mathbf{u}_{i=1..k}$ as column vectors and $\mathbf{PP}^T$ represents the projection matrix onto the normal subspace. The anomaly detection is finally performed in the residual traffic, looking for large changes in the squared norm of residuals, $||Y_{\text{residual}}||^2$.

Let us evaluate and compare the performance of the OSBD method and the PCA approach. We shall use two testing datasets, composed of 720 consecutive SNMP measurements from the Abilene and the GEANT networks. The learning datasets for the OSBD method consist of one hour of anomaly-free SNMP measurements, gathered one hour before the testing datasets. In the case of the PCA approach, the method is directly applied to each complete testing dataset $\mathbf{Y}_{1..720} = \{Y_{t_1}, Y_{t_2}, .., Y_{t_{720}}\}^T$. For the sake of false alarm and correct detection rates evaluation, the set of "true" anomalies is manually identified in each testing dataset. Manual inspection declares an anomaly in an OD flow if the unusual deviation intensity of the guilty OD flow leads to an increase of traffic (i) larger than 1.5% of the total amount of traffic on the network and (ii) larger than 1% of the amount of traffic carried by the links routing this guilty OD flow, for each of these links. This rule is based on the conclusions about large traffic changes drawn in [33]. Hence, only large volume anomalies are considered as "true anomalies". 40 measurements of the Abilene testing dataset are affected by at least one significant volume anomaly. In the case of the GEANT testing dataset, 36 anomalous measurements are identified. Different from the PCA approach, the OSBD method is applied to the SNMP measurements of each testing dataset in an on-line fashion, sequentially running the test defined in (9) for every new "incoming" SNMP measurement $Y_{t_1}, Y_{t_2}, .., Y_{t_{720}}$. For the detection purpose, it is crucially important to have a good estimate of γ_t. This parameter is easily estimated from the learning dataset by using the maximum likelihood estimate of noise variance in residuals U_t [35]. Since this parameter can slowly vary in time, its value is updated during the test: at time t, if no anomaly has been declared in the last hour, γ_t is estimated by its value one hour before.



(a) ROC Curves for Abilene (b) ROC Curves for GEANT

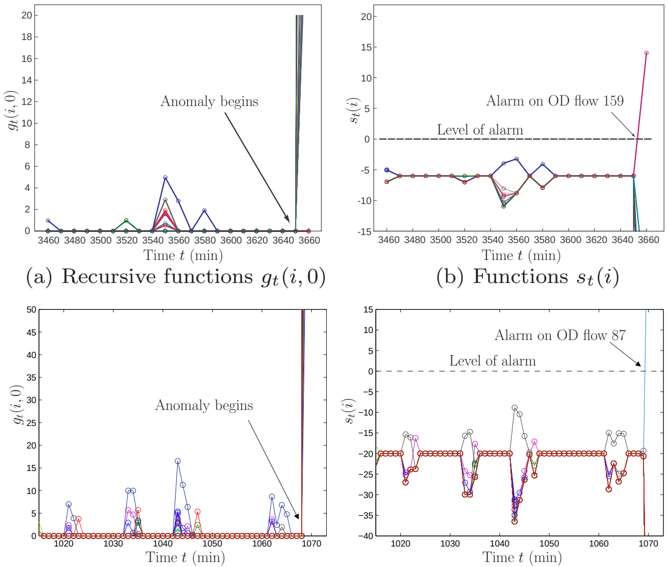
Figure 7: Correct detection rate vs false alarm rate for the OSBD method (solid line) and the PCA approach, considering a different number of k first principal components $\mathbf{u}_k$ to model the normal subspace.

Figure 7 depicts the ROC curves for the OSBD and the PCA methods in the Abilene and the GEANT datasets, showing the correct detection rate β for different values of the false alarm rate α, corresponding to different values of the detection threshold. In the PCA approach, a different number of k first principal components $\mathbf{u}_k$ is used to model the normal subspace. Results obtained with the PCA approach in the Abilene dataset are quite far from those obtained with the OSBD method; the PCA test presents more than 2 times lower detection rates for a reasonable false alarm rate, below 5%. For example, for a false alarm rate $\alpha = 1\%$, the OSBD method correctly detects almost 80% of the anomalies, while this value drops to nearly 40% for the best performance of the PCA approach (using 1 principal component $\mathbf{u}_1$ to model the normal subspace). Results are quite similar for the GEANT dataset, but in this case the best performance of the PCA approach is attained using 3 principal components $\mathbf{u}_{i=1..3}$ to model the normal subspace. Figure 7 also evidences the lack of consistency of the PCA approach as regards the number of principal components used to model the anomaly-free traffic; for the same dataset, results are quite different when this number slightly varies. For the different datasets, the number of principal components that provides better results also differs, which makes it difficult to generalize results. As it is shown in recent works [29], the PCA approach has to be highly tuned for each particular dataset in order to provide reliable results, making it inapplicable in a general real scenario. In fact, the main problem with current in-house methods is the difficulty to generalize their results.

