PAPER Special Section on Information and Communication System Security

A Traffic Decomposition and Prediction Method for Detecting and Tracing Network-Wide Anomalies

Ping DU[†a)], **Shunji ABE**[†], **Yusheng JI**[†], ***Members***, **Seisho SATO**[††], ***Nonmember***, ***and*** **Makio ISHIGURO**[††], ***Member***

SUMMARY Traffic volume anomalies refer to apparently abrupt changes in the time series of traffic volume, which can propagate through the network. Detecting and tracing these anomalies is a critical and difficult task for network operators. In this paper, we first propose a traffic decomposition method, which decomposes the traffic into three components: the *trend component*, the *autoregressive (AR) component*, and the *noise component*. A traffic volume anomaly is detected when the AR component is outside the prediction band for multiple links simultaneously. Then, the anomaly is traced using the projection of the detection result matrices for the observed links which are selected by a shortest-path-first algorithm. Finally, we validate our detection and tracing method by using the real traffic data from the third-generation Science Information Network (SINET3) and show the detected and traced results.

key words: *anomaly detection, anomaly tracing, autoregressive (AR) model, Kalman filter*

1. Introduction

As shown in Fig. 1, traffic volume anomalies refer to apparent spikes (marked by the red circles in Fig. 1) in the time series of the traffic data, which might be caused by flash crowds or attacks. Since anomalies can create congestion in a network, it is important for the network operators to detect when a volume anomaly is occurring and to trace its propagation path. Rapidly and accurately detecting and tracing anomalies are critical for the efficient operation of large computer networks.

How to detect the time points at which a network is experiencing anomalies is a complex task. There are two challenges in the detection task. One is that the average traffic volume varies over time in a day. In the example in Fig. 1, the average traffic volume in the work time (e.g., 2:00–5:00 pm) is more than thrice that of sleep time (e.g., 2:00–5:00 am). The varying traffic volume makes it difficult to detect anomalies using a simple threshold. Another is that it is difficult to discriminate anomalies from stochastic traffic fluctuations, especially on the links with a large volume of noisy traffic.

Several approaches have been proposed for anomaly detection. In [2], [3], wavelet analysis methods were proposed, in which the traffic was decomposed into low-band, mid-band, and high-band components, and anomalies were detected by using the threshold of the mid-band component of the traffic. The demerit is that a wavelet-based method is not suitable for real-time detection, but suitable for off-line analyses. In [4], [5], anomalies were detected by applying thresholds in time series model of the network traffic. The shortcoming of these methods is that it is difficult to find an exact model to forecast the real traffic. In [6], detection was done via a signature matching mechanism, in which an anomaly was detected when the feature vector of the current traffic matches a template anomaly feature vector generated based on the traffic history. This method is unable to detect an anomaly out of existing anomaly templates.

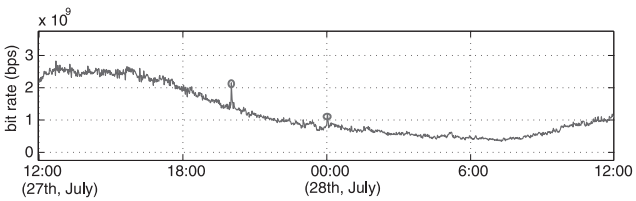

Fig. 1 Examples of traffic volume anomalies. (Traffic data were observed on the link from Tokyo1 to Nagoya over a 24-hr period starting from 12:00, July 27, 2007 in SINET3 [1]. The anomaly points are marked with red circles.)

Besides anomaly detection, to effectively control the anomalies, another task is to trace the sources of detected possible anomalies. Despite a number of the techniques that have been proposed for detecting traffic anomalies in a single-link [2]–[7], network-wide traffic anomalies remain poorly studied. To our knowledge, only A. Lakhina et al., [9] proposed a subspace separation method based on the Principle Component Analysis, by which the space of a whole set of network traffic were separated into different subspaces, representing the normal and anomalous traffic behaviors. Although it can diagnose anomalous behavior in any flow, it requires a high computational cost due to the high-dimensional matrix analysis and it is also suitable for real-time operation.

In this paper, we propose a traffic decomposition and AR prediction method to detect and trace anomalies. The method decomposes the traffic into three components: (1) the *trend component*, which captures the gradual changes in the traffic volume in a time series, (2) the *autoregressive (AR) component*, which consists of predictable anomalous traffic and stochastic traffic fluctuations, and (3) the *noise component*, which is assumed to be a white noise process with zero mean. A potential anomaly can be quickly detected when the AR component is outside the prediction

Manuscript received July 31, 2008.

[†]The authors are with National Institute of Informatics, Tokyo, 101–8430 Japan.

[††]The authors are with The Institute of Statistical Mathematics, Tokyo, 106–8569 Japan.

a) E-mail: duping@nii.ac.jp

DOI: 10.1587/transinf.E92.D.929

band. Furthermore, when taking into account that a traffic anomaly propagates through the network, simultaneous anomalies should be observed on more than one link when the traffic propagation delay is negligible. A volume spike on only one link will also be judged as a normal fluctuation of the traffic. When an anomaly originates in an ingress node of the backbone, it will propagate on the links along the route path obtained by a shortest-path-first algorithm. We project the detection result matrices on the time axis to diagnose which links and how many links are simultaneously experiencing anomalies. By marking the diagnosed results on the shortest route tree, the traffic anomalies propagation path can be traced.

The contributions of this paper are: (1) a general traffic decomposition and prediction method to detect anomalies, (2) a projection method to trace the possible anomalies in the backbone network, and (3) the validation of the detection and tracing methods using the real data from the SINET3 backbone network.

