Anomaly Detection and Traffic Shaping under Self-Similar Aggregated Traffic in Optical Switched Networks

Wei Yan
Dept. of Electrical & Computer Engineering,
New Jersey Institution of Technology
University Heights, Newark, NJ 07102

Edwin Hou
Dept. of Electrical & Computer Engineering
New Jersey Institution of Technology
University Heights, Newark, NJ 07102

Nirwan Ansari
Dept. of Electrical & Computer Engineering
New Jersey Institution of Technology
University Heights, Newark, NJ 07102

Abstract — Recent work in traffic analysis has shown that modern network produces traffic streams that are self-similar over several time scales from microseconds to minutes. Simulation studies have demonstrated that self-similarity leads to larger queueing delays and higher drop rates than the Markovian Short Range Dependence (SRD) traffic. At the same time, the dramatic expansion of applications on modern network gives rise to a fundamental challenge for network monitoring and security. Therefore, it is critical to reduce the degree of second order scaling for better network performance and detect traffic anomalies efficiently. In this paper, we propose an approach which can capture the traffic anomalies and decrease the degree of Long Range Dependence at the conjunction of the optical packet switching backbone network. In this method, a traffic shaping technique is proposed and a reference model is generated based on the well-behaving traffic for anomaly detection. Further, we apply the compensation bursty parameter for smoothing the deviation error caused by burstiness difference existing in the traffic data sets. The simulation results show that our work can decrease the degree of self-similarity and detect the anomaly-behaving traffic efficiently.

Keywords — self-similar; network security; traffic anomalies; traffic shaping; Hurst parameter;

I. Introduction

A number of recent measurements and studies of actual traffic from modern networks demonstrated that real traffic has statistical self-similarity and is Long Range Dependent [1]. The traditional models such as Poisson or Markovian, which are short-range dependent, are basically not applicable to self-similar traffic. On the other side, the dramatic expansion of networking applications makes network security a pressing issue. Therefore, the monitoring of traffic anomalies is important to the security of modern networks. Traffic anomalies result from link corruption, buffer overflow and illegal ports scanning etc., Since traffic anomalies have no certain rules, the capture of them is fundamental to enhance the robustness and survivability of computer networks. In this paper, we propose a scheme to identify traffic anomalies and decrease the degree of self-similar at the one network device. In particular, we focus on the aggregated traffic on the edge devices at the conjunction of optical switched network, such as edge routers and edge switches. There are two reasons for that. Firstly, high-speed aggregated traffic is relatively stable than the traffic nearby the source nodes, which can make traffic analysis and measurement more precisely. Secondly, aggregated traffic is relatively easier to be extracted and isolated from the whole network, which help not to have much effect on the whole network's performance. Our monitoring module generates the reference traffic model based on the incoming traffic. The monitoring module can detect traffic anomalies by large anomaly deviation from ordinary-behaving traffic with this reference model. Since the trace on every time bin is with different burstiness character, in order to being fair, we apply the bursty compensation parameter to the calculation of deviation. Further, for the outgoing aggregated traffic, we use a traffic-shaping algorithm to decrease the degree of self-similarity.

This rest of paper is organized as follows. In section 2 we briefly introduce the definition of self-similarity and Sup_FRP method for generating the self-similar packet inter-arrival time model. Section 3 describes the architecture of edge device with monitoring module. In section 4 we propose a traffic-shaping algorithm to decrease the Hurst Parameter of output aggregated traffic. A method of detect the abnormal traffic is presented in section 5 and we conclude in section 6.

II. Self-Similarity Traffic Model

A. Definition of Self-Similarity

A number of empirical studies [1][2] has shown that the network traffic is self-similar in nature. For a stationary time series x, let the m-aggregated time series be:

$$x^{(m)} = \{x_k^{(m)}, k = 0,1,2,...\}$$

by summing the original time series over non-overlapping blocks of size m and averaging each block, where k labels the block and $x^{(m)}$ is expressed as

$$x_k^{(m)} = \frac{1}{m} \sum_{i=km-(m-1)}^{km} x_i$$

The stationary time series x is called to be exactly self-similar if for m = 1, 2, ...,

$$r^{(m)}(k) = r(k) \sim k^{-\beta}$$

where $k \rightarrow \infty$, $0 < \beta < 1$ and Hurst parameter $H = 1 - \beta/2$ ($\frac{1}{2} < H < 1$). The stationary time series x is called to be asymptotically self-similar if for $m \rightarrow \infty$

$$r^{(m)}(k) = r(k) \sim k^{-\beta}$$

The variance-time plot and R/S plot are two of most used methods to calculate Hurst parameter H.