The last important observation is that the OSBD method provides highly accurate results with a remarkably short learning-step, reinforcing the stability properties of the underlying parametric anomaly-free-traffic model and the robustness of the approach. On the contrary, the PCA approach provides a completely data-driven model for anomaly-free traffic, resulting in the aforementioned shortcomings.

5.4. Performance Evaluation of the Sequential Anomaly Detection and Isolation Algorithm

Le us first demonstrate the ability of the SSB algorithm to detect and isolate an OD flow volume anomaly from SNMP measurements in two different networks, the commercial Tier-2 network and the Abilene network. Figure 8 shows a typical realization of functions $s_t(i)$ and $g_t(i,0)$ defined in (14) and (15) respectively. Functions $s_t(i)$ are used



(a) Recursive functions $g_t(i, 0)$

(b) Functions $s_t(i)$

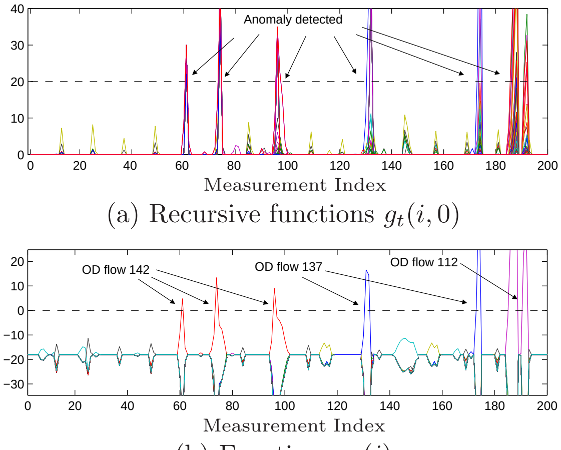
(c) Recursive functions $g_t(i, 0)$

(d) Functions $s_t(i)$

Figure 8: Typical realizations of anomaly detection/isolation functions for a Tier-2 network (a,b) and Abilene (c,d).

to “monitor” the OD flows; when $s_t(i)$ exceeds the threshold 0, OD flow i is declared anomalous. The anomaly in the Tier-2 network begins at time 3660 min, and at time 1070 min in Abilene. Note that after this time, several recursive functions $g_t(i, 0)$ rapidly grow in both network scenarios. Each function $g_t(i, 0)$ is associated with OD flow i and when this function increases, it means that OD flow i is suspected of carrying an abnormal amount of traffic. Contrary to $g_t(i, 0)$, only function $s_t(159)$ associated to anomalous OD flow 159 increases and finally exceeds the threshold 0 in the Tier-2 network. In the case of Abilene, the anomaly is correctly isolated in OD flow 87. Hence, functions $s_t(i)$ permit to isolate the anomalous OD flow among all the OD flows associated to functions $g_t(i, 0)$ that have rapidly increased. The volume anomalies detected in the examples of figure 8 correspond to abrupt and massive volume augmentations, and thus functions $s_t(i)$ only need one observation to detect and isolate the anomalous OD flow. Since the underlying sampling rates of both datasets are 10’ and 5’ for the Tier-2 and the Abilene networks respectively, the detection delay corresponds to 10’ and 5’, respectively. Note however that our algorithm in not intrinsically tied to any particular sampling rate, thus this detection delay would be even shorter if the sampling rates were higher. An interesting observation of this evaluation is that the SSB algorithm achieves accurate results in both datasets, even though the respective anomaly-free traffic behaviors are quite different between these two networks.

Let us now compare the performance of the SSB algorithm to continuously detect and isolate volume anomalies in real-time against two sequential methods in the literature: the Kalman-Based method (KB) presented in [31], and a sequential implementation of the previously described PCA method that we will reference as the Sequential PCA method (SPCA). This sequential extension of the PCA approach comes from the authors of the former PCA method in [23, 25], but the method was never evaluated in their anomaly detection work [25]. The idea is straightforward; the principal components and the corresponding projection matrix $\mathbf{PP}^T$ are built off-line from a certain time window $[t_1, t_n]$ of SNMP measurements $\mathbf{Y}_{1..n}$; subsequently, every new arriving measurement Y_t at time $t > t_n$ is processed on-line using this projection matrix.



(a) Recursive functions $g_t(i, 0)$

(b) Functions $s_t(i)$

Figure 9: On-line volume anomaly detection and isolation in Abilene, using the SSB method. The time between consecutive measuremmets is 5 minutes.

The PCA and the subspace methods can also be used to detect single OD flow volume anomalies. In the subspace framework, a particular volume anomaly represents a displacement of the SNMP measurements vector Y from the normal subspace $\mathcal{S}$ into a particular direction. The authors of [25] propose to find the single largest OD flow volume anomaly that best describes this deviation by simply using a greedy search algorithm. We apply this approach to isolate volume anomalies with the SPCA method as well. The KB method is only used for volume anomaly detection as presented in [31], thus we do not intend to use it for anomaly isolation. Similar to Section 5.2, we use the enhanced extension of the recursive traffic model presented in [8] for the KB method. In order to use the SSB method to continuously detect and isolate volume anomalies, the algorithm statistics are reset to 0 after each anomaly detection, i.e., $g_t(i, 0)$ is set to 0 after a change detection at time t, $\forall i = 1..m$. Figure 9 shows how the SSB method works on-line, continuously detecting and isolating volume anomalies in Abilene.