The paper is organized as follows. In Sect. 2, we introduce our traffic decomposition and prediction method for detecting anomalies. Section 3 introduces our tracing method. Section 4 evaluates the AR prediction method using a comparison with two other popular methods, the Exponential Weighted Moving Average (EWMA) and the Non-Seasonal Holt-Winters (NSHW) [4], [5] methods for traffic anomaly detection. Section 5 validates our proposed methods by using the real traffic data from the SINET3 network. Finally, Sect. 6 concludes this paper and outlines our future work.

2. Detecting Anomalies

The third-generation Science Information Network (SINET3) (shown in Fig. 2), which is our studied network in this paper, consists of 12 core nodes and more than 60 edge nodes to provide high-speed communication environment for more than 700 universities and research institutes. It also provides connectivity to other domestic networks such as JPIX and JPNAP, and to foreign networks such as those in New York and Los Angeles through international lines. The backbone capacity is a maximum of 40 Gbps. The traffic propagation delay in SINET3 is less than 1 ms, which is negligible compared to the sampling interval of the traffic volume, which is one minute in our research.

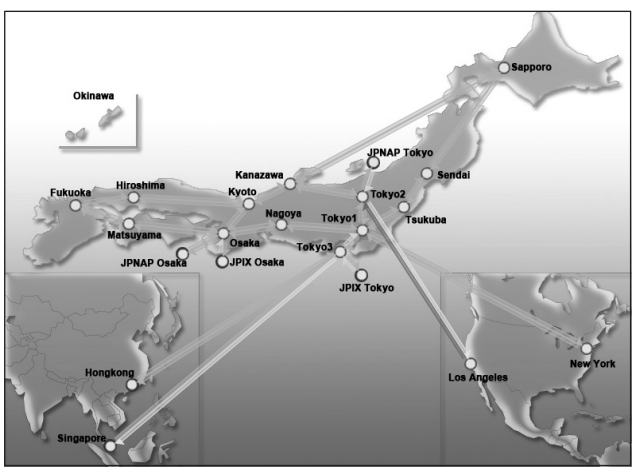

Fig. 2 Network topology of SINET3.

2.1 Traffic Decomposition

On July 26, 2007, we observed a set of anomalous points on the time series of the traffic from New York to core node Tokyo1 in SINET3. Since the sum of the input traffic equals the sum of the output traffic for each node, we might say that the anomalous traffic will also occur on some of the links out of node Tokyo1. The time series of the traffic on these links are shown in Fig. 3, in which "IR" and "IN" are the two edge nodes connected to node Tokyo1. As shown in Fig. 3, the mean traffic level varies considerably. Although the mean traffic level is very useful for predicting the gradual change tendency of the traffic volume, it is useless for analyzing the short-term changes, such as anomalies. We use a polynomial trend model of k order in this research to estimate the mean traffic volume. After removing the mean traffic, the residua in the time series can be modeled using the autoregressive (AR) process of order m (depicted as $AR(m)$) [11], [12].

Let y_n denote the observed traffic volume at time interval n. We can decompose it into three components: the trend component tr_n, the AR component p_n, and the noise component ω_n. They can be described as:

$$y_n = tr_n + p_n + \omega_n, \tag{1}$$

$$\Delta^k tr_n = \upsilon_{n1}, \tag{2}$$

$$p_n = \sum_{i=1}^{m} a_i p_{n-i} + \upsilon_{n2}, \tag{3}$$

where $\omega_n \sim N(0, \varsigma^2)$, $\upsilon_{n1} \sim N(0, \tau_1^2)$, and $\upsilon_{n2} \sim N(0, \tau_2^2)$, which are all white noise with a zero mean; and $\{a_i\}$ is the AR coefficient parameter. By replacing the difference operator Δ using backward shift operator B with $\Delta = I - B$ [11], Eq. (2) can be written as:

$$tr_n = \sum_{i=1}^{k} (-1)^{i+1} \binom{k}{i} tr_{n-k} + \upsilon_{n1}$$

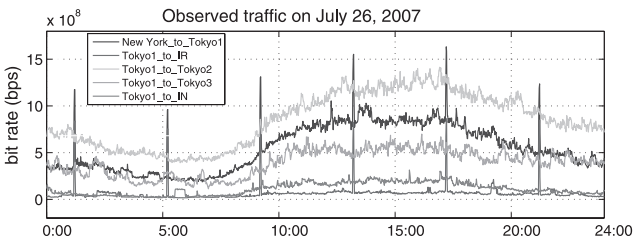

Fig. 3 Time series of traffic volumes on the links connected Tokyo1 on July 26, 2007.

$$= \sum_{i=1}^{k} c_k tr_{n-k} + \upsilon_{n1}, \tag{4}$$

where $c_k = (-1)^{i+1}\binom{k}{i}$.

We define a $(k + m - 1)$-dimensional vector $x_n = (tr_n, \cdots, tr_{n-k+1}, p_n, p_{n-1}, \cdots, p_{n-m+1})^t$ to denote the state at time interval n. The state-system described in Eqs. (1), (3), and (4) can be depicted as state-space Eq. (5), which consists of two models: the first describing the evolution of the state as a hidden Markovian process and the second is of the form state plus noise.

$$\begin{aligned} x_n &= \mathbf{F}x_{n-1} + \mathbf{G}\upsilon_n, \\ y_n &= \mathbf{H}x_n + \omega_n, \end{aligned} \tag{5}$$

where

$$\begin{aligned} \mathbf{F} &= \begin{bmatrix} \mathbf{F_1} & 0 \\ 0 & \mathbf{F_2} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{G_1} & 0 \\ 0 & \mathbf{G_2} \end{bmatrix}, \\ \mathbf{H} &= \begin{bmatrix} \mathbf{H_1} & \mathbf{H_2} \end{bmatrix}, \quad \upsilon_n = \begin{bmatrix} \upsilon_{n1} & \upsilon_{n2} \end{bmatrix}, \end{aligned} \tag{6}$$