B. Self-Similar Traffic Model for Simulation

We generate the self-similar traffic model of packet inter-arrival times according to [3]. Superposition of the Fractal Renewal Point Process (Sup_FRP) method is proposed by Ryu et al. and is constructed as the superposition of M i.i.d fractal renewal point processes. We have simulated 6 incoming traffic traces. These traces are the inputs for our monitoring module, as shown in Table 1.

Table 1. Six Input Traces for Simulation

Input Trace	Variance	H	$\hat{H}$	
			Variance-time	R/S
Trace1	0.000000076	0.9	0.8543	0.8295
Trace2	0.000497700	0.6	0.5881	0.5775
Trace3	0.000001306	0.65	0.6038	0.6356
Trace4	0.000000044	0.8	0.8073	0.8011
Trace5	0.000000001	0.9	0.8728	0.8641
Trace6	0.000000897	0.55	0.5323	0.5225

III. Function Architecture of Edge Device of Optical Switched Network

Optical switched network is exploited by the expeditious demand for bandwidth. Two different kinds of paths exist in optical switched network: data path and burst control packet (BCP) path [6]. Data path are of very high speed because of no electronic conversion and no data storage, whereas BCP path need time to process the packet source address, destination address and Qos classes. In backbone optical packet networks, traffic will be aggregated at various nodes like edge routers, edge switches etc., On the other hand, the performance of network is seriously degraded because of the self-similarity nature of the traffic [1][2]. For example, the probability of packet loss still decays very slowly even with the increasing the depth of Fiber Delay Lines (FDL) which are very expensive currently. Hence the study of the statistics of the aggregated data at those points is critical to network performance. Since the speed to process the control packet in BCP path is much slower than the transmission speed of data packet in data path, edge devices have to decouple the data path from the BCP path [6]. We have taken advantage of the process time gap to monitor the traffic and apply the traffic-shaping algorithm. Here we assume here that every incoming packet includes BCP section and data section.

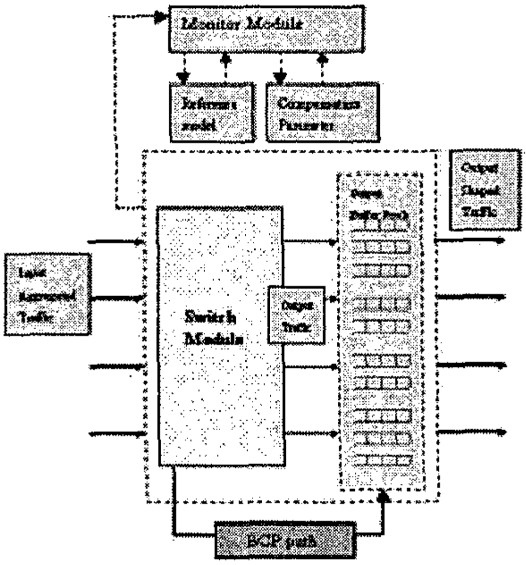

Figure 1. Function Architecture of Edge Switch

Fig 1. is a function architecture of one kind of edge devices, edge switch of optical switched network. It includes a switch module, queue buffer pools, monitor module and m inlet and outlet optical links. Input traffic is aggregated at inlet links whereas output aggregated traffic appears at the output side of switch module. After going through the queue buffer pools, shaped aggregated traffic is forwarded to core optical networks. Each outlet link has a queue buffer pool, which is composed of several FIFO buffers. Each buffer is assigned for one Qos class. We buffer several incoming packets with same source address, destination address and Qos class into a bigger one and transfer it as a whole. The monitoring module records the input aggregated traffic and split it into data sets of time bins. Meanwhile, the reference model is generated to detect traffic anomalies along with the compensation parameter.

IV. Traffic-Shaping Algorithm

One burst assembly mechanism is reported in [7]. However, this method only considered constant timeout value and did not propose the Qos classes solution. Our scheme takes consideration of different Qos classes *qi* (i = 1..n) and expound dynamic timeout value. As shown in fig. 1, the jth (j = 1..m) buffer pool is assigned for the jth outlet link of switch module. Each buffer pool is composed of n Qos buffers. The switch fabric in the device transfers every incoming packet to the *qi* buffer of a certain outlet based on the packet control section, which includes source information, destination information and qi. If the ith buffer of the jth buffer pool is not overflow or timeout Tij is not expire, the incoming packet is buffered. Otherwise, the buffered packet is sent out and T_{ij} is reset. Two triggers for sending out the packet in the qi buffer of every buffer pool are buffer overflow and timeout expiration. Timeout T_{ij} is started any time when a packet arrives in the empty qi buffer of j^{th} buffer pool. Since the traffic is changeable in the network all the time, we calculate the number of timeout triggers out of every 100 triggers in every qi buffer in j^{th} buffer pool, that is $P_{i,j}$\{timeout\}. Let

$P_{i,j}\{timeout\}=N_{(i,j)timeout}/(N_{(i,j)timeout}+N_{(ij)buffer\ overflow}))$.

