

On WAN usage patterns in local environment

Chandrashekhar G Dethe^{a*} & D G Wakde^b

^aDepartment of Electronics, ^bPrincipal SSGM College of Engineering, Shegaon 444 203

Received 4 December 2003; rev recd 12 July 2004; accepted 05 August 2004

Today's complex network traffic is the aggregation of traffic generated by individual sources on network. Therefore, access pattern of users is of immense importance in order to model and simulate the network traffic using user characteristics. In this study, we identify parameter matrix to characterize the individual user and compute their correlation analysis in order to find their independence, useful criterion for modeling.

Keywords: Wide area network. Networks, Network traffic

IPC Code: Int. Cl.⁷: H 04 L 12/28

1 Introduction

Today's network traffic is a highly complex process. This makes it difficult to get better model, which can simulate the network traffic. Often, we get highly parameterized traffic models for such complex traffic. A good traffic model may not track the traffic but reveals some surprises in the behaviour of traffic, which are interesting to study.

In order to model the aggregate network traffic, it is important to look at the traffic generated by individual users. Since the aggregate traffic is the superposition traffic of individual sources, it is necessary to exploit the superposition of sources to characterize the aggregate traffic. Even though, an individual user may not affect the aggregate traffic significantly but group of users can certainly modulate the traffic characteristics. Thus, given the users and their characteristics we hope to come up with a model to simulate the aggregate traffic.

Section 2 gives details about the packet traces used. Section 3 discusses previous studies on traffic measurements. Section 4 describes the characterization of individual users. Correlation among the various parameters is discussed in Section 7. Findings of user-based analysis are given in Section 8. Aggregate traffic parameters and their analysis are explained in Section 9. Correlation analysis of

aggregate traffic parameters and corresponding results are discussed in Section 10 and 11, respectively.

2 Trace Details

Traffic traces used in this experimental work are downloaded from www.moat.nlanr.net/trace.html. The WAND research group captured these traces at University of Auckland Internet uplink. The University of Auckland ITSS Department operates an OC3c ATM link. An ATM virtual channel (VPI = 10, VCI = 103) is used to connect the university to the global Internet. The connection has a packet peak rate of 2 Mbits/s set in each direction (4 Mbits/s when you aggregate the two trace files).

3 Related Work

In the past, we came across many studies attempting to get parsimonious model for network traffic. An important breakthrough was the study by Paxson¹ who exposed the limitations of Poisson model for computer network traffic. Poisson model is considered as parsimonious model for tele-traffic, but fails to predict computer network traffic. Another important study to model computer network traffic was the Packet Train model given by Jain and Routhier². But measurements as reported by Jain and Routhier² were taken on a token ring network and suitability of Packet Train model² is not known for today's complex traffic. Leland *et al.*³ were the first to demonstrate the self-similar nature of network traffic. Many others⁴⁻¹² follow this study giving different models, which account for self-similar behaviour of

*Author for correspondence
E-mail: cgdethe@ssgmce.ac.in

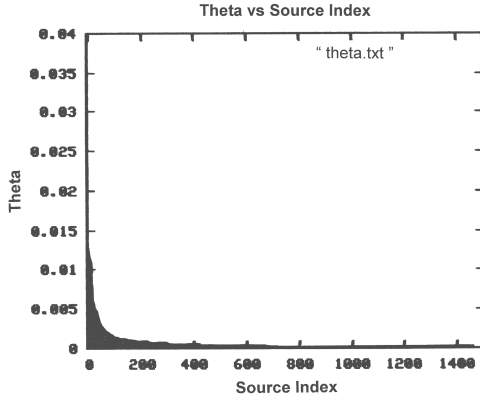


Fig. 1 — Ranked user utilization

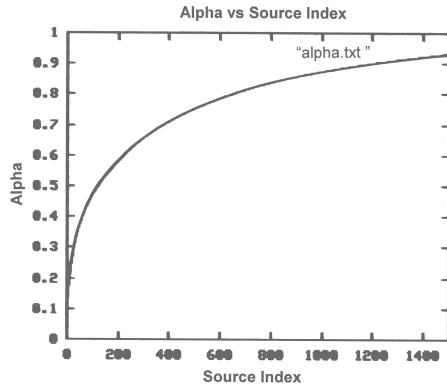


Fig. 2 — Cumulative utilization of ranked users

the network traffic. In particular, we are motivated by Packet Train model² and the network traffic. In particular, we are motivated by Packet Train model² and the flow based traffic analysis by Claffy *et al.*¹¹. In this study, aggregate traffic is characterized by defining a flow and corresponding flow timeout that was a user defined parameter. But it did not consider individual user's impact on the aggregate traffic. In our study, we characterize individual user and its contribution to the aggregate traffic.