in which $\mathbf{F_1}$, $\mathbf{G_1}$, and $\mathbf{H_1}$ correspond to the trend component, and $\mathbf{F_2}$, $\mathbf{G_2}$, and $\mathbf{H_2}$ correspond to AR component. They are depicted as

$$\begin{aligned} \mathbf{F_1} &= \begin{bmatrix} c_1 & c_2 & \cdots & c_k \\ 1 & & & \\ & \ddots & & \\ & & 1 & \end{bmatrix}, \quad \mathbf{G_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \\ \mathbf{H_1} &= \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}, \end{aligned} \tag{7}$$

and

$$\begin{aligned} \mathbf{F_2} &= \begin{bmatrix} a_1 & a_2 & \cdots & a_m \\ 1 & & & \\ & \ddots & & \\ & & 1 & \end{bmatrix}, \quad \mathbf{G_2} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \\ \mathbf{H_2} &= \begin{bmatrix} 1 & 0 & \cdots 0 \end{bmatrix}. \end{aligned} \tag{8}$$

The state-space representation Eq. (5) can be fast calculated using Kalman filtering [12]. The Kalman filter has two distinct steps: prediction and estimation. The prediction step uses the linear least-squares estimator from the previous state x_{n-1} to estimate the current state x_n. In the estimation step, currently observed traffic y_n is used to refine this prediction for the current state x_n. Let V_n be the covariance of x_n. In the following equations, the notation $\hat{x}_{n|n-1}$ represents the prediction of x_n at time $n-1$ and $\hat{x}_{n|n}$ denotes the estimation of x_n at time n. Similarly, the notation $\hat{V}_{n|n-1}$ represents the prediction of V_n at time $n-1$ and $\hat{V}_{n|n}$ denotes the estimation of V_n at time n.

[Prediction Step]

$$\hat{x}_{n|n-1} = \mathbf{F}\hat{x}_{n-1|n-1}. \tag{9}$$

$$\hat{V}_{n|n-1} = \mathbf{F}\hat{V}_{n-1|n-1}\mathbf{F}^t + \mathbf{G}R\mathbf{G}^t, \tag{10}$$

where $R = \begin{bmatrix} \tau_1^2 & 0 \\ 0 & \tau_2^2 \end{bmatrix}$.

[Estimation Step]

$$K_n = \hat{V}_{n-1|n-1}\mathbf{H}^t(\mathbf{H}\hat{V}_{n-1|n-1}\mathbf{H}^t + \varsigma^2)^{-1}. \tag{11}$$

$$\hat{x}_{n|n} = \hat{x}_{n|n-1} + K_n(y_n - \mathbf{H}\hat{x}_{n|n-1}). \tag{12}$$

$$\hat{V}_{n|n} = (I - K_n\mathbf{H})\hat{V}_{n|n-1}. \tag{13}$$

Here, K_n is the Kalman gain matrix and Eq. (11) can be obtained by minimizing the conditional mean-squared estimation error $E[\| \hat{x}_{n|n} - \hat{x}_{n|n-1} \|^2]$.

The trend component order k and the AR component order m are selected based on the Akaike Information Criterion (AIC) [11] through experimentation, which is not addressed in this paper. In our research, the trend component order k and the AR component order m are set to 2 as a default. Figure 4 is an example of the traffic component decomposition. The observation results show that the AR component is a suitable metric to describe the short-term changes in the traffic volume when we study the anomalies.

2.2 Anomaly Detection on One Single Link

Let the time series $Y_n = \{y_1, \ldots, y_n\}$ denote the observed traffic history. We define $\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{O}_{1\times k} & \mathbf{H_2} \end{bmatrix}$, where $\mathbf{O}_{1\times k}$ is a $1 \times k$ zero matrix. Then, $p_n = \tilde{\mathbf{H}}x_n$. Let $\hat{p}_{n|n-1}$ be the mean prediction of p_n, $\hat{p}_{n|n}$ be the estimation of p_n at time n, and $\hat{d}_{n|n-1}$ be the predicted covariance of p_n at time $n-1$. They can be calculated as:

$$\begin{aligned} \hat{p}_{n|n-1} &= E(p_n|Y_{n-1}) \\ &= E(\tilde{\mathbf{H}}x_n|Y_{n-1}) \\ &= \tilde{\mathbf{H}}\hat{x}_{n|n-1}, \end{aligned} \tag{14}$$

$$\begin{aligned} \hat{d}_{n|n-1} &= Cov(p_n|Y_{n-1}) \\ &= Cov(\tilde{\mathbf{H}}x_n|Y_{n-1}) \\ &= \tilde{\mathbf{H}}Cov(x_n|Y_{n-1})\tilde{\mathbf{H}}^t \\ &= \tilde{\mathbf{H}}\hat{V}_{n|n-1}\tilde{\mathbf{H}}^t, \end{aligned} \tag{15}$$

and

$$\hat{p}_{n|n} = \tilde{\mathbf{H}}\hat{x}_{n|n}. \tag{16}$$

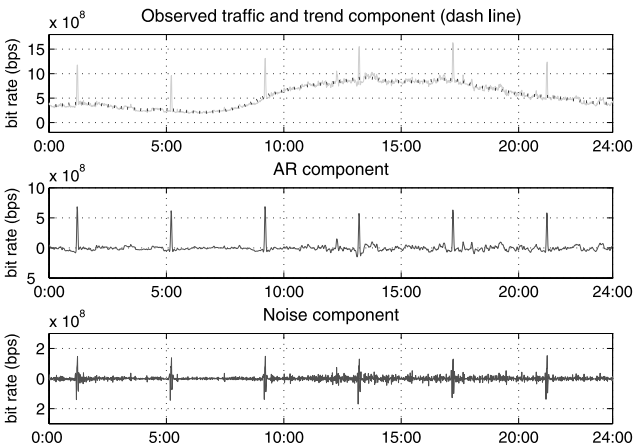

Fig. 4 Example of traffic decomposition. The observed traffic is on the link from New York to Tokyo1 on July 26, 2007.

