Methodology for Modeling the Impact of Traffic Burstiness on High-Speed Networks

Tibor Gyires
Applied Computer Science Department
Illinois State University, Normal, Illinois 61790-5150
tbgyires@ilstu.edu

ABSTRACT

Recent results published in [1,3,4,7] prove that high-speed network traffic is more bursty and its variability cannot be predicted as assumed previously. According to the authors, network traffic has similar statistical properties in a certain amount of time called self-similarity. One of the consequences is that combining the various flows of data, as it happens for example in ATM virtual paths, does not result in the smoothing of traffic. Combining bursty data streams will also produce bursty combined data flow. The studies imply that the methods and models used in traditional network design require modifications. The objective of this paper is to present a new methodology that estimates and measures the impact of *bursty* traffic on network links and network devices applying discrete even simulation techniques. By using the new methodology, we can get more accurate results in measuring the momentarily utilization of links, response time, and the queuing performance of switches and routers in a large network. We illustrate the methodology with a Comnet model [2,9].

1. INTRODUCTION

The results published in [1,3,4,7] make evident that high-speed network traffic is more bursty and its variability cannot be predicted as assumed previously. According to these results, network traffic has similar statistical properties on many time scales called self-similarity. Traffic that is bursty on many or all time scales can be described statistically using the notion of self-similarity. Self-similarity is often associated with objects in fractal geometry: objects that appear the same regardless of the scale at which they are viewed. In bursty traffic the term self-similarity is used in a distribution sense: when viewed at varying scales, the traffic's distribution remains unchanged. Self-similar traffic has observable bursts on all time scales. Therefore, measuring burstiness is the same as characterizing the self-similarity of the network traffic.

One of the consequences is that combining the various flows of data, as it happens for example in ATM virtual paths, does not result in the smoothing of traffic. Combining bursty data streams will also produce bursty combined data flow. The studies imply that the methods and models used in traditional network design require modifications. Self-similar traffic makes traditional assumptions incorrect. For instance, it has been assumed that linear increase in buffer sizes results in an exponential decrease in packet loss. In self-similar traffic the decrease of packet loss is far less than predicted by traditional models. It has also been proven that in case of self-similar traffic, the addition of few new circuits can cause a large loss of data packets.

Many methods for measuring traffic burstiness are based on the estimate of the Hurst parameter H in addition to the bursts' mean and variance: The higher the value of H, the higher the burstiness. Most of the papers (e.g., [3, 5, 6]) on traffic modeling focus on analytical models. They compare the distributions of the model-generated traffic and the real measured traffic to test the theory that the real traffic results from model like systems. These papers do not investigate the impact of burstiness on the various network components.

The objective of our paper is to introduce a new methodology to measure the impact of the burstiness on network links and network devices. We also measure the Hurst parameter based on real traffic traces, but instead of using analytical models we integrate the Hurst parameter with the discrete event simulation system Comnet [2,9]. Traditional analytical methods cannot cope with the effects of random variance. The assumptions required by analytical methods ignore the effects of queuing, event interdependence, and random variance when analyzing complex, high-speed communication networks. By using discrete event simulation methodology, we can get more realistic and accurate results in measuring network performance parameters, such as the momentarily utilization of the links, response time, and the queuing performance of switches and routers along the traffic paths. Our method can model and measure the harmful consequences of aggregated bursty traffic, and predict its impact on the network's performance parameters.

The second section reviews the definition and consequences of burstiness. The section also describes measuring methods of traffic burstiness. The third section presents the traffic model. The section includes the results

0-7803-5731-0/99$10.00 ©1999 IEEE

of the model response time for increasing intensity of burstiness. Appendix A graphically illustrates the model results; Appendix B specifies the details of our implemented method.

2. TRAFFIC BURSTINESS

Each transaction between a client and a server consists of active periods followed by inactive periods. Transactions consist of groups of packets sent in each direction. Each group of packets transmitted without interruption is called a burst. In bursty traffic the variance of the bursts' length can be very high. High burstiness means that long periods of heavy traffic are typically followed by long periods of light traffic. During the heavy traffic the buffer queues in switches and routers build up and significant losses can occur. Although these long periods of heavy traffic are typically followed by long periods of light traffic the damage is already done [3].

The burstiness of the traffic can be characterized by the following time parameters (See Appendix B):

Transaction Interarrival Time (TIAT): The time between the first packet in a transaction and the first packet of the next immediate transaction.

Burst Interarrival Time ($1/\lambda$, λ: arrival rate of bursts): The time between bursts.