4 Individual User Characterizations

We have defined three parameters to characterize user behaviour. These are Utilization (θ), Duty_On cycle (η), and Burstiness (ζ). These parameters are defined as given below:

4.1 Utilization (θ)

It is defined as the ratio of traffic, in bits, put by a user to the total traffic in bits during the entire trace duration. Thus, if B_i is the traffic introduced by i^{th} user then,

$$\theta_i = \frac{B_i}{\sum_{j=1}^n B_j} \quad \dots(1)$$

where n is total number of users. The parameter θ helps us to identify the dominant users in terms of putting the number of bits in the total traffic. In fact, we can rank the users by defining α which is the cumulative utilization, as follows:

$$\alpha = \sum_i \theta_i \quad \dots(2)$$

Ranked user utilization θ and cumulative utilization α are shown in Fig. 1 and 2.

4.2 Duty_On Cycle (η)

It is defined as the ratio of ON time of a user to sum of ON and OFF times for a user. ON time is the period over which the user is active. ON time is calculated using the bit volume process of a user. Details of ON/OFF time process and bit volume process are discussed in Section 5.

If $U_i^{(j)}$ denotes the j^{th} ON time and $D_i^{(j)}$ denotes the j^{th} OFF time of i^{th} user then,

$$\eta_i = E[\eta_i^{(j)}] \quad \eta_i^{(j)} = \frac{U_i^{(j)}}{U_i^{(j)} + D_i^{(j)}} \quad \dots(3)$$

We have large number of hosts present on the network but not all of them will be accessing the network at a time, i.e. active. Some of them will be in the active state and rest will be in the idle state. Different users have different access patterns and the parameter η shows how much active a user is.

4.3 Burstiness (ζ)

An important characteristic of a user, which has the strong impact on the aggregate traffic behaviour, is its burstiness. We define Zeta (ζ) as a measure of traffic burstiness as given below:

$$\zeta_i = E[\zeta_i^{(j)}] \quad \zeta_i^{(j)} = \frac{\sqrt{\text{Var}(N_i^{(j)})}}{E[N_i^{(j)}]} \quad \dots(4)$$

where $(N_i^{(j)})$ is a random variable representing byte arrivals in a window of size T time units for j^{th} session of i^{th} host (session is defined in Section 6). Thus, a burstier user with large bit volume is likely to have more impact on network traffic. This may lead to degradation in network performance by introducing delays, packet loss, etc. Hence, we find this is a useful individual performance parameter to model the user.