We define $[\hat{p}_{n|n-1} - \beta \cdot \sqrt{\hat{d}_{n|n-1}}, \hat{p}_{n|n-1} + \beta \cdot \sqrt{\hat{d}_{n|n-1}}]$ as the prediction band of AR component p_n, where β is the scaling factor. When the prediction error $|\hat{p}_{n|n} - \hat{p}_{n|n-1}| \geq \beta \cdot \sqrt{\hat{d}_{n|n-1}}$ (in other words, $\hat{p}_{n|n}$ is outside this band), there is a possible anomaly.

An example of the prediction method is shown in Fig. 5, in which the cases of $\beta = 1.96$ and $\beta = 3$ are shown as solid and dash lines, respectively. It can be seen that the $\beta = 3$ line can catch our concerned major anomalies, while the $\beta = 1.96$ line can catch more abrupt change points. How to choose β is not only a tradeoff between the number of detected anomalies and false alarms, which will be discussed in more detail in Sect. 4, but also should be decided by the task of the network analysis. In the analysis of tracing anomaly that follows, we set $\beta = 3$ to catch and analyze major anomalies that are more likely to cause congestion.

Another interesting result can be seen in Fig. 5 is that the variance in the AR component is not changed over time after a short period of change from the start, so we can assume the AR process to be a stationary process and denote $\sqrt{\hat{d}_{n|n-1}}$ as σ. Furthermore, according to the Appendix, the AR component can be assumed to follow a Gaussian distribution. So, the anomaly detection rule can be described as

$$|\hat{p}_{n|n} - \hat{p}_{n|n-1}| \geq \delta_\alpha, \tag{17}$$

where δ_α denotes the threshold for the prediction error at the $1 - \alpha$ confidence level. The thresholds $\delta_{0.05} = 1.96\sigma$ ($\beta = 1.96$) and $\delta_{0.003} = 3\sigma$ ($\beta = 3$) can assure a $1 - \alpha = 0.95$ confidence level and a $1 - \alpha = 0.997$ confidence level, respectively.

2.3 Anomaly Detection on Multiple Links

Let l denote the number of observed links and t denote the number of successive time intervals of interest. Let matrices $\mathbf{Y}$ and $\mathbf{A}$ of size $l \times t$ denote the time series of observed traffic volumes and AR components respectively. In each of these two matrices, each column i denotes the time series of the ith link and each row j represents an instance of all observed links at time j.

We define $\hat{A}_{ij|j-1}$ as the estimator of A_{ij}. Accordingly, the prediction band is calculated as $[\hat{A}_{ij|j-1} - \beta\sigma_i, \hat{A}_{ij|j-1} + \beta\sigma_i]$, where σ_i is the standard deviation of the AR component of link i. When A_{ij} is outside this prediction band, there is a possible anomaly. We define two $l \times t$ matrices, $\mathbf{P}$ and $\mathbf{Q}$, to record the detected results:

$$\mathbf{P} = \begin{bmatrix} P_{11} & \cdots & P_{1t} \\ \vdots & \ddots & \vdots \\ P_{l1} & \ldots & P_{lt} \end{bmatrix},$$
$$\mathbf{Q} = \begin{bmatrix} Q_{11} & \cdots & Q_{1t} \\ \vdots & \ddots & \vdots \\ Q_{l1} & \ldots & Q_{lt} \end{bmatrix}, \tag{18}$$

in which

$$P_{ij} = \begin{cases} 1, & \text{if } |A_{ij} - \hat{A}_{ij|j-1}| \geq \beta\sigma_i \\ 0, & \text{otherwise} \end{cases},$$
$$Q_{ij} = \begin{cases} 2^{i-1}, & \text{if } |A_{ij} - \hat{A}_{ij|j-1}| \geq \beta\sigma_i \\ 0, & \text{otherwise} \end{cases}. \tag{19}$$

In addition, considering that a traffic volume anomaly propagates through the network, one anomaly should be observed on more than one link at the same time. A volume spike occurring only on one link will also be judged as an occasional fluctuation of traffic. Projection of matrices $\mathbf{P}$ and $\mathbf{Q}$ on the time axis is an effective technique for detecting whether one anomaly is observed on multiple links at the same time. We define two vectors, p and q, to represent the projected matrices $\mathbf{P}$ and $\mathbf{Q}$ on the time axis. They can be calculated using the following equation.

$$p = \left[\sum_i P_{i1} \quad \cdots \quad \sum_i P_{it} \right],$$
$$q = \left[\sum_i Q_{i1} \quad \cdots \quad \sum_i Q_{it} \right]. \tag{20}$$

When p_t equals n, it indicates that there is an anomaly occurring at n links at time t simultaneously. Next, we denote q_t into the binary system to judge which links are experiencing an anomaly. If the ith bit of q_t equals 1, this indicates that there is an anomaly on the ith link at time t; otherwise there is no anomaly. For example, if $q_t = (3)_{10} = (0011)_2$, this indicates that the first and second links have an anomaly at time t.

We apply the detection method to the traffic links shown in Fig. 3, and the projections of the detected matrices $\mathbf{P}$ and $\mathbf{Q}$ are shown in Fig. 6. Here we set $\beta = 3$ for our purposes. In the figure, at the points marked by the red circles, $p = 2$ indicates there are two links experiencing anomalies at those points. Furthermore, $q = (3)_{10} = (00011)_2$

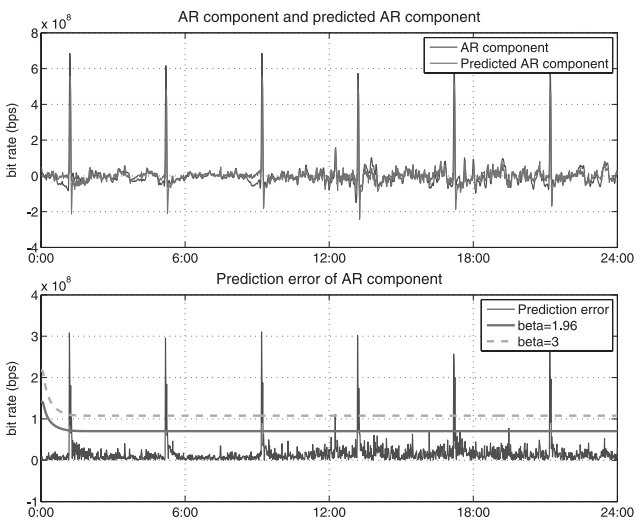

Fig. 5 Example of prediction error for AR component for traffic shown in Fig. 4.