The testing dataset used for the evaluation consists of 864 consecutive SNMP measurements from the Abilene network. Instead of manually identifying the set of true volume anomalies, we introduce synthetic volume anomalies into this set. Indeed, in order to test the volume anomaly isolation algorithms, we need to know exactly which is the anomalous OD flow. Additionally, we need to be sure that a volume anomaly only occurs at a partic-

Method	Detected	False Alarms	Isolated
SSB	93.9 %	1.4 %	90.8 %
KB	90.8 %	1.3 %	n/a
SPCA ($\mathbf{u}_1$)	76.9 %	1.9 %	73.9 %
SPCA ($\mathbf{u}_{i=1..3}$)	53.9 %	1.7 %	49.2 %

Table 2: Results of the detection and isolation for 864 SNMP measurements in Abilene, composed of 65 OD flow volume anomalies.

ular OD flow at one time, so as to fulfill the simplifying hypothesis of single OD flow anomalies considered in Section 4. We follow a similar procedure as that described in [31] to introduce 63 large synthetic volume anomalies. The basic idea of this procedure consists in extracting the long-term trend from each OD flow, adding a Gaussian noise to these "smoothed" OD flows and finally adding the synthetic volume anomalies to this "anomaly-free" smoothed dataset. These anomalies correspond to short-lived volume changes in particular single OD flows. We additionally add two short-lived volume anomalies that span multiple OD flows at the same time, in order to analyze the response of the single OD flow volume anomaly isolation algorithms in that case.

Table 2 presents the comparative performance of the three algorithms. As before, the training dataset for the SSB method consists of 1 hour of anomaly-free SNMP measurements. As in [23], the training dataset for the SPCA method consists of 1 week of SNMP measurements, gathered immediately before the testing dataset and not necessarily free of volume anomalies. The PCA decomposition as proposed in [23, 25] is directly applied to unfiltered data, thus we follow this approach. Similarly to [31], the learning dataset for the KB method consists of 24 hours of anomaly-free direct OD flow measurements. The detection thresholds for the three methods are set so as to achieve a false alarm rate of about 1% in the testing dataset. As we have previously stated in Section 4 and considering the observations in [43], the isolation threshold of the SSB method is set to the same value as the detection threshold, i.e., $h_{\mathrm{i}} = h_{\mathrm{d}}$ in equation (17). In order to appreciate the sensitivity of the SPCA method to the dimensionality of the normal subspace, we consider two different representations for $\mathcal{S}$, using 1 and 3 principal component(s) respectively.

The SSB method correctly detects 61 out of the 65 volume anomalies, producing a total of 12 false alarms on the 864 measurements of the testing dataset. From the 61 detected anomalies, 59 are correctly identified in the particular anomalous OD flows. The two volume anomalies that are not correctly isolated correspond to those anomalies that span multiple OD flows simultaneously. In this case the algorithm certainly produces an alarm, but the isolation step can not correctly distinguish between the anomalous OD flows. In the following section we discuss an approach to solve this problem.

Detection results are similar for the KB method, which correctly detects 59 anomalies with only 11 false alarms. Obtained results are less accurate with the SPCA approach and many anomalies go undetected. Using 1 principal component to construct the normal subspace, the SPCA method correctly detects 50 volume anomalies while triggering 16 false alarms. The detection threshold of the SPCA approach can be tuned so as to correctly detect 89% of the anomalies, but the false alarm rate climbs to approximately 6% in that case, a value almost 5 times bigger than the rest of the methods. The SPCA method has a similar problem to isolate multiple OD flows anomalies, in this case because the greedy search we used only looks for single OD flow anomalies. However, studies in [29] show that correctly identifying the anomalous OD flows with the PCA approach is inherently difficult. Results are quite poor when using 3 principal components to model $\mathcal{S}$, only detecting 35 volume anomalies and isolating 32. These results are consistent with the sensitivity analysis and the highlighted shortcomings of the PCA approach presented in [29, 28].

6. Discussion

In this section we shall focus on complexity and implementation issues of the presented methods, discussing advantages and disadvantages of our proposals with respect to previous works, as well as some possible extensions for the anomaly isolation algorithm.

6.1. Complexity Analysis

Numerical complexity and memory storage are central issues for on-line anomaly detection. Most of previous works on network-wide anomaly detection have conceived methods for off-line detection [25, 33], mining anomalies in large snapshots of data rather than treating every single measurement sequentially. These methods can be used for diagnosis of volume anomalies after their occurrence, but are rather useless for an ISP if anomaly mitigation or any other kind of countermeasure is the objective. On the contrary, our both SB methods can be used for on-line anomaly detection, and thus we should assess their complexity. Let us compare the numerical complexity of these algorithms against those used for comparison in Section 5.4, the KB method and the SPCA approach.

