Sequential Non-Bayesian Network Traffic Flows Anomaly Detection and Isolation

Lionel Fillatre[1], Igor Nikiforov[1], Sandrine Vaton[2], and Pedro Casas[2]

[1] Institut Charles Delaunay/LM2S, FRE CNRS 2848,
Université de Technologie de Troyes,
12 rue Marie Curie Troyes 10010 France
(e-mail: firstname.lastname@utt.fr)

[2] Computer Science Department, TELECOM Bretagne
Technopôle Brest-Iroise - CS 83818 - 29238, Brest, France
(e-mail: firstname.lastname@telecom-bretagne.eu)

Abstract. Sequential detection (and isolation) of unusual and significant changes in network Origin-Destination (OD) traffic volumes from simple link load measurements is considered in the paper. The ambient traffic, i.e. the OD traffic matrix corresponding to the non-anomalous network state, is unknown and it is considered here as a nuisance parameter because it can mask the anomalies. Since the OD traffic matrix is not recoverable from the simple link load measurements, the anomaly detection is an ill-posed decision-making problem. The method discussed in this paper consists of finding a linear parsimonious model of ambient traffic (nuisance parameter) and detecting/isolating anomalies by using an invariant decision algorithm. An optimal sequential algorithm has been discussed in our previous publication, the main goal of the present paper is to discuss a simple "snapshot" algorithm based on the last vector of observations.

Keywords. Statistical change detection/isolation, nuisance parameters, network traffic flows, parsimonious model, invariant decision algorithm.

1 Introduction and motivation

The traffic demand over a network is typically described by a traffic matrix that captures the amount of traffic transmitted between every pair of ingress and egress nodes in a network, also called the Origin-Destination (OD) flows. A volume anomaly is a sudden change in an OD flow's traffic (for example, due to denial-of-service, viruses/worms, external routing reconfigurations, etc.) that spans multiple physical links of the network. The reliable detection/isolation of these unusual and significant changes in the OD traffic matrix is an important issue for network operation. The direct measurements (flow-level data) need high hardware requirements to be collected and processed network-wide (Coates et al., 2002). Consequently, the Simple Network Management Protocol (SNMP), which is a widely deployed standardized protocol, is preferred in practice to measure link loads and obtain some information on the traffic matrix. The challenge lies in the ill-posed

nature of the problem: the number of unknown OD flows is much larger than the number of SNMP measurements (Ahmed et al., 2007).

Several approaches are proposed to remedy this problem. The first group of methods consists in detecting anomalies in SNMP measurements without taking in consideration the traffic matrix. Such methods typically use times series, ARIMA models among others, (Krishnamurthy et al. (2003); Thottan and Ji (2003); Tartakovsky et al. (2006a, 2006b); Ahmed et al. (2007)) to model the SNMP measurement evolution in time and to detect deviations. In Lakhina et al. (2004), the authors propose to decompose the SNMP measurements on a Principal Component Analysis (PCA) basis. These methods can detect anomalies by monitoring each link but they do not exploit the routing matrix, i.e. the linear mathematical relationship between the OD traffic matrix and the SNMP measurements.

The second group of methods (Cao J. *et al.* (2000); Coates *et al.* (2002); Tebaldi *et al.* (1998)) exploits this linear mathematical relationship. These approaches typically assume that the traffic matrix is well approximated by a known statistical model. Such a method requires a well known prior to be efficient, which is not always feasible in practice. In Zhang *et al.* (2005), the authors study a large number of methods based on several kinds of model for SNMP measurements (wavelets, PCA among others) and OD flows (ARIMA time series) to detect anomalies. A major drawback of these methods is the absence of theoretical results on the optimality of the studied methods with respect to the statistical hypotheses testing theory: no results are proposed about the design of an optimal test which maximises the probability to detect anomalies.

Finally the last group of methods consists in using the Kalman filtering technique (Soule *et al.* (2005)) to track the evolution of the traffic matrix in time and to detect changes in the OD flows. Strictly speaking, the ill-posed nature of the measurement model makes the Kalman filter not observable and the Kalman filtering efficiency strongly depends on the initialization, which is a serious limitation in practice.

The contributions of this study are the following. First, a parsimonious linear model of non-anomalous OD flow volumes ("ambient" traffic) is proposed. This model can be used in two ways, either to estimate the OD flow volumes or to eliminate the non-anomalous "ambient" traffic from the SNMP measurements in order to provide residuals sensitive to anomalies. Secondly, since a few anomaly-free SNMP measurements (at most one hour of measurements) is sufficient to obtain a reliable model of the OD flows, the proposed method is well adapted to the case of non stationary measurements and to dynamic routing which is a typical situation in practice. Finally, an optimal invariant detection algorithm is proposed to detect anomalies directly from SNMP measurements.