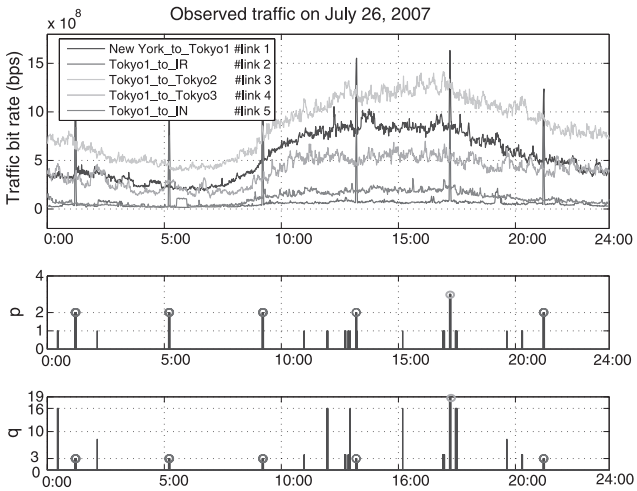

Fig. 6 Projection of detected matrices **P** and **Q** for observed traffic on July 26, 2007. (shown in Fig. 3)

indicates that #*link* 1 (New York to Tokyo1) and #*link* 2 (Tokyo1 to IR) are experiencing anomalies. At the point marked by the green circle, $p = 3$ indicates that there are three links experiencing anomalies at that point. In addition, $q = (19)_{10} = (10011)_2$ indicates that #*link* 1 (New York to Tokyo1), #*link* 2 (Tokyo1 to IR) and #*link* 5 (Tokyo1 to IN) are experiencing anomalies. The other points are judged as stochastic fluctuations.

3. Tracing Anomalies

Besides anomaly detection, another important issue is how to trace the sources of these detected anomalies. Our task in this paper is to trace the anomaly propagation path in the SINET3 backbone network shown in Fig. 7, in which 12 core nodes are connected by 28 unidirectional links. The key question is how many links and which ones we should inspect that can provide the necessary information for tracing anomalies. Suppose there is a traffic volume anomaly entering the SINET3 backbone at node 6 (e.g., through the link from Los Angeles to Tokyo2 in Fig. 2), it may propagate to other 11 core nodes. Taking into account that the traffic between each two core nodes will follow a route path obtained by a shortest path algorithm, the links on these shortest route paths are those links we should inspect for tracing an anomaly.

We develop an algorithm for our research that is based on Dijkstra's [17] algorithm to generate the shortest path tree starting at an ingress node s, which is shown in Table 1. It is obvious that the shortest route path tree will consist of 11 unidirectional links and we only need to inspect 11 unidirectional links to trace an anomaly. Otherwise, we need to inspect 28 links to trace an anomaly without the shortest route path tree. By marking the diagnosed result on the shortest route path tree, the propagation path of traffic anomalies can be traced.

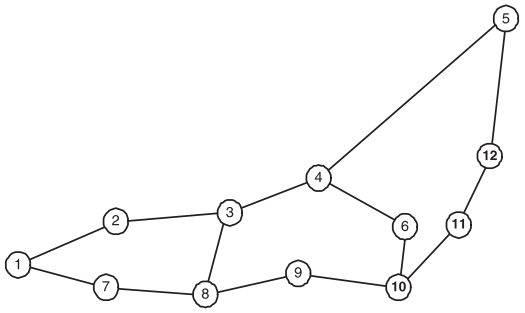

Fig. 7 Topology of core nodes in SINET3 backbone network.

Table 1 Algorithm for generating shortest route path tree.

G	whole core node set; $G = \{i\}_{i=1}^{12}$ for Fig. 7;
S	marked core node set;
Q	unmarked core node set;
$w(u, v)$	OSPF cost between neighbor node pair u and v;
$d[v]$	total cost of the path from the ingress node s to node v;
$p[v]$	the previous node of v along its shortest path.
1:	$d[s] = 0$; $S = \emptyset$; $Q = G$;
2:	**for** $\forall v \in G$ **do**
3:	$d[v] = \infty$;
4:	**end for**
5:	**while** $(Q \neq \emptyset)$ **do**
6:	$u = \{u : d[u] = min\{d[i]; i \in Q\}; u \in Q\}$;
7:	$S = S \bigcup \{u\}$; $Q = Q - \{u\}$;
8:	**for** $\forall v = \{v : w(u, v) \neq \infty; v \in Q\}$ **do**
9:	**if** $(d[v] > d[u] + w(u, v))$
10:	$d[v] = d[u] + w(u, v)$; $p[v] = u$;
11:	**end if**
12:	**end for**
13:	**end while**

4. Evaluation of AR Prediction Method

In this section, we evaluate the performance of our proposed AR prediction method for detecting possible anomalies on a single traffic link. For comparisons, we will start by introducing two other methods, the Exponential Weighted Moving Average (EWMA) [4], [5] (also called exponential smoothing) and the Non-seasonal Holt-Winters (NSHW) [4], [5] methods, which are currently being widely used in network analyses, such as in RRDTOOLS [4].