P_{ij}\{timeout\} is used to calculate the value of $P_{i,j}$\{timeout\} in next 100 triggers. If $P_{ij}\{timeout\} > ½$, which means that the traffic load is not high, the timeout value should decrease and versa vice. Thus we let timeout $T_{i+1,j} = 2 *T_{i,j} * P_{i,j}\{timeout\}$. In our algorithm, the work load is 0.8, the buffer size b is set to be 2560 bits, the packet control section's length 32 bits, packet data length constant and initial timeout T_0 0.0025s and simulation time 4 seconds. The destination outlets are randomly chosen.

Table 2. Simulation of Output Traces

Output Trace	Variance	H	$\hat{H}$	
			Variance-time	R/S
Input aggregated traffic	0.0000000001	0.9	0.854309	0.829470
Output aggregated traffic	0.0000000014	0.9	0.833476	0.805432
Shaping aggregated traffic	0.0000010720	N/A	0.771348	0.700006

As shown in Table 2, Hurst parameter of input aggregated traffic is the maximum value of the 6 input traces and the Hurst parameter of output aggregated traffic is about 0.833476, which is approach to the maximum value of the 6 input traces. That means that simply aggregating self-similar traffic traces will not decrease LRD efficiently. However, with our shaping algorithm, the Hurst parameter is 0.771348,decreasing about 0.083.

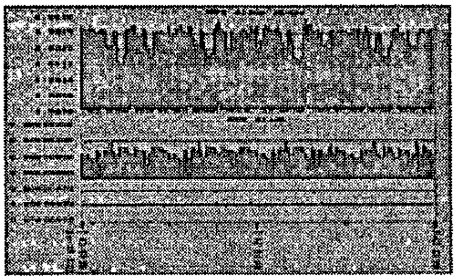

Figure 2. End-to-End Delay in 4 seconds

From the fig. 2, we can see that for 4 second simulation time, the traffic shaped algorithm can decrease the burstiness of end-to-end delay (ETE delay) efficiently, which is very beneficial to network performance. We also compare the traffic-shaping algorithm with timeout mechanism and traffic-shaping algorithm without timeout mechanism. We define rate of burstiness R (R = largest bursty value/mean) and found rate of burstiness with timeout mechanism, Rtimeout is about 1.3 while R no timeout approaches to 2.0.

V. Anomaly Traffic Detection

In addition to decrease the Hurst parameter H, we also expound a method to identify traffic anomalies. Our method is based on the multi-time scaling nature of self-similarity. If $X_{(t)}$ is self-similar with Hurst parameter H, then for $\forall a > 0$, $t \geq 0$,

$$X(at) =_d a^{H_x} X(t)$$

where $=_d$ means equality of finite dimensional distributions and a (a>0) is called scale factor. Thus $X_{(t)}$ and its time scaled version $X_{(at)}$, after normalizing must follow the same distribution[4].

In [5], a technique to detect errors in network traffic was proposed. This scheme detected errors by comparing the distribution deviation between traffic sets with reference trace. The assumption for that scenario is that Hurst parameter will keep relatively stable in a period of time slot. However, sometime ordinary traffic behavior appears bursty even in very small time bins. Therefore, abnormal traffic detection only by multi-time scaling is not enough for core networks. In order to make more accurately detection, it is necessary for detection schemes to take consideration of the disparities of different data sets. In this paper, we introduce an algorithm to detect traffic anomalies, which adopts bursty compensation parameter and can be tolerant to the changing degree of self-similarity in consecutive time bins.

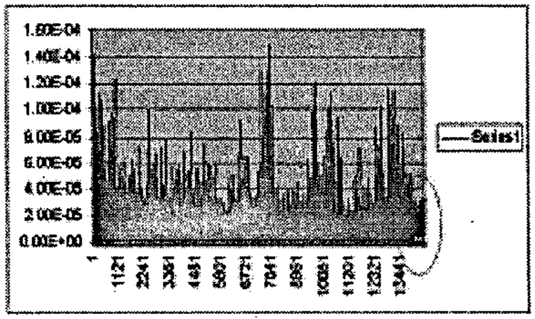

Figure 3. Traffic with Abnormal Behavior

Fig. 3 is the packet inter-arrival time trace of one segment of input aggregated traffic at the edge switch described in section 3. It includes a time series of abnormal traffic, which is marked with red circle. We built up two time scales trace of that segment, $X_{(at)}$ and $X_{(t)}$, where a >1 and $X_{(at)}$ is aggregate traffic distribution on time scale "at". (In our work, a = 10.) Based on $X_{(at)}$, we use multi-time scaling of self-similarity to generate reference model $X_{r(t)}$.

$$X_{r(t)} = (a^{-H})^{*} X_{(at)}$$

In our algorithm, we divided it into 10 data sets, 400 values with each set, and need to know the histogram of each data set. From the figure, we observe that each data set appear different burstiness character. In such data sets with heavy-tailed distributions, most of the observations are small, but most of the contribution to the sample mean or variance comes from the few large observations [4]. Thus we must consider this feature when figuring out the histograms. The scope we choose for histogram in each data set is the 1.0σ~2.0σ from average, where σ is standard deviation.