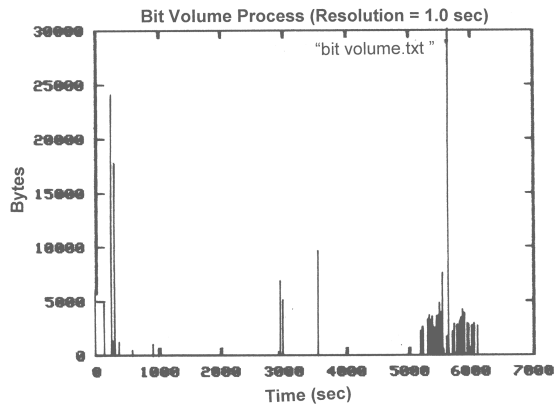


Fig. 3 — Bit volume process

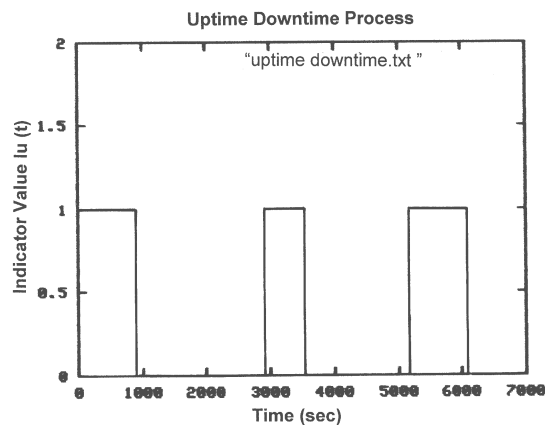


Fig. 4 — ON/OFF process of a user

5 Bit Volume and ON/OFF Time Process

5.1 Bit Volume Process

Consider the access pattern of a user along time axis. Now divide the time axis in fixed size windows of T units, we get number of bits put by user in each of these windows (called as resolution) for calculating the bit volume of that user. Thus, bit volume will be zero if user is silent, and will be equal to number of bits put by the user if it is active during the resolution period, T . This can be put as follows:

$$B[nT] = \begin{cases} 0 & \text{if } n=0 \\ \text{bits introduced in } [(n-1)T, nT] & \text{otherwise} \end{cases} \quad \dots (5)$$

A typical Bit Volume Process for a user is shown in Fig. 3.

5.2 ON/OFF Time Process

From bit volume process we can depict user's access pattern into cycles of ON and OFF periods.

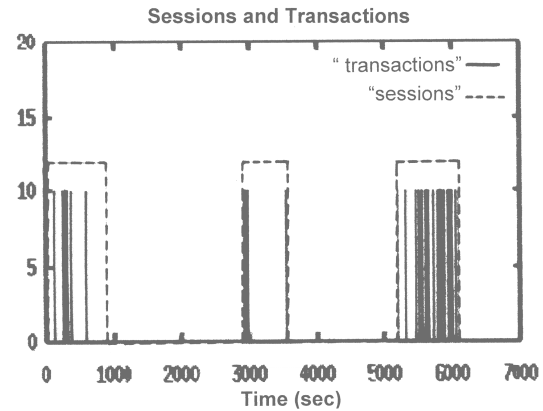


Fig. 5 — Sessions and transactions

Particularly, we call a user is in ON state if its bit volume process is non-zero for at least user defined time period from the given time instant. It is said to enter into OFF state if user fails to put traffic within user-defined period from the given time instant. Thus, ON/OFF time cycles forms a random process. ON/OFF process for a given user is shown in Fig. 4.

6 Sessions and Transactions

A session can be defined as the time period over which a user is executing some networked applications. A typical user operates in multiple sessions. If a user does not perform any network activity for sometime (typically called as session timeout period) then we say that user has entered in OFF state. Each session has number of transactions, which are typically TCP connections, established between two users. Fig. 5 shows sessions and transactions within the sessions for a user.

7 Correlation Analysis

Having got the values of utilization (θ), Duty_On Cycle (η) and Burstiness (ζ) for all the users, we have computed the correlation matrices for different resolutions, as shown in Table 1.

From the above tables we observe that there exists significant negative correlation between η and ζ . Also at higher resolution there is negative correlation between θ and ζ .

8 Results (User Based Analysis)

- Out of total 12947 users (sources), which were encountered, only 1194 users contribute to 90 per cent of the traffic.

Table 1 — Correlation matrix of θ , η and ζ

	θ	η	ζ
θ	1.00	0.04	-0.10
η	0.04	1.00	-0.52
ζ	-0.10	-0.52	1.00

$T_{\text{timeout}} = 5 \text{ min.}$

θ	η	ζ	
θ	1.00	0.06	-0.12
η	0.06	1.00	-0.55
ζ	-0.12	-0.55	1.00

$T_{\text{timeout}} = 10 \text{ min.}$

θ	η	ζ	
θ	1.00	0.07	-0.13
η	0.07	1.00	-0.60
ζ	-0.13	-0.60	1.00

$T_{\text{timeout}} = 20 \text{ min.}$

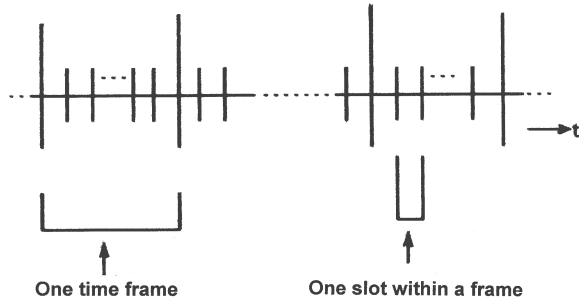


Fig. 6 — Time window and slots along trace duration

- Machine sources contribute to 0.3 per cent of traffic.
- Utilization is uncorrelated to Burstiness and Duty_On.
- Users active for longer durations are less likely to be bursty.