A. EWMA

Assume y_n is the analyzed traffic, then the prediction can be given by

$$\hat{y}_n = \lambda y_{n-1} + (1 - \lambda)\hat{y}_{n-1}, \tag{21}$$

where $0 \leq \lambda \leq 1$ is the weight parameter of past values. According to [5], [9], the values of $0.2 \leq \lambda \leq 0.3$ are suitable for traffic prediction. So, in this paper, we will also set 0.25 as the default value of λ. Anomalies are detected by thresholding $|y_n - \hat{y}_n|$.

B. NSHW

The Non-Seasonal Holt-Winters method is an advanced version of the EWMA method, in which the prediction $\hat{y}_n$ is

divided into two components: a trend component t_n and a smoothing component s_n. The prediction $\hat{y}_n$ is defined as $\hat{y}_n = s_n + t_n$, in which

$$s_n = \lambda y_{n-1} + (1-\lambda)\hat{y}_{n-1},$$
$$t_n = \mu(s_n - s_{n-1}) + (1-\mu)t_{n-1}, \quad (22)$$

where $\lambda \in [0,1]$ and $\mu \in [0,1]$. Here, we will set 0.25 as the default value of λ and 0.0035 as the default value of μ [3]. Anomalies can also be detected by thresholding $|y_n - \hat{y}_n|$, which is the same as for the EWMA method.

To evaluate the performance of AR prediction, EWMA and NSHW methods, anomalies from either real traffic data or synthetic anomaly generator can be used [5], [9], [10]. Because the exact statistical properties of anomalies are too difficult to be simulated, and A. Soule et al. [10] has shown that the synthetic anomalies may result in a **reversed** result to the real anomalies for evaluating anomaly detection methods, we will use the real traffic collected on different links of SINET3 to evaluate the AR prediction method comparing with EWMA and NSHW methods.

As has been previously introduced, all the EWMA, NSHW, and AR prediction methods rely on the selection of a threshold of the prediction error that is to decide whether there is an anomaly. The test result is a tradeoff between the detection rate and false alarm rate, where the detection rate is defined as the fraction of the true anomalies detected and the false alarm rate is defined as the fraction of the normal data points over the threshold. Whether the prediction errors at the anomalous points sharply stand out from the normal data decides the performance of a detection method.

In our evaluations, we take into account all the possible thresholds for each method and plot the results in the Receiver Operation Characteristic (ROC) [18] curves. In a ROC curve, the false alarm rate is plotted on the x-axis and the detection rate is plotted on the y-axis.

In Figs. 8 and 9, we compare the AR prediction method with the EWMA and NSHW methods using the real traffic collected on the links from Tokyo2 to Tokyo1 and from Tokyo1 to Nagoya in July 2007. From the left plot in each figure, we can clearly see that the AR prediction method performs the best, because its ROC curves climb more rapidly towards the upper left of the graph than the EWMA and NSHW methods do. This means that the AR prediction method can detect a higher fraction of the true anomalies with fewer false alarms. The NSHW method performance is a little better than that of EWMA method. For a false alarm rate of less than 0.15%, the AR prediction method misses no anomalies, while the EWMA and NSHW methods catch all the anomalies at a false alarm rate of about 6%–7%, which is more than 40 times the value of the AR prediction method.

The results of the plot of on the right of each figure show the tradeoff by the scaling factor β of the AR prediction method described in Sect. 2. When $\beta = 1.6$, the system catches all the anomalies with a false alarm rate of less than 0.15%. When $\beta = 1.96$, the system catches about 50% of the anomalies. When $\beta = 3.0$, the system can catch about 25% of the anomalies. How to choose β is decided by the

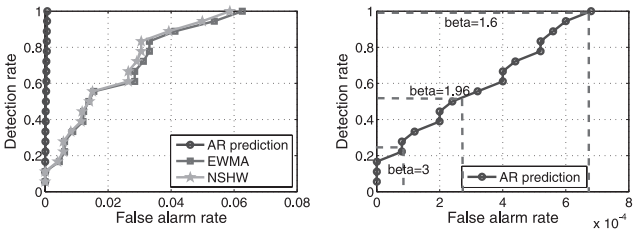

Fig. 8 ROC curves using real traffic collected on link from Tokyo2 to Tokyo1 in July 2007.

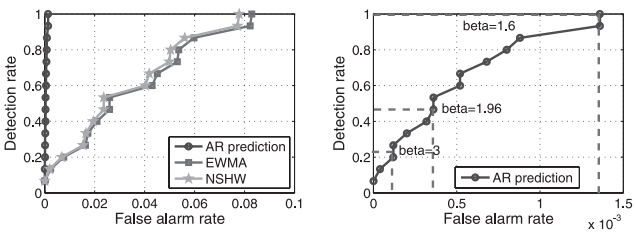

Fig. 9 ROC curves using real traffic collected on link from Tokyo1 to Nagoya in July 2007.

task of the network operation. If the network operators are only interested in the major anomalies that are more likely to cause congestions in the network, maybe a large β (e.g. $\beta = 3$) is preferred.

5. Validation of Tracing

In this section, we validate our proposed anomaly detection and tracing methods by an example of multi-sources and multi-hops anomalies in the SINET3 network. We observe that there were some possible traffic volume anomalies on the traffic from Los Angeles to Tokyo2 on the link from 2007-7-27 12:00 to 2007-7-28 12:00 and try to trace the sources for these possible anomalies. First, we execute the algorithm in Table 1 to generate the shortest route path tree starting at node 6 (node Tokyo2); the tree is shown in Fig. 10. For each link between a two core nodes pair in the tree, we assign a link number, which is marked in Fig. 10. The link from Los Angeles to Tokyo2 is marked as #*link* 12.