The OSBD method stores two matrices in memory, the matrix $\Phi^{-\frac{1}{2}}$, with $\Phi = R\Sigma R^T$, and the projection matrix $P_H^{\perp} = I_r - H\left(H^T H\right)^{-1} H^T$, with $H = \Phi^{-\frac{1}{2}} RS$. This represents a total of $3r^2/2$ variables ($P_H^{\perp}$ is symmetric), where r is the number of links in the network. The computation of $\Phi^{-\frac{1}{2}}$ and $P_H^{\perp}$ involves matrix multiplications and inversions, and thus the associated cost is $\mathcal{O}(r^3)$. There is an additional cost in the learning phase of the spline-based methods, related to the tomogravity estimate used to design the splines basis S. The cost of the tomogravity method is similar to that of the least-squares method, which implies $\mathcal{O}(m^3)$ operations to estimate an $m \times 1$ vector. All these matrices are computed off-line during the learning phase and do not affect the scalability and on-line applicability of the method. The on-line application involves three consecutive operations at every time t: the whitening of the SNMP measurements vector $Z_t = \Phi^{-\frac{1}{2}} Y_t$, the projection of the obtained vector onto the left null space of H, and the computation of the norm of this projection. All these operations have a complexity $\mathcal{O}(r^2)$.

Memory usage is similar in the case of the SSB method. The matrix $\Phi^{-\frac{1}{2}}$ is also stored, but instead of saving the projection matrix $P_H^{\perp}$ the rejector W is kept in memory, built from the first $r - q$ eigenvectors of $P_H^{\perp}$. Given the recursive structure of the SSB method, m additional variables are kept in memory, which corresponds to the m recursive functions $g_t(i, 0)$, $i = 1, \dots, m$. For anomaly isolation purposes, the m anomaly signatures $\mathbf{v}_k \in \mathbb{R}^{(r-q)\times 1}$ are also stored. The singular value decomposition (SVD) of $P_H^{\perp}$ has a computation complexity of $\mathcal{O}(r^3)$, and as before, the construction of the splines basis involves $\mathcal{O}(m^3)$ operations. In the on-line detection/isolation phase, residuals $U_t = W Z_t$ are firstly computed and then used to update the m recursive functions $g_t(i, 0)$ according to (15) and (16). Finally, the m functions $s_t(i)$ used for anomaly isolation are computed according to (14). These steps involve approximately $\mathcal{O}(r^2)$ operations for anomaly detection and $\mathcal{O}(m^2)$ additional operations for anomaly isolation.

The SPCA method keeps the symmetric projection matrix $\mathbf{C} = (I_r - \mathbf{P}\mathbf{P}^T)$ in memory, which accounts for $r^2/2$ variables. The anomaly isolation in the SPCA method consists of a greedy search for a particular anomaly signature, each represented by a normalized column of the routing matrix $\mathbf{r}_k \in \mathbb{R}^{r \times 1}$ that must also be saved in memory. The construction of $\mathbf{P}\mathbf{P}^T$ relies on computing the SVD of the SNMP measurements matrix $\mathbf{Y} \in \mathbb{R}^{n \times r}$, where n is the number of consecutive SNMP measurements considered, a number usually much bigger than r; for example, $n = 1008$ and $r = 49$ in [25]. This SVD has a numerical complexity of $\mathcal{O}(nr^2)$. The use of the SPCA for anomaly detection involves the projection of the SNMP measurements vector onto the anomaly subspace and the computation of the norm of this projection, with a numerical complexity of $\mathcal{O}(r^2)$. As regards anomaly isolation, the greedy search consists of constructing m possible anomaly explanations (with a cost of $\mathcal{O}(r^2)$ operations each), thus additionally adding $\mathcal{O}(mr^2)$ operations.

Finally, the KB method complexity corresponds to that of the standard Kalman filter recursive equations. We refer the reader to the original paper of the KB method [31] for additional details. The method must store in memory an $m \times m$ state transition diagonal matrix that models the evolution of the anomaly-free traffic matrix, the routing matrix R, and the noise covariance matrices associated with the observation and the evolution processes; the latter is also a diagonal matrix. This accounts for a total of $2(r^2 + m)$ variables in memory. The recursive nature of the Kalman filter implies to keep in memory two additional matrices, the $m \times r$ Kalman gain matrix and the $m \times m$ prediction error covariance matrix. The learning process of the KB method consists of a recursive Expectation Maximization (EM) approach. There are many different EM algorithms, but in all cases the resolution involves matrix operations with a numerical complexity of $\mathcal{O}(m^3)$ for the estimation of an $m \times 1$ vector. The use of the KB for on-line anomaly detection implies to update the Kalman gain, the estimation covariance error and the residual error. This involves matrix multiplications and inversions, and thus the associated cost is $\mathcal{O}(m^3)$.

Method	n° vars. mem.	n° ops. learn	n° ops. on-line	
OSBD	$\mathcal{O}(r^2)$	$\mathcal{O}(m^3)$	$\mathcal{O}(r^2)$	n/a
SSB	$\mathcal{O}(mr)$	$\mathcal{O}(m^3)$	$\mathcal{O}(r^2)$	$\mathcal{O}(m^2)$
SPCA	$\mathcal{O}(mr)$	$\mathcal{O}(nr^2)$	$\mathcal{O}(r^2)$	$\mathcal{O}(mr^2)$
KB	$\mathcal{O}(m^2)$	$\mathcal{O}(m^3)$	$\mathcal{O}(m^3)$	n/a

Table 3: Numerical complexity and memory usage for different on-line anomaly detection algorithms. On-line operations are divided into detection operations and isolation operations.