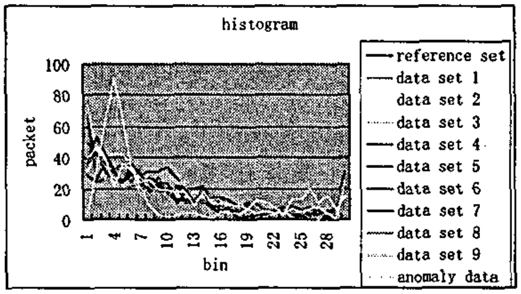

Figure 4. Histograms of 10 Data Sets and Reference Model

For each data set, the histogram of 30 equivalent time bins is computed. We compare the histogram of each data set with the histogram of reference model $X_{r(t)}$ by deviation. Fig. 4 presents the histograms of 10 data sets and reference model. Note that the traffic anomaly exists in the 10th data set, which is of large distinction from others. Since the trace on every data segment is with different burstiness character, in order to be fair, we apply the smoothed burstiness parameter δ and bursty compensation parameter ε to compensate the deviation error ω resulting from the burstiness. Because the few largest observations in every data set are so bursty, we define smoothed burstiness parameter δ_i of ith data set:

δ = (mean of *1%* largest observations)/mean of whole data set)

Based on δ, bursty compensation parameter ε_i of ith data set is defined as $\varepsilon_i = \delta_{mean}/\delta_i$. Finally, the compensation weight ω_i is

$$\omega_i = \{(X_1 \cdot \varepsilon_1 - Z_1)^2/Z_1\} + \{\Sigma (X_j \cdot \varepsilon_j - Z_j)^2/Z_k\} + \{(X_k \cdot \varepsilon_k - Z_k)^2/ Z_k\}$$

where j = 2 ... k-1, Z is the number of packets lies in every time bin of a data set and k is the number of data sets.

Table 3. δ, ε and ω of 10 Data Sets

Data sets	1	2	3	4	5	6	7	8	9	10
δ	6.43	5.61	6.66	4.89	4.69	5.02	6.81	4.83	4.81	4.92
ε	0.85	0.97	0.82	1.16	1.16	1.09	0.81	1.13	1.15	1.11
ω	-20.5	0	-43.8	+5.28	+2.52	+16.5	-14.9	+2.6	+5.6	-2.34

According to table 3, deviation without compensation trace is adjusted by ω. It is clearly in Figure 6 that ω smoothes the deviation error caused by burstiness of traffic within normal-behavior data sets, whereas has little effect on that in abnormal traffic.

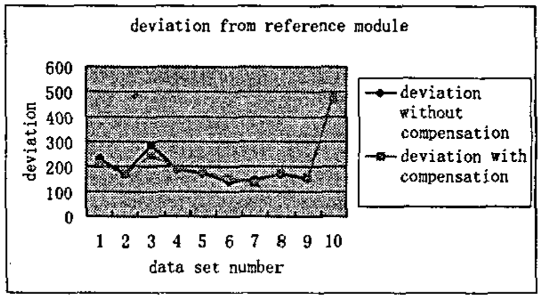

Figure 5. Deviation w/o Compensation

6. Inclusion

In this paper, we propose a scheme of aggregated traffic analysis to identify traffic anomalies efficiently and decrease the degree of self-similar at the one network device. Our monitoring module generates the reference traffic model. The monitoring module can detect traffic anomalies by large anomaly deviation from ordinary-behaving traffic with compensation weight efficiently.

REFERENCES

[1] V. Paxson and S. Floyd, "Wide-area traffic: the failure of Poisson modeling," *Proceedings of ACM Sigcomm'94*, pp. 257 - 268, 1994.

[2] W.Leland, et al., "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*,1994, 2(1):1-15

[3] B.Ryu and S. Lowen, "Fractal Traffic Models for Internet Simulation," *IEEE Symposium on Computers and Communications (ISCC)*, Juan-Les-Pins, France, 2000.

[4] Kihong Park, Walter Willinger, "Self-similar network traffic and performance evaluation," *John Wiley &Sons Inc*, pp. 91, 2000

[5] Schleifer, W., Männle, M. "On-line error detection through observation of traffic self-similarity," *IEE Proceedings of Communications*, Volume 148, Issue 01, pp. 38-42, February 2001

[6] A. Ge et al., "On Optical Burst Switching and Self-Similar Traffic," *IEEE Communications Letters*, vol. 4, no.3, pp.98-100, March 2000