Next, we apply the detection method to these 11 traffic links shown in Fig. 10. Here we set $\beta = 3$. The detection results are shown in Fig. 11. In the figure, at the point marked by the red circle, $p = 7$ indicates that there are 7 links experiencing anomalies at that point. Furthermore, $q = (2545)_{10} = (100111110001)_2$ at that point indicates that #*links* 1, 5, 6, 7, 8, 9, and 12 are experiencing anomalies. We mark these detected links with red lines on the shortest route tree, and get the anomaly propagation path shown in Fig. 12 (a).

At the point marked by the green circle, $p = 8$ indicates that there are 8 links experiencing anomaly. Furthermore, $q = (3569)_{10} = (110111110001)_2$ indicates that #*links* 1,5,6,7,8,9,11, and 12 are showing anomalies. We also mark these detected links with red lines on the shortest route tree, and get the anomaly propagation path shown

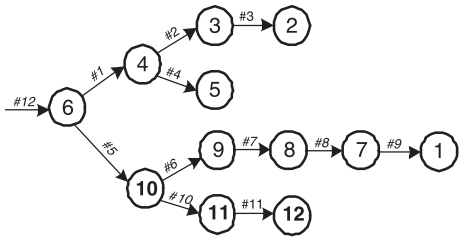

Fig. 10 Shortest route path tree starting at node 6.

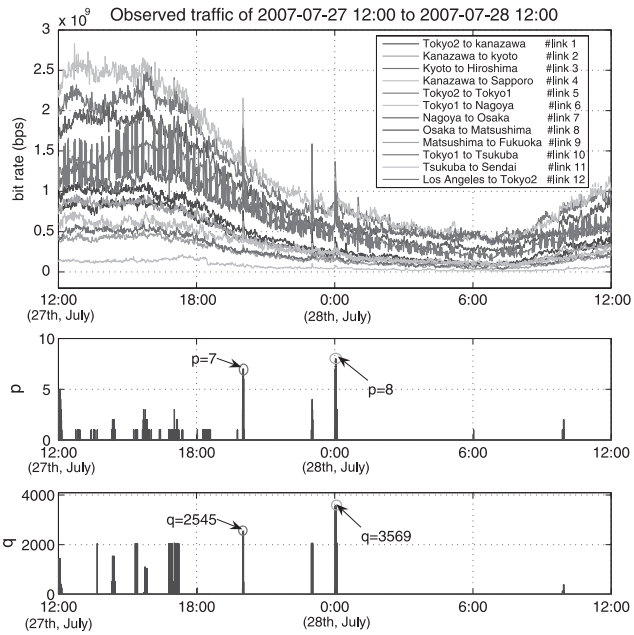

Fig. 11 Detection results for traffic on the links from 2007-7-27 12:00 to 2007-7-28 12:00.

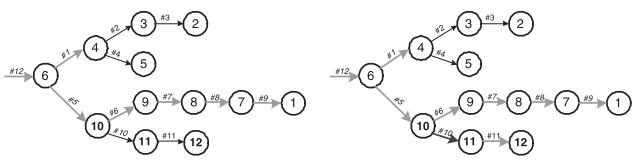

Fig. 12 Propagation paths for traffic anomalies in Fig. 11 which are marked with (a) red circle and (b) green circle. Red lines are traced propagation paths and blue line is deduced propagation path.

in Fig. 12 (b). Although the anomaly is detected on #*links* 12, 5 and 11, it is not detected on #*link* 10 (marked with the blue line). This is because that the corresponding spike in the #*link* 10 traffic is so small that it is difficult to detect even from a visual inspection. However, from the anomaly propagation path, we may think that #*link* 10 is also experiencing an anomaly at that point.

6. Conclusions and Future Work

In this paper we have proposed a traffic decomposition and prediction method to detect and trace traffic volume anomalies in the SINET3 network. The traffic is decomposed into three components: the trend component, the AR component, and the noise component, in which AR component consists of stochastic fluctuations and anomalous traffic. The traffic volume anomalies were detected when the AR component is outside the prediction band on some observed links simultaneously. Then we projected the detection result matrices on the time axis to diagnose which links and how many were experiencing the same anomalies. By marking the diagnosed result on the shortest route tree, which was generated through the algorithm listed in this paper, the traffic anomalies propagation path was traced.

Finally, our detection and tracing methods have been validated on the third-generation Science Information Network (SINET3). Our future work is to propose a more sensitive method to detect an anomaly even when its spike is hidden in the stochastic fluctuations during propagation.

References

[1] S. Urushidani, S. Abe, J. Matsukata, Y. Ji, K. Fukuda, M. Koibuchi, and S. Yamada, "Overview of SINET3 — Next-generation science information network," Progress in Informatics, no.4, pp.51–61, 2007.
[2] P. Barford and D. Plonka, "Characteristics of network traffic flow anomalies," Proc. ACM SIGCOMM Internet Measurement Workshop 2001, 2001.
[3] P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," Proc. ACM SIGCOMM Internet Measurement Workshop 2002, 2002.
[4] J. Brutlag, "Aberrant behavior detection in time series for network monitoring," USENIX Fourteenth System Administration Conference LISA XIV, pp.139–146, Dec. 2000.
[5] B. Krishnamurthy, S. Sen, Y. Zhang, and Y. Chen, "Sketch-based change detection: Methods, evaluation, and applications," Internet Measurement Conference, 2003.
[6] F. Feather, D. Siewiorek, and R. Maxion, "Fault detection in an Ethernet network using anomaly signature matching," ACM SIGCOMM, 1993.
[7] M. Thottan and C. Ji, "Anomaly detection in IP networks," IEEE Trans. Signal Process., vol.51, no.8, pp.2191–2204, Aug. 2003.
[8] C. Hood and C. Ji, "Proactive network fault detection," IEEE INFOCOM, 1997.
[9] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," Proc. ACM SIGCOMM 2004, 2004.
[10] A. Soule, K. Salamatian, and N. Taft, "Combining filtering and statistical methods for anomaly detection," Proc. ACM IMC 2005, 2005.
[11] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel, Time Series Analysis, Third ed., Prentice-Hall International, 1994.
[12] M. Pourahmadi, Foundations of Time Series Analysis and Prediction Theory, John Wiley & Sons, 2001.
[13] R. Jain, The Art of Computer Systems Performance Analysis, John Wiley & Sons, 1991.
[14] I.M. Chakravarti, R.G. Laha, and J. Roy, Handbook of Methods of Applied Statistics, Volume I, pp.392–394, John Wiley & Sons, 1967.
[15] N. Venables and M. Smith, "An introduction to R," http://cran.r-project.org/doc/manuals/R-intro.pdf
[16] S. Shapiro and B. Wilk, "An analysis of variance test for normality (complete samples)," Biometrika, vol.52, no.3/4, pp. 591–611, 1965.
[17] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, Introduction to Algorithms, Second ed., MIT Press and McGraw-Hill, 2001.
[18] T. Fawcett, "An introduction to ROC analysis," Pattern Recognit. Lett., vol.27, pp.861–874, 2006.