9 Aggregate Traffic Analysis

9.1 Procedure

In order to compute aggregate traffic statistics we divided the trace time axis in time windows of equal length. Also each time window is divided into an

integral number of slots. Window duration and slot length are user-defined parameters. Trace duration time axis along with time window and time slots is shown in Fig. 6.

Corresponding to each time-window along the trace duration we have defined following parameters in order to characterize aggregate traffic:

- Number of Packets — A packet count is maintained for each slot in a given time window. Thus, if P_i is the packet count in i^{th} slot of j^{th} window then,

$$P_j = \sum_{i=1}^N P_i \quad \dots(6)$$

where P_j is total number of packets in j^{th} window and N is number of slots in window.

Similarly, a byte count B_j is maintained on per window basis, which gives total bytes in j^{th} time window. This count is used further in the definitions of following parameters:

- Bit-Rate — From the byte count B_j we calculate Bit-Rate per time window by averaging over the window length. Thus Bit-Rate BR_j for the j^{th} window of length T can be given as follows:

$$BR_j = \frac{B_j}{T} \quad \dots(7)$$

Also we have maintained Bit-Rate difference BRD_j as follows:

$$BRD_j = BR_j - BR_{j-1} \quad \dots(8)$$

- Burstiness — We have defined burstiness in the aggregate traffic over two granularities:
 - Packet based burstiness
 - Byte based burstiness

Byte based burstiness $BBST_j$ is defined as follows:

$$BBST_j = E[BBST_j^{(i)}] \quad BBST_j^{(i)} = \frac{\sqrt{\text{Var}(B_j^{(i)})}}{E[B_j^{(i)}]} \quad \dots(9)$$

With similar analogy we can define packet based burstiness for a given packet count P_j

$$PBST_j = E[PBST_j^{(i)}] \quad \dots(10)$$

- Number of ON Sources — We have maintained a count for the number of active sources within a user defined time window. For this purpose we

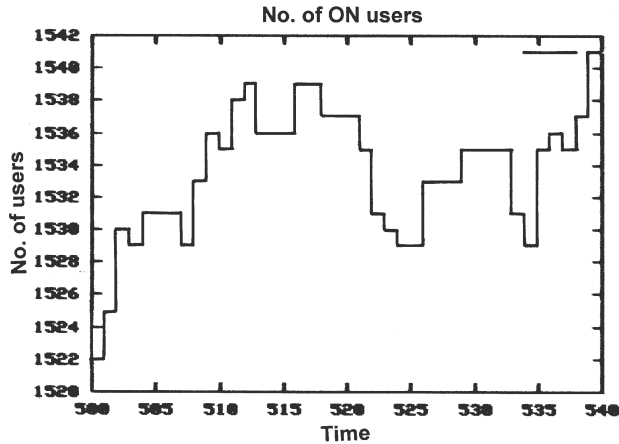


Fig. 7 — ON users process

Table 2 — Correlation matrix for time window = 30 s and Timeout = 300 s

	BR	BRD	NOP	PBST	BBST	NONU
BR	1.0	0.55	0.57	-0.03	0.22	0.05
BRD	0.55	1.0	0.34	0.01	0.30	0.01
NOP	0.57	0.34	1.0	-0.26	-0.06	0.11
PBST	-0.03	0.01	-0.26	1.0	0.38	-0.11
BBST	0.22	0.30	-0.06	0.38	1.0	-0.13
NONU	0.05	-0.01	0.11	-0.11	-0.13	1.0

Table 3 — Correlation matrix for time window = 60 s and Timeout = 300s

	BR	BRD	NOP	PBST	BBST	NONU
BR	1.0	0.60	0.53	-0.03	0.28	0.07
BRD	0.60	1.0	0.37	-0.01	0.34	-0.02
NOP	0.53	0.37	1.0	-0.33	-0.05	0.13
PBST	-0.03	-0.01	-0.33	1.0	0.34	-0.15
BBST	0.28	0.34	-0.05	0.34	1.0	-0.22
NONU	0.07	-0.02	0.13	-0.15	-0.22	1.0

have to define timeout period to find whether a user is in ON state or not. Timeout period is a user configurable parameter. The number of ON users in a given time window is equal to sum of ON users in previous window and newly arrived users in current window minus the users not able to put traffic within timeout