Table 3 builds a raw summary of the numerical complexity and memory storage restrictions for the algorithms discussed above. Memory usage is similar in all cases, with a slightly higher requirement for the KB approach. While the SPCA method works with an $n \times r$ matrix in the learning phase, the SB and the KB methods use $m \times m$ matrices and thus they require more operations for learning issues. As regards on-line applicability, we see that the KB method is largely more expensive than the rest of the algorithms for anomaly detection, which comes directly from using the Kalman filter with large matrices. Finally, anomaly isolation involves a similar number of operations for the SSB and the SPCA methods. The important conclusion that can be drawn from table 3 is that the SB algorithms that we propose in this work have both similar or even smaller numerical complexity for on-line anomaly detection/isolation than those proposed to date.

6.2. Implementation Issues

We shall now discuss some important issues related to a real implementation of the proposed algorithms in a large-scale operational network. Table 4 presents a comparative analysis of some implementation-significant features between the SB algorithms, the KB method, and the SPCA method. Let us discuss each of the compared items.

All the methods use SNMP measurements as input data for anomaly detection, making it possible, at least a priori, to detect volume anomalies in OD flows without necessity of direct flow monitoring technology. This is a key feature regarding the development of light monitoring systems. However, the KB method needs anomaly-free (A-F) direct flow measurements for calibration purposes, loosing this advantage. The learning data for the SB methods consists of anomaly-free SNMP measurements, while the SPCA method uses SNMP measurements not necessarily free of anomalies for calibration (collected "raw" data). There is a major difference in the duration of the learning step, which we will see has important consequences. As we have shown in the evaluation section, the SB methods just need one hour of SNMP measurements to achieve reliable results. The KB method uses 24 hours of OD flow measurements to calibrate the underlying anomaly-free-traffic model, and the SPCA method uses as much as 1 week of SNMP measurements to build the normal and anomalous subspaces. The use of raw SNMP measurements in the SPCA approach is certainly useful, but as it has already been shown in previous works [29, 28], there is an undeniable associated risk of learning contamination, which is definitely magnified by the lengthly learning step. The remarkably short learning step of the SB methods makes it easy for network operators to calibrate the underlying spline-based model without risks of contamination, as it is quite easy to collect 1 hour of SNMP measurements free of volume anomalies.

The assumptions involved in deriving the SB anomaly-free-traffic model are quite strong with respect to the rest of the algorithms. Nevertheless, the validation of the SB model in three different large-scale networks shows that these assumptions are correctly verified in quite different network topologies and traffic scenarios (commercial traffic as well as research-oriented traffic). The KB method makes little assumptions on the underlying traffic model and assumes the classical Kalman filter hypotheses to be correctly verified. In practice, the Kalman filter is well-known for being robust to model imprecisions, and thus we claim that the KB assumptions are weak. The SPCA method is a pure data-driven method and makes no assumptions about traffic characteristics. However and as it is pointed out in [29], there are quite significant assumptions in the heuristics used for anomaly isolation that have no a-priori justification and can unduly trigger alarms in some OD flows much more frequently than others.

The numerical complexity analysis previously performed shows that both SB methods as well as the SPCA method are easily scalable with the size of the network, while poor scalability can be expected from the KB method.

There is no discussion about the impacts of routing modifications over the SPCA method in the former papers [23, 25] and a constant routing matrix is used, both in the theoretical development and in the evaluation. The authors of [31] claim that the KB method can be easily extended to work with time-varying routing matrices, but no discussion is provided on the involved challenges and current proposal does not support dynamic routing. The main challenge with routing modifications is that intradomain routing modifications can modify the incoming OD traffic flows distribution due to interdomain traffic shifts. In fact, it is well known that hot potato routing can induce interdomain routing changes due to intradomain routing modifications. In this sense, all algorithms must be recalibrated when an intradomain routing modification occurs, and the only methods that have a learning period length in the time scale of a routing modification are the SB methods, thus we claim that the SB anomaly detection methods can partially support routing modifications.