Appendix: Distribution Tests of AR Component

To test whether the AR components follow a Gaussian distribution, we first use a normal quantile-quantile (Q-Q) plot [13] to graphically compare the distribution of the AR component samples of the traffic shown in Fig. 3 to a normal distribution. When the theoretical Gaussian distribution matches the distribution of the measured AR component samples, the quantile-quantile plot will produce a straight line. The Q-Q plots results in Fig. A· 1 show that curves of AR component samples of $T1\_T2$, $T1\_T3$ and $T1\_IN$ fit the theoretical Gaussian distribution fairly well, where, "$T1\_T2$" represents the traffic from Tokyo1 to Tokyo2, "$T1\_T3$" represents the traffic from Tokyo1 to Tokyo3, and "$T1\_IN$" represents the traffic from Tokyo1 to IN.

For the AR component samples of NY_T1 (traffic from New York to Tokyo1) and $T1\_IR$ (the traffic from Tokyo1 to IR), the curves in the pink area also fit the theoretical Gaussian distribution very well. As has been analyzed, the AR component consists of stochastic traffic fluctuations and anomalous traffic. In the figures for NY_T1 and $T1\_IR$, the pink rectangular areas correspond to the stochastic traffic fluctuations and the other areas correspond to anomalous traffic. We would like to say that the stochastic traffic fluctuation follows the Gaussian distribution. In the next test, we will focus on testing the distribution of the AR components in the pink rectangular areas.

Furthermore, we use a one-sample Kolmogorov-Smirnov test (KS-test) [14], [15] to determine whether the distribution of the AR component samples and the Gaussian distribution (with the same mean and variance) significantly differed. The test results $D = 0.05$, 0.12, 0.06, 0.07, and 0.09 for NY_T1, $T1\_IR$, $T1\_T2$, $T1\_T3$ and $T1\_IN$ respectively, which are the maximum vertical deviations between the cumulative fraction plots of the distribution of the AR component samples and that of the Gaussian distribution, also indicate that there is very little discrepancy between the AR component and Gaussian distributions.

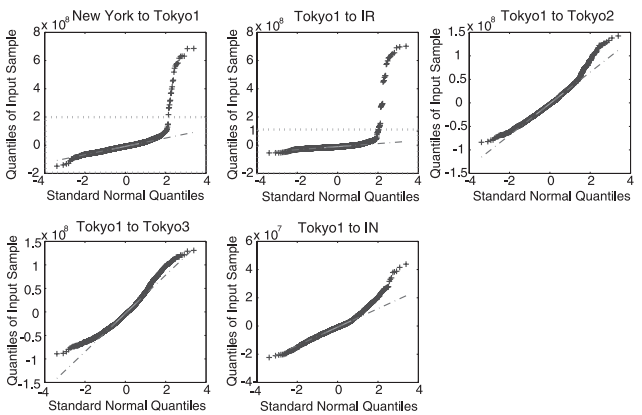

Fig. A· 1 Q-Q normal test for AR components of observed traffic.

Ping Du received B.E. and M.E. degrees from University of Science and Technology of China in 2000 and 2003, respectively. He received a Ph.D. from the Graduate University for Advanced Studies in Japan in 2007. Now, he works as a researcher at the National Institute of Informatics of Japan. His research interests include optical network, network security etc, Internet traffic analysis.

Shunji Abe received B.E. and M.E. degrees from Toyohashi University of Technology, Japan, in 1980 and 1982, respectively. He received a Ph.D. from the University of Tokyo in 1996. In 1982 he joined Fujitsu Laboratories Ltd., where he engaged in research on broadband circuit switching system, ATM switching system, ATM traffic control, and network performance evaluation. He worked at the National Center for Science Information Systems, Japan NACSIS) from 1995 to 1999. Since 2000 he has worked at the National Institute of Informatics of Japan as an associate professor. He is currently interested in the Internet traffic analysis, network performance evaluation, optical switching system architecture, and mobile IP system architecture.

Yusheng Ji received B.E., M.E. and D.E. in electrical engineering from the University of Tokyo in 1984, 1986 and 1989 respectively. She jointed the National Center for Science Information Systems in 1990. Currently she is an associate professor at the National Institute of Informatics, and the Graduate University for Advanced Studies. Her research interests include network architecture, traffic control and performance analysis for quality of service provisioning in high speed networks. She is also a member of IPSJ and IEEE.

Seisho Sato received B.Econ. from the University of Tokyo in 1991 and received M.Econ. and D.Econ. from the Tokyo Institute of Technology in 1993 and 2000 respectively. Currently he is an associate professor at The Institute of Statistical Mathematics. He is interested in time series analysis.

Makio Ishiguro received B.L.A., M.L.A. and D.E. in electrical engineering from the University of Tokyo in 1969, 1971 and 1984 respectively. Currently he is a professor at The Institute of Statistical Mathematics. He is interested in time series analysis and information criteria.