2 SNMP Measurements

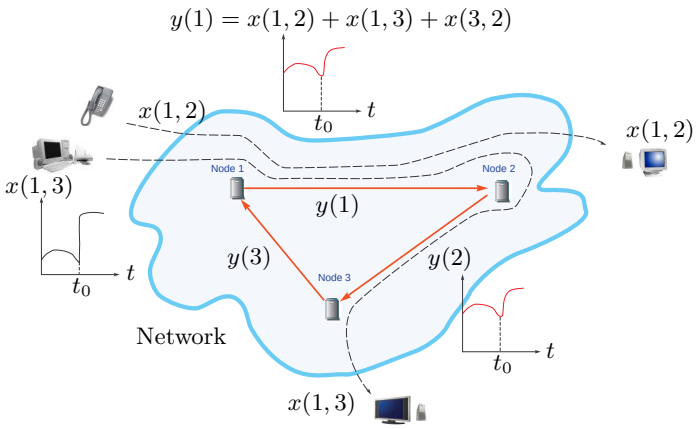

Fig. 1. Detection of unusual changes in OD traffic volumes. A step-wise anomaly appears at time t_0 in the OD flow $x(1,3)$ and it is routed on links $y(1)$ and $y(2)$.

Let us consider a network composed of r nodes and n monodirectional links (Lakhina et al. (2004); Zhang *et al.* (2005)) where $y(\ell)$ denotes the volume of traffic (typically in bytes) on the link ℓ at time t. To simplify the notations, the subscript t is omitted. The link volumes are provided by SNMP measurements. Let $x(i,j)$ be the OD traffic demand from node i to node j at time t. This situation is shown in Fig. 1. The traffic matrix $X = \{x(i,j)\}$ is reordered in the lexicographical order as a column vector $X = (x(1), \ldots, x(m))^T$ where X^T denotes the transpose of the matrix X and $m = r^2$ is the number of OD flows. Let us define an $n \times m$ routing matrix $A = (a(\ell,k))$ where $0 \leq a(\ell,k) \leq 1$ represents the fraction of OD flow k volume that is routed through link ℓ. This leads to the linear model $Y = A\,X$ where $Y = (y(1), \ldots, y(n))^T$. Without loss of generality, the known matrix A is assumed to be full row rank, $\mathrm{rank}\,(A) = n$.

3 Problem statement : volume anomaly detection

The problem consists in detecting (/isolating) a significant volume anomaly in an OD flow $x(i,j)$ by using only SNMP measurements $y(1), \ldots, y(n)$. For example, in Fig. 1, it is necessary to detect a sudden increase of the traffic volume $x(1,3)$ by using $y(1), y(2), y(3)$ (typical figures of anomaly to be detected in SNMP measurements are presented in Lakhina et al. (2004)). As it has been mentioned above, the main problem with the SNMP measurements is that $n \ll m$. To overcome this difficulty a parsimonious linear

model of non-anomalous traffic has been used. The derivation of this model includes two steps: *i)* description of the ambient traffic by using a spatial stationary model and *ii)* linear approximation of the model by using piecewise polynomial splines. The idea of the spline model is that the non-anomalous (ambient) traffic at each time t can be represented by using a known family of basis functions superimposed with unknown coefficients : i.e. $X_t \approx B\boldsymbol{\mu}_t$, where the $m \times q$ matrix B is assumed to be known and $\boldsymbol{\mu}_t \in \mathbb{R}^q$ is a vector of unknown coefficients such that $q < n$. Finally, it is assumed that the model residuals together with the natural variability of the OD flows follow a spatial Gaussian distribution, which leads to the following equation:

$$X_t = B\boldsymbol{\mu}_t + \boldsymbol{\xi}_t \tag{1}$$

where $\boldsymbol{\xi}_t \sim \mathcal{N}(0, \Sigma)$ is a Gaussian noise, with the $m \times m$ spatial diagonal covariance matrix $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$. The advantages of a parametric model are the following ones. First, a non-parametric basis, typically the PCA basis, can be used to generate the matrix B but this solution needs direct OD flow measurements (infeasible in practice) and the PCA basis depends on the period when the measurements are made (see Ringberg *et al.* (2007)). Secondly, the parametric detection method performances are better than the non-parametric ones provided that the adopted model is accurate enough. Hence, the link load measurement model is given by the following linear equation :

$$Y_t = A\,B\boldsymbol{\mu}_t + A\boldsymbol{\xi}_t = G\boldsymbol{\mu}_t + \boldsymbol{\zeta}_t + [\boldsymbol{\eta} = \text{possible anomaly}], \tag{2}$$

where $Y_t = (y(1), \ldots, y(n))^T$ and $\boldsymbol{\zeta}_t \sim \mathcal{N}(0, A\Sigma A^T)$. Without any loss of generality, the resulting matrix $G = A\,B$ is assumed to be full column rank. Typically, when an anomaly occurs on OD flow j at time t_0 (change-point), the vector $\boldsymbol{\eta}$ has the form $\boldsymbol{\eta} = \epsilon\,\mathbf{a}(j)$ where $\mathbf{a}(j)$ is the j-th normalized column of A and ϵ is the intensity of the anomaly. The goal is to detect the presence of an anomalous vector $\boldsymbol{\eta}$ not explicable by the ambient traffic model $X_t \approx B\boldsymbol{\mu}_t$ and to isolate the contaminated OD flow j.