A similar analysis can be done regarding the application of the methods to non-stationary OD flows. Non-stationarities in traffic flows may render the underlying anomaly-free-traffic model non-longer adecuate, motivating a model recalibration. The key issue is how to detect when a new recalibration must be done. In [7], authors propose a very simple heuristic to achieve this task for the underlying models of the SPCA and KB methods. The idea is to monitor the *innovation process* i_t of the traffic model, namely the difference between the measured SNMP link counts Y_t and the link counts obtained from the estimated TM, namely $\hat{Y}_t = R\hat{X}_t$. The decision rule is straightforward: if the innovation process is above certain threshold, a recalibration is triggered. To avoid unnecessary and expensive recalibrations due to short-lived volume anomalies, authors propose to monitor i_t during periods of 24 hours, and only perform a new calibration if i_t has exceeded the threshold more than some fraction of the time. A similar heuristic could be directly applied to the SB methods. However, there are some clear drawbacks of this approach. The first problem is related to long-lived anomalies, which may not be filtered even with a 24hs window of measurements. In fact, in this case it is not possible to distinguish between an anomaly and a model that has drifted. The second problem is that the recalibration could come many hours late, seriously affecting the performance of the detection algorithm. Our SB methods have once again the lead in this subject, due to the short and "cheap" (SNMP-based) learning period of the underlying model. A very simple heuristic to avoid drifting from an accurate model would be to proceed in a similar way to Section 5.3: simply recalibrate the model if no anomaly has been declared in the last hour. Evaluations about the temporal stability of the SB model showed that this is not necessary even for several consecutive days in the real datasets that we used. Even so, we have shown that if necessary, our method can effectively be recalibrated every hour, and thus we claim that the SB anomaly detection methods support non-stationary traffic.

The last item we discuss concerns missing data; all algorithms use SNMP measurements as input, which has known practical limitations due to missing data and synchronization problems when collecting SNMP readings network-wide. In fact, the simultaneous collection of SNMP readings is practically impossible in very large-scale networks. The SPCA and the SB methods assume temporal independence between consecutive SNMP measurements, and thus the only impact that missing data has is a delayed verdict. In practice, it is easy to verify that all SNMP router readings are available at time t before applying the detection/isolation tests; in case there are missing readings at time t, the methods have to delay the analysis until the following time step where data is complete. As regards desynchronized readings, the problem is similar to

Feature Under Comparison	SB	KB	SPCA
Input Data	SNMP	SNMP	SNMP
Learning Data	SNMP A-F	TM A-F	SNMP
Learning Period Length	1 hs.	24 hs.	1 week
Assumptions	strong	weak	significant
Scalability	yes	poor	yes
Dynamic Routing	partially	no	no
Non-Stationary Traffic	yes	partially	partially
Missing Data	yes	no	yes

Table 4: Implementation issues in on-line anomaly detection/isolation.

missing data, and the best the algorithms can do is to delayed the analysis as before. Both problems condition the smallest feasible time scale on which the proposed methods might be used, but this is an implementation issue that depends on the particular network and thus it is impossible to give an order of this smallest time-scale. A possible solution to alleviate the problem of missing and desynchronized SNMP readings is to use oversampling: oversampling is commonly used in signal processing to reduce the effect of noisy measurements. However, this analysis is beyond the scope of current paper. As regards the KB method, it strongly relies on the temporal dependence between consecutive SNMP measurements, and thus it can be heavily influenced by missing data. The Kalman filter can be modified so as to cope with missing data, but current KB implementation [31] does not support this practical limitation.

6.3. Multiple Anomaly Isolation

To conclude with the discussion section, we propose some possible extensions to the presented anomaly isolation algorithm. In this paper we have assumed the same simplifying hypothesis as in [25], considering only "localized" anomalies, namely anomalies in a single OD flow at a time. However, the isolation algorithm can be extended, at least in theory, to identify multiple consecutive OD flow volume anomalies. The multiple hypotheses $\mathcal{H}^k_{t_0}$ in (13) can be rewritten so as to consider multiple combinations of consecutive anomalous OD flows as additional hypotheses to test. For example, suppose that we want to detect single OD flow volume anomalies as well as volume anomalies that span two OD flows at the same time. In this case, we have to add to $\mathcal{H}^k_{t_0}$ all the hypotheses that consider a volume anomaly at OD flow i and at OD flow j at the same time, for $0 \leqslant i \neq j \leqslant m$. This accounts for $\mathcal{C}^m_2 = m!/2!(m-2)! \approx m^2/2$ additional hypotheses to test. In this case, the set of anomaly signatures is composed not only by the m single normalized columns of the routing matrix $\mathbf{r}_k$, $k = 1..m$ but also by $\mathcal{C}^m_2$ matrices that include the two normalized columns of the routing matrix associated with the two anomalous OD flows. This procedure is the same as the one discussed in [25], but the idea comes from the former work of the PCA approach for fault diagnosis [21]. The problem with this approach is that the number of hypotheses to deal with, and consequently the number of decision functions $s_t(i)$ to compute grows highly and becomes very difficult to manage in a practical implementation. It is important to stress that the PCA approach [21, 25] suffers from exactly the same problem as regards anomaly isolation, as the heuristics employed have a numerical complexity in the same order as our methods. The isolation of multiple consecutive anomalous OD flows is out of the scope of this paper.

7. Conclusions

In this paper we have addressed the problem of network-wide volume anomaly detection and isolation in large-scale IP networks. The following list highlights the main characteristics of the proposed solution and our major contributions to the field:

(1) Presented methods rely on coarse-grained, easily available SNMP data to detect and isolate volume anomalies in traffic OD flows. This is a main advantage in order to develop light monitoring systems without the necessity of direct flow measurement technology, particularly in the advent of the forecast massive traffic to analyze in the near future.