Packet Interarrival Time ($1/\delta$, δ: arrival rate of packets): The time between packets in a burst.

2.1. Consequences of burstiness

Measuring and modeling network traffic with methods that do not reflect the burstiness of the traffic will result in analysis that significantly underestimate performance measures, such as link utilization, response time, packet delay, maximum queue size, or buffer size in routers and switches. Various papers [1,4,5,7,10] discuss the impact of the burstiness on network congestion. Their conclusions are that:

- congested periods can be quite long with losses that are heavily concentrated,
- linear increases in buffer size do not result in large decreases in packet drop rates, and
- a slight increase in the number of active connections can result in a large increase in the packet loss rate.

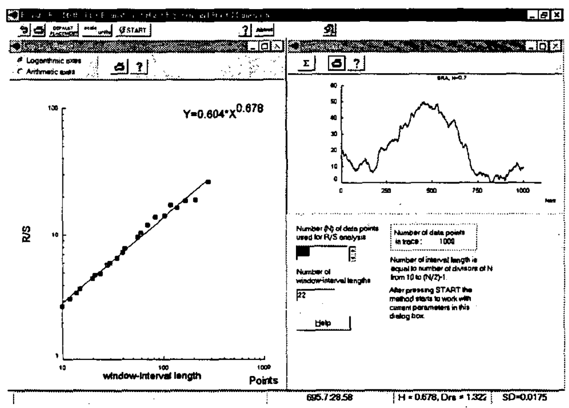
Results show that packet traffic "spikes" (which cause actual losses) ride on longer-term "ripples", that in turn ride on still longer-term "swells" [1].

Another area where burstiness can affect network performance is a link with priority scheduling between classes of traffic. In an environment where the higher priority class has no enforced bandwidth limitations, (other than the physical bandwidth) interactive traffic might be given priority over bulk-data traffic. If the higher priority class is bursty over long time scales, then the bursts from the higher priority traffic could obstruct the lower priority traffic for long periods of time.

The burstiness may also have an impact on networks where the admission control mechanism is based on measurements of recent traffic, rather than on policed traffic parameters of individual connections. Admission control that considers only recent traffic patterns can be misled following a long period of fairly low traffic rates.

2.2. Methods for measuring burstiness

Most of the methods (Variance, Aggregated Variance, Higuchi, Variance of Residuals, Rescaled Adjusted Range (R/S), Whittle Estimator, Periodogram, Residuals of Regression [7]) for measuring burstiness are based on the estimate of the Hurst parameter H, $0.5<H<1$. Some are more reliable than others. The reliability depends on several factors, e.g., the estimation technique, sample size, time scale, traffic shaping or policing, etc. Based on published measurements we chose the Rescaled Adjusted Range (R/S) method implemented in the Benoit package [14] to estimate the Hurst parameter of real traffic traces. Below, we include a screen output of the package as an example:



In our model the Hurst parameter characterizes the variance of the number of bytes sent and received in a certain time interval between a client and a server. The higher the value of H, the higher the burstiness, and consequently the worse the queuing performance of switches and routers along the traffic path. Appendix B specifies the details of the integration of the Hurst parameter in the Comnet modeling environment.

3. TRAFFIC MODEL

In order to build the baseline model, we collected traffic traces in a large corporate network by the Concord Network Health network analyzer system. We took measurements from various broadband and narrowband links including 45Mbps ATM, 56Kbps, and 128 Kbps frame relay connections. [15]. Concord Network Health can measure the traffic in certain fixed time intervals in the order of seconds and minutes. Our measurements were taken in 5-minute intervals. It can measure the number of bytes and packets sent and received, total number of bytes sent in both directions, latency, dropped packets, discard eligible packets, etc. For illustration we gathered traffic traces from a 56Kbps frame relay connection between a remote site and a server in the corporate office and imported the traffic data to our model. The topology is as follows:

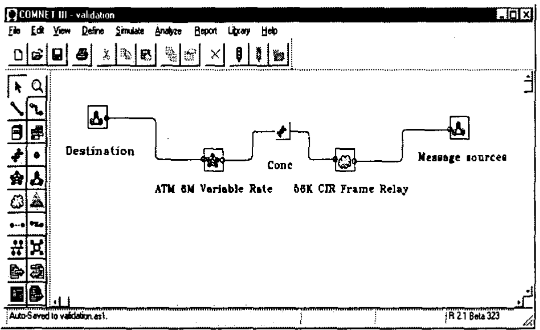


Fig 1. Part of the real network topology where the measurements were taken

The "Message sources" icon is a subnetwork that represents a remote site with a local area network, a local router, and a client A. The client is sending 440 bytes per second, on average to server B in the "Destination" subnetwork similarly to the real network. The Hurst parameter of the real traffic trace was H=0.55. The length of a message is determined by the Pareto distribution to generate bursty traffic. The Pareto distribution's location parameter is set to 208.42, the shape parameter is set to 1.9 that correspond to the mean number of bytes 440 and Hurst parameter 0.55. (The details of the calculations can be found in Appendix B). The Frame Relay icon represents a frame relay cloud with 56Kbps committed information rate (CIR). The "Conc" router connects the frame relay network to a 6Mbps ATM network with variable rate control (VBR). The "Destination" icon denotes a subnetwork with server B. The result of the model shows almost identical average utilization of the frame relay link (0.035~3.5%) to the utilization of the real measurements (0.031~3.1%), i.e., we start from a model that closely represents the real network (See Figure 2 below.) It can also be shown from the model's traffic trace that for the model generated messages the Hurst parameter H=0.56, i.e., the model generates almost the same bursty traffic as the real network.

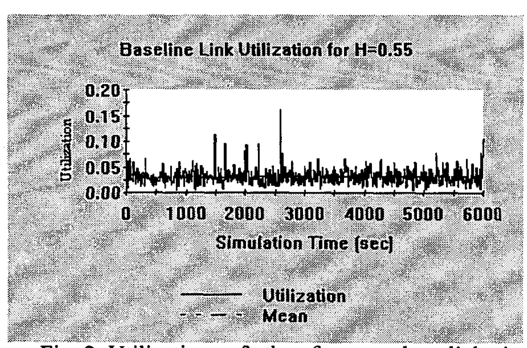


Fig. 2 Utilization of the frame relay link in the baseline model

3.1. Model results

After validating the baseline model we compared the response time for increasing simulated burstiness. The model traffic is generated with message sizes determined by various Hurst parameters (H=0.55, H=0.75, and H=0.95) and fixed size messages for comparison. The model demonstrates that increasing burstiness results in extremely high peaks in response time. We have to emphasize that the results are based on 5-minute polling time intervals. The interval is too wide to reveal the peaks of link utilization and response time in shorter time intervals, such as 5 seconds or even less. Although we could simulate very short time intervals we still need to validate the model based on short polling time intervals. We are going to repeat the same measurement when shorter time intervals are available in Concord Network Health. Therefore, the results below are only for demonstration purpose to illustrate the new methodology. However, it can be anticipated that the momentarily peaks will be even higher as the polling time intervals get shorter.

The "Message Source" icons in the figure below symbolize remote offices with clients. Each client transmits the same number of bytes per second as in the baseline model above but with different burstiness: H=0.55, H=0.75, H=0.95, and with fixed size:

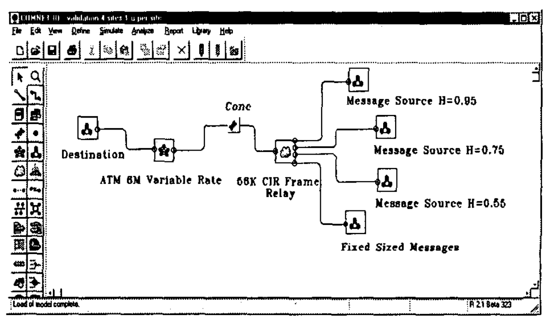


Fig 3. Network topology of bursty traffic sources with various Hurst parameters

The following table shows the simulated response time between the clients and the server in the different cases. It shows that the higher the burstiness the higher the response time. The charts in Appendix A clearly demonstrate the extremely high peaks in response times.

	Fixed messages	H=0.55	H=0.75	H=0.95
Average response time (ms)	75.960	65.61	87.880	311.553
Peak response time (ms)	110.06	3510.9	32418.7	112458.1
Standard deviation	0.470	75.471	716.080	4341.24

Table 1. Relation between response time and burstiness

We also measured the utilization of links and the number of cells dropped at a router's input buffer in the ATM network due to surge of bursty cells. We simulated the aggregated traffic of approximately 600 users each sending the same number of bytes in a second as in the measured real network. The ATM link is a DS-3 6Mbps connection between the Frame Relay cloud and the "Destination" subnetwork. The results of the simulation cannot be included in this paper due to space limitations. They can be obtained from the author upon request and will also be published in another paper.