The problem of the anomaly detection/isolation based on a sequential decision rule (Nikiforov (1995, 2000, 2003)) has been discussed in our previous publication (see Fillatre et *al.* (2007b)) for a time correlated noise sequence $(\boldsymbol{\zeta}_t)_{t\geq 1}$. In case of a strong autocorrelation in the residual process $(\boldsymbol{\zeta}_t)_{t\geq 1}$ and a large signal-to-noise ratio, the statistical characteristic of the snapshot algorithm is very close to the characteristic of the optimal sequential test because the mean detection delay is almost equal to one observation. For this reason let us consider now a simple and efficient detection scheme based on the step by step hypotheses testing sometimes called “snapshot”.

4 Snapshot detection : hypotheses testing

Since the matrix $\Phi = A\Sigma A^T$ is known, the testing problem consists of choosing between the two alternatives at time t:

$$\mathcal{H}_0 = \{Z_t \sim \mathcal{N}(\boldsymbol{\theta} + H\boldsymbol{\mu}_t, I_n);\ \boldsymbol{\theta} = 0,\ \boldsymbol{\mu}_t \in \mathbb{R}^q\} \tag{3}$$

$$\mathcal{H}_1 = \{Z_t \sim \mathcal{N}(\boldsymbol{\theta} + H\boldsymbol{\mu}_t, I_n);\ \boldsymbol{\theta} \neq 0,\ \boldsymbol{\mu}_t \in \mathbb{R}^q\}, \tag{4}$$

with $Z_t = \Phi^{-\frac{1}{2}} Y_t$, $H = \Phi^{-\frac{1}{2}} G$, $\boldsymbol{\theta} = \Phi^{-\frac{1}{2}}\boldsymbol{\eta}$, $\Phi^{-\frac{1}{2}}$ is the square-root matrix of Φ^{-1}, Φ^{-1} is the inverse of Φ and I_n is the identity matrix of size n. Here $\boldsymbol{\mu}_t$ is considered as a nuisance vector parameter since i) it is completely unknown, ii) it is of no interest for the system in charge to detect the anomaly $\boldsymbol{\theta}$ and iii) it can mask the anomalies.

Let $\mathcal{K}_\alpha = \{\phi : \sup_{\boldsymbol{\mu}_t \in \mathbb{R}^q} \Pr_{\boldsymbol{\theta}=0,\boldsymbol{\mu}_t}(\phi(Z_t) = \mathcal{H}_1) \leq \alpha\}$ be the class of tests $\phi : \mathbb{R}^n \mapsto \{\mathcal{H}_0, \mathcal{H}_1\}$ with upper-bounded maximum false alarm probability, where the probability $\Pr_{\boldsymbol{\theta},\boldsymbol{\mu}_t}$ stands for the vector of observations Z_t being generated by the distribution $\mathcal{N}(\boldsymbol{\theta} + H\boldsymbol{\mu}_t, I_n)$ and α is the prescribed probability of false alarm. The power function β is defined as the probability of detection: $\beta(\boldsymbol{\theta}; \boldsymbol{\mu}_t) = \Pr_{\boldsymbol{\theta}\neq 0,\boldsymbol{\mu}_t}(\phi(Z_t) = \mathcal{H}_1)$. The subtlety of the above mentioned hypotheses testing problem consists of choosing between $\mathcal{H}_0$ and $\mathcal{H}_1$ with the best possible performance indexes (α, β) while considering $\boldsymbol{\mu}_t$ as a nuisance parameter.