(2) We have introduced an original linear, parsimonious, spline-based traffic model to describe the anomaly-free behavior of the traffic in a large-scale IP network. This spatial traffic model has several applications and advantages with respect to previous traffic matrix models: (i) being parsimonious by conception, it allows to solve the fundamentally ill-posed nature of the traffic matrix estimation problem from link SNMP measurements; (ii) it is non-data-driven and as we have verified through extensive evaluation with real data, it remains stable in time, at least for several days; (iii) the model is easy to calibrate and needs a very small amount of anomaly-free data to provide reliable results; and most importantly, (iv) this parsimonious parametric model makes it possible to remove the anomaly-free traffic from the anomaly detection problem, motivating our original approach of treating the detection and isolation of volume anomalies as a statistical change detection/isolation problem with a nuisance parameter. This a-priori simple characteristic allows to construct optimal algorithms for volume anomaly detection and isolation.

(3) We have developed different methods for volume anomaly detection and isolation with a paramount advantage with respect to previous works in the field, that of having solid optimality properties in terms of detection mean delay, false alarm rate and false isolation rate. This represents a major breakthrough in the field and the most important contribution of the paper. We argue that optimality support is fundamental in the conception of general algorithms, not tied to any particular network or evaluation.

(4) Using extensive data from three real backbone networks we have shown that the theoretical optimality properties of the proposed algorithms are verified in practice, providing results that outperform current network-wide anomaly detection/isolation methods in a wide variety of network topologies and traffic scenarios.

(5) The complexity analysis has shown that our algorithms are more efficient than current methods to perform anomaly detection and isolation in real time with even better results. We believe that a real implementation of our optimal algorithms could be envisaged without any modifications to current technology.

It is worth noting that the presented approaches can be easily extended to the detection and isolation of more general traffic anomalies, provided that a statistical parametric model is available. We expect that the proposed solutions in this work will stimulate in the future the development of anomaly detection algorithms with a solid theoretical background, allowing a robust growth of the network monitoring field. We believe that the results of decision theory applied to the field of network monitoring are still not sufficient and worthy to extend. This paper contributes to bridging the gap between these two fields.

Acknowledgements

We would like to thank the anonymous reviewers for their detailed comments that helped to improve the paper.

References

[1] Cisco Systems, "Global IP Traffic Forecast and Methodology, 2006-2011", white paper available at http://www.cisco.com, 2007 - updated 2008.

[2] Cisco Systems, "The Exabyte Era", white paper available at http://www.cisco.com, 2007 - updated 2008.

[3] I. Cunha, F. Silveira, R. Oliveira, R. Teixeira, and C. Diot, "Uncovering Artifacts of Flow Measurement Tools", in *Proc. of Passive and Active Measurement Conference*, 2009.

[4] P. Casas, L. Fillatre and S. Vaton, "Multi Hour Robust Routing and Fast Load Change Detection for Traffic Engineering", in *Proc. IEEE ICC*, 2008.

[5] Y. Zhang, M. Roughan, N. Duffield and A. Greenberg, "Fast Accurate Computation of Large-Scale IP Traffic Matrices from Link Load Measurements", in *Proc. ACM SIGMETRICS*, 2003.

[6] A. Soule, K. Salamatian, A. Nucci, and N. Taft,"Traffic Matrix Tracking using Kalman Filters", in *LSNI*, 2005.

[7] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, and C. Diot,"Traffic Matrices: Balancing Measurements, Inference, and Modeling", in *Proc. ACM SIGMETRICS*, 2005.

[8] P. Casas, S. Vaton, L. Fillatre, and T. Chonavel, "Efficient Methods for Traffic Matrix Modeling and On-line Estimation in Large-Scale IP Networks", in *Proc. ITC21*, 2009.

[9] C. Hood and C. Ji, "Proactive network fault detection", in *Proc. INFOCOM*, 1997.

[10] I. Katzela and M. Schwartz, "Schemes for fault identification in communications networks", in *IEEE/ACM Trans. on Networking*, vol. 3, no. 6, pp. 753-764, 1995.

[11] A. Ward, P. Glynn and K. Richardson, "Internet service performance failure detection", in *Performance Evaluation Review*, 1998.

[12] J. Jung, B. Krishnamurthy and M. Rabinovich, "Flash crowds and denial of service attacks: Characterization and implications for CDNs and webs", in *WWW-02*, 2002.

[13] L. Xie et al., "From Detection to Remediation: A Self-Organized System for Addressing Flash Crowd Problems", in *Proc. IEEE ICC-08*, 2008.

[14] P. Barford, J. Kline, D. Plonka, and A. Ron, "A Signal Analysis of Network Traffic Anomalies", in *SIGCOMM Internet Measurement Workshop*, 2002.

[15] J. Brutlag, "Aberrant Behavior Detection in Time Series for Network Monitoring", in *Proc. 14th Systems Administration Conference*, 2000.

[16] C.M. Cheng, H. Kung, K.S. Tan, "Use of Spectral Analysis in Defense Against DoS Attacks", in *Proc. IEEE GLOBECOM*, 2002.

[17] B. Krishnamurthy, S. Sen, Y. Zhang, and Y. Chen, "Sketch-based Change Detection: Methods, Evaluation, and Applications", in *Proc. USENIX/ACM IMC*, 2003.