Table 4 — Correlation matrix for time window = 90 s and Timeout = 300 s

	BR	BRD	NOP	PBST	BBST	NONU
BR	1.0	0.61	0.49	-0.06	0.30	0.07
BRD	0.61	1.0	0.35	-0.08	0.37	-0.03
NOP	0.49	0.35	1.0	-0.40	-0.09	0.15
PBST	-0.06	-0.08	-0.40	1.0	0.35	-0.20
BBST	0.30	0.37	-0.09	0.35	1.0	-0.30
NONU	0.07	-0.03	0.15	-0.20	-0.30	1.0

period since they were active. If N is the number of ON users in j^{th} window then,

$$N = [N \text{ in } (j-1) + \text{New Arrivals in } j] - [\text{Users went into OFF state in } j]$$

Thus the number of ON users will look like a counting process. This is shown in Fig. 7.

10 Correlation Analysis (Aggregate Trace)

The parameters characterizing the aggregate traffic, which are defined in Section 9 were obtained for different sizes of time window, varying from 5 to 90s. The correlations among them were computed which are shown in Tables 2-4.

From Tables 2-4 we observe that Bit Rate (BR) is correlated to Byte based burstiness (BBST), to Number of packets (NOP) and to Bit Rate Difference (BRD). Also NOP and BBST are correlated to BRD. There is negative correlation between NOP and PBST. Interestingly, we observe that BBST is not correlated to NOP. The number of ON users is uncorrelated to all these parameters. Significance of these observations is given in Section 11.

11 Results (Aggregate Traffic Analysis)

- Increase in Bit rate does not lead to heavy bursty traffic. This is because extra bytes coming in time window get divided into different time slots of that window.
- Negative correlation between NOP and PBST indicates that extra packets coming in a time window are distributed evenly in all the time slots of that time window and thus help in reducing packet based burstiness.
- NOP is not correlated to BBST. This is due to the fact that packets are of variable lengths.

References

- 1 Paxson V & Floyd S, Wide area traffic: The failure of Poisson modeling, *IEEE/ACM Trans Networking*, **3** (June 1995) 226-245.
- 2 Jain R & Routhier S, Packet trains: Measurements and a new model for computer network traffic, *IEEE Selected Areas Commun*, **SAC-4** (September 1986) 986-995.
- 3 Leland W E, Taqqu M S, Willinger W & Wilson D V, On self similar nature of Ethernet traffic (extended version), *IEEE/ACM Trans Networking*, **2** (February 1994) 1-15.
- 4 Shu Y, Zhang L & Wang L, Traffic Predictions using FARIMA models, *Proc ICC'99*, **2** (1999) 891-895.
- 5 Crovella M E & Bestavros A, Self-similarity in World Wide Web traffic: Evidence and possible causes, *IEEE/ACM Trans Networking*, **5** (December 1997) 835-845.
- 6 Tsybakov B & Georganas N D, Self-similar process in communication networks, *IEEE Trans Information Theory*, **44** (1998) 1713-1725.
- 7 Grossglauser M & Bolot J C, On the relevance of long-range dependence in network traffic, *ACM Trans Networking*, **7**(8) (1999).
- 8 Addie R G, Zukerman M, & Neame T D, Broadband traffic modeling: Simple solutions to hard problems, *IEEE Commun Magazine*, (August 1998) 88-95.
- 9 Thompson K, Miller G J, & Wilder R, Wide area Internet traffic patterns and characteristic, *IEEE Networks*, **11** (December 1997) 10-23.
- 10 Kihong Park G K & Crovella M, On the relationship between file sizes, transport protocols, and self-similar network traffic, *Proc ICNP'96*, (October 1996) 171-180.
- 11 Claffy K, Braun W & Polyzos G C, A parameterizable methodology for Internet traffic flow profiling, *IEEE Selected areas Commun*, **13** (8) (October 1995) 1481-1494.
- 12 Paxson V, Empirically derived analytical models of wide area TCP connections, *IEEE/ACM Trans Networking*, **2** (4) (August 1994) 316-336.