4. CONCLUSION

The paper presented a discrete event simulation methodology to measure various network performance parameters while transmitting bursty traffic. It has been proved in recent studies that combining bursty data streams will also produce bursty combined data flow. The studies imply that the methods and models used in traditional network design require modifications. Traditional analytical methods cannot cope with the effects of random variance. By using discrete event simulation methodology we can get more realistic and accurate results in measuring network performance parameters. We implemented a new methodology and the Hurst parameter in the discrete event simulation package Comnet. The methodology has been proved to be an appropriate model to characterize bursty traffic. Using real network traces we built and validated a model that could measure and illustrate the impact of bursty traffic on the utilization of links and on response times between clients and servers in large, high-speed networks.

ACKNOWLEDGEMENT

The author would like to thank Bela Gyires for his useful help in preparation of this article.

Appendix A.

Measurements for response times

The following charts illustrate the relation between various Hurst parameters clearly demonstrating the extremely high peaks in response times.

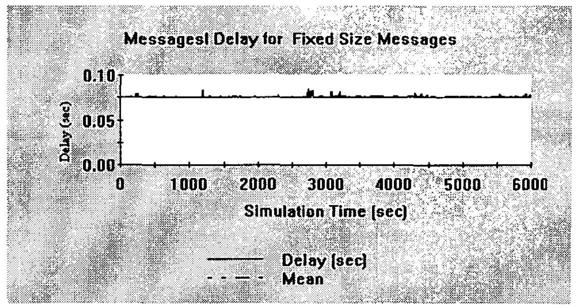


Fig. 4. Response time for fixed size message:

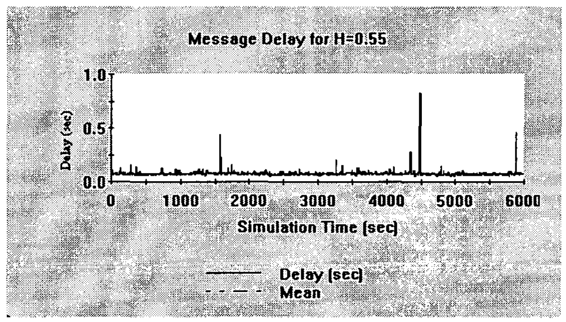


Fig. 5. Response time for H=0.55 (longer response time peaks)

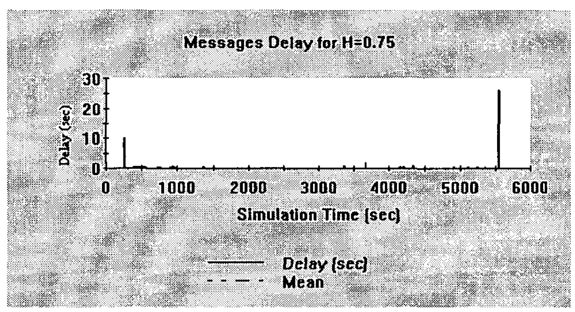


Fig. 6. Response time for H=0.75 (Long response time peaks)

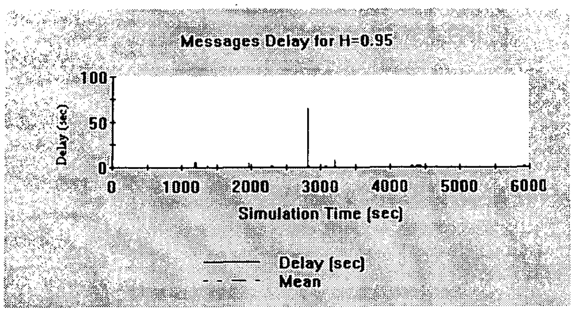


Fig. 7. Response time for H=0.95 (Extremely long response time peak)

Appendix B.

Integration of the methodology with the Comnet modeling tool

We make use of some of the calculations in the analytical model called M/Pareto presented in [3,13]. The M/Pareto model has been proved to be an appropriate model to characterize bursty traffic. It is a Poisson process for burst arrivals with rate λ. A burst generates packets with arrival rate δ. Each burst, from the time of its interval, will continue for a Pareto-distributed time period. The probability that a Pareto-distributed random variable X exceeds threshold x is

$$P(X > x) = (1+x)^{-\gamma} \quad 1 < \gamma < 2$$

The mean number of packets in a burst is $\delta/(\gamma - 1)$ and its variance is infinite. Assuming a t time interval, the mean number of packets in the time interval t is

$$\mu = \lambda t \delta / (\gamma - 1), \text{ and}$$

$$\lambda = \mu(\gamma - 1)/t\delta$$

It is shown in [3] that for the Hurst parameter H the following equation holds:

$$H = (3-\gamma)/2, \text{ i.e.,}$$

$$(1) \quad \gamma = 3 - 2H$$

We implement the Hurst parameter and a modified version of the M/Pareto model in the discrete event simulation system Comnet [2,9].