5 Anomaly detection methodology

It is easy to see that the problem remains invariant under the group of translations $G = \{g : g(Z) = Z + H\mathbf{c}\,,\ \mathbf{c} \in \mathbb{R}^q\}$. Let $\mathcal{S}$ be the family of surfaces $\mathcal{S} = \{S_c\ :\ c > 0\}$ with $S_c = \{\boldsymbol{\theta}\ :\ \|P_H^\perp \boldsymbol{\theta}\|_2^2 = c^2\}$. Then, it is shown (Fillatre and Nikiforov (2007a)) that the test

$$\phi^*(Z_t) = \begin{cases} \mathcal{H}_0 \ \text{ if } \ \Lambda(Z_t) = \|P_H^\perp Z_t\|_2^2 < \lambda_\alpha \\ \mathcal{H}_1 \ \text{ else} \end{cases}, \tag{5}$$

where $P_H^\perp = I_n - H(H^T H)^{-1} H^T$ and the threshold λ_α is chosen to satisfy the false alarm bound α, $\Pr_{\boldsymbol{\theta}=0,\boldsymbol{\mu}_t}(\Lambda(Z_t) \geq \lambda_\alpha) = \alpha$, is Uniformly Best Constantly Powerful (UBCP)[1] in the class $\mathcal{K}_\alpha$ over the family of surfaces $\mathcal{S}$. The statistics Λ is distributed according to the χ^2 law with $n - q$ degrees of freedom. This law is central under $\mathcal{H}_0$ and non-central under $\mathcal{H}_1$ with the non-centrality parameter $\boldsymbol{\theta}^T P_H^\perp \boldsymbol{\theta}$.

[1] A test $\phi^* \in \mathcal{K}_\alpha$ is UBCP on $\mathcal{S}$ if 1) $\beta_{\phi^*}(\boldsymbol{\theta}') = \beta_{\phi^*}(\boldsymbol{\theta}'')$, $\forall \boldsymbol{\theta}', \boldsymbol{\theta}'' \in S_c$; 2) $\beta_{\phi^*}(\boldsymbol{\theta}) \geq \beta_\phi(\boldsymbol{\theta})$, $\forall \boldsymbol{\theta} \in S_c$, $\forall c > 0$ for any test $\phi \in \mathcal{K}_\alpha$ which satisfies 1).

6 Conclusion

The results of numerical experiments with real data will be shown during the presentation of the paper. They confirm the quality of the proposed methods.

References

Ahmed, T., Coates, M. and Lakhina, A. (2007): Multivariate online anomaly detection using kernel recursive least squares, in *Infocom*.

Cao J. *et al.* (2000): Time-varying network tomography: router link data, *Journal of Americal Statistical Association*, vol. 95, no. 452, pp. 1063–1075.

Coates, M., Hero, A., Nowak, R. and Yu, B. (2002) Internet tomography, *IEEE Signal Processing Mag.*, May.

Fillatre, L. and Nikiforov, I. (2007a) Non-bayesian detection and detectability of anomalies from a few noisy tomographic projections, *IEEE Trans. Signal Processing*, vol. 55, no. 2, pp. 401–413.

Fillatre, L., Nikiforov, I. and Vaton, S. (2007b): Sequential Non-Bayesian Change Detection-Isolation and Its Application to the Network Traffic Flows Anomaly Detection, in Proceedings of the 56th Session of ISI, Lisboa, 22-29 August.

Krishnamurthy, B., Sen, S., Zhang, S. and Chen, Y. (2003): Sketch-based change detection: methods, evaluation, and applications, in *IMC*.

Lai, T. L. (2000) Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems, *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 595–608.

Lai, T. L. (2001) Sequential analysis : some classical problems and new challenges (with discussion), *Statistica Sinica*, vol. 11, pp. 303–408, 2001.

Lakhina, A. *et al.*, Diagnosing network-wide traffic anomalies, in *SIGCOMM*, 2004.

Nikiforov, I. (1995): A generalized change detection problem, *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 171–187.

Nikiforov, I. (2000): A simple recursive algorithm for diagnosis of abrupt changes in random signals, *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2740–2746.

Nikiforov, I. (2003): A lower bound for the detection/isolation delay in a class of sequential tests, *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 3037–3046.

Ringberg, H. *et al.* (2007) Sensitivy of PCA for traffic anomaly detection, in *SIGMETRICS*.

Soule, A.*et al.* (2005): Traffic matrices: balancing measurements, inference and modeling, in *SIGMETRICS*.

Tartakovsky, A., Rozovskii, B., Blažek, R. and Kim, H. (2006a): A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods, *IEEE Trans. Signal Processing*, vol. 54, no. 9, pp. 3372–3382.

Tartakovsky, A., Rozovskii, B., Blažek, R. and Kim, H. (2006b): Detection of intrusions in information systems by sequential change-points methods, *Statistical methodology*, vol. 3, no. 3, pp. 252–293, July.

Tebaldi, C. *et al.* (1998): Bayesian inference on network traffic using link count data," *J. Amer. Statist. Assoc.*, vol. 93, no. 442, pp. 557–576.

Thottan, M. and Ji, C. (2003): Anomaly detection in IP networks, *IEEE Trans. Signal Processing*, vol. 51, no. 8, pp. 2191–2204.

Zhang, Y. *et al.* (2005): Network anomography, in *IMC'05*.