[18] C.C. Zou, W. Gong, D. Towsley and L. Gao "The Monitoring and Early Detection of Internet Worms", in *IEEE/ACM Trans. on Networking*, vol. 13, n. 5, pp. 961-974, 2005.

[19] H. Wang, D. Zhang and K.Shin, "Detecting SYN flooding attacks", in *Proc. IEEE INFOCOM'02*, 2002.

[20] M. Roughan, T. Griffin, M. Mao, A. Greenberg, and B. Freeman, "Combining Routing and Traffic Data for Detection of IP Forwarding Anomalies", in *ACM SIGCOMM NeTs Workshop*, 2004.

[21] R. Dunia and S. J. Qin, "Multi-Dimensional Fault Diagnosis Using a Subspace Approach", in *Proc. American Control Conference*, 1997.

[22] R. Dunia and S. J. Qin, "A Subspace Approach to Multidimensional Fault Identification and Reconstruction", in *American Institute of Chemical Engineers Journal*, pp. 1813-1831, 1998.

[23] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. Kolaczyk, and N. Taft, "Structural Analysis of Network Traffic Flows", in *Proc. ACM SIGMETICS*, 2004.

[24] A. Lakhina, M. Crovella, and C. Diot, "Characterization of Network-Wide Anomalies in Traffic Flows", in *Proc. USENIX/ACM IMC*, 2004.

[25] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing Network-Wide Traffic Anomalies", in *Proc. ACM SIGCOMM*, 2004.

[26] A. Lakhina, M. Crovella and C. Diot, "Mining Anomalies Using Traffic Feature Distributions", in *Proc. ACM SIGCOMM*, 2005.

[27] X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone and A. Lakhina, "Detection and Identification of Network Anomalies Using Sketch Subspaces", in *Proc. USENIX/ACM IMC*, 2006.

[28] T. Ahmed, M. Coates, and A. Lakhina, "Multivariate Online Anomaly Detection Using Kernel Recursive Least Squares," in *Proc. IEEE Infocom*, 2007.

[29] H. Ringberg, A. Soule, J. Rexford and C. Diot, "Sensitivity of PCA for Traffic Anomaly Detection", in *Proc. ACM SIGMETRICS*, 2007.

[30] M. Thottan and C. Ji, "Anomaly Detection in IP Networks", in *IEEE Trans. on Signal Processing*, vol. 51, no. 8, pp. 2191-2204, 2003.

[31] A. Soule, K. Salamatian and N. Taft, "Combining Filtering and Statistical Methods for Anomaly Detection", in *Proc. USENIX/ACM IMC*, 2005.

[32] A. Tartakovsky *et al.*, "A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods", in *IEEE Trans. on Signal Processing*, vol. 54, no. 9, pp. 3372-3382, 2006.

[33] Y. Zhang, Z. Ge, A. Greenberg, and M. Roughan, "Network Anomography", in *Proc. USENIX/ACM IMC*, 2005.

[34] G. Nürnberger, "Approximation by spline functions", Springer-Verlag, 1989.

[35] C. Rao, "Linear Statistical Inference and its Applications", John Wiley & Sons, 1973.

[36] E. Lehman, "Testing Statistical Hypotheses, Second Edition", Chapman & Hall, 1986.

[37] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large", in *Trans. American Math. Soc.*, vol. 54, pp. 426-482, 1943.

[38] Fouladirad M. and I. Nikiforov (2005), Optimal statistical fault detection with nuisance parameters. Automatica, 2005, vol. 41, p. 1157- 1171.

[39] L. Fillatre and I. Nikiforov, "Non-bayesian detection and detectability of anomalies from a few noisy tomographic projections", in *IEEE Trans. on Signal Processing*, vol. 55, no. 2, pp. 401-413, 2007.

[40] M. Basseville and I. Nikiforov, "Detection of abrupt changes: theory and applications", Prentice Hall, 1993.

[41] I. Nikiforov, "A generalized change detection problem", in *IEEE Trans. on IT*, vol. 41, pp. 171-187, 1995.

[42] T. Oskiper and H. Poor, "Online activity detection in a multiuser environment using the matrix CUSUM algorithm", in *IEEE Trans. on IT*, vol. 48, pp. 477-493, 2002.

[43] I. Nikiforov, "A simple recursive algorithm for diagnosis of abrupt changes in random signals", in *IEEE Trans. on IT*, vol. 46, no. 7, pp. 2740-2746, 2000.

[44] I. Nikiforov, "A lower bound for the detection/isolation delay in a class of sequential tests", in *IEEE Trans. on IT*, vol. 49, no. 11, pp. 3037-3046, 2003.

[45] Abilene Observatory, http://abilene.internet2.edu/observatory/

[46] Y. Zhang, "Abilene Data", http://www.cs.utexas.edu/~yzhang/

[47] S. Uhlig et al, “Providing Public Intradomain Traffic Matrices to the Research Community”, in ACM Sigcomm Computer Communication Review, 2006.

[48] TOTEM Toolbox, http://totem.run.montefiore.ulg.ac.be/