The Concord Network Health system can measure the traffic in certain time intervals at network nodes, such as routers and switches. It can also measure the number of bytes and packets sent and received per second, latency, dropped packets, discard eligible packets, etc. It cannot measure the number of packets in a burst and the duration of the bursts as it is assumed in the M/Pareto model above. Therefore, we slightly modify our traffic model according to the data available

Comnet models a transaction by a message source and a destination, the size of the message, and the communication devices, and links along the path. The rate at which messages are sent is specified by an interarrival time distribution. The M/Pareto model's Poisson distribution represents the number of message arrivals in a certain time interval. In Comnet this information is expressed as the time interval between successive arrivals, hence we use the Exponential distribution as the traditional way to specify interarrival time. Using the Exponential distribution will result in an arrival pattern characterized by the Poisson distribution. The interarrival time in the model is one second.

The length of a message is specified by the Pareto distribution defined by two parameters in Comnet: the location and the shape. The location parameter corresponds to the packet arrival rate δ in a burst. The shape parameter corresponds to the γ parameter of the M/Pareto model calculated in the relation (1) as

$$\gamma = 3 - 2H$$

In the M/Pareto model each burst will continue for a Pareto-distributed time period. The Concord Network Health cannot measure the number of packets in a burst, hence, we assume that a burst is equivalent to the number of bytes in a message sent or received in a second. Because the time period of each burst is proportional to the length of the message, we further assume that the length of the messages is also Pareto-distributed. So we derive the packet arrival rate δ in a burst not from the mean number of packets in a burst, but from the mean length of messages denoted by M. The mean of the Pareto distribution is implemented in Comnet [16] as:

$$M = \delta\gamma/(\gamma - 1)$$

and

$$\delta = M * (\gamma - 1)/\gamma$$

The relation (1) allows us to model bursty traffic based on real traffic traces by performing the following steps:

a. Collect traffic traces using the Concord Network Health network analyzer
b. Compute the Hurst parameter H by making use of the Benoit package with the traffic trace as input
c. Use the Exponential and Pareto distributions in the COMNET modeling tool with the parameters calculated above to specify the interarrival time and length of messages
d. Generate traffic according to the modified M/Pareto model and measure network performance parameters.

The traffic generated according to the steps above is bursty with parameter H calculated from real network traffic.

6. REFERENCES

[1] Leland, W., Taqqu, M., Willinger, W., and Wilson, D. W., "On the Self-Similar Nature of the Ethernet Traffic," IEEE Transactions on Networking, Vol. 2, no. 1, February 1994, pp.1-15.

[2] "Comnet III Application Notes: Modeling ATM Networks with Comnet III" CACI Products Company, 3333 North Torrey Pines Ct., La Jolla, California 92037, 1998.

[3] R. G. Addie, M. Zukerman, and T. D. Neame, "Broadband Traffic Modeling: Simple Solutions to Hard Problems," IEEE Communications Magazine, August, 1998, pp.88-95.

[4] P. J. Davy, "Self-similar processes with non-negative stationary increments," submitted to J. of Stat. Planning and Inference.

[5] S. Molnar, A. Vidacs, and A. Nilsson, "Bottlenecks on the Way towards fractal characterization of Network traffic: Estimation and Interpretation of the Hurst Parameter."

[6] V. Paxson, S. Floyd, "Wide Area Traffic: The Failure of Poisson Modeling," IEEE/ACM Transactions on Networking, Vol. 3, NO. 3, June 1995, pp.226-244.

[7] M. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for loon-range dependence: an empirical study," Fractals, Vol. 3, No. 4 1995, pp. 785-788.

[8] "Enterprise Profiler," CACI Products Company, 3333 North Torrey Pines Ct., La Jolla, California 92037, 1998.

[9] "COMNET III, Planning for Network Managers," Release 1.4, 1997 CACI Products Company, 3333 North Torrey Pines Ct., La Jolla, California 92037, 1997.

[10] T. Neame et al., "Investigation of traffic models for high speed data networks," Proceedings of ATNAC '95, 1995.

[11] Gyires, T., "Smart Permanent Virtual Circuits in Asynchronous Transfer Mode Networks," invited paper in the Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics, October 12-15, Orlando, Florida, pp. 4460-4464.

[12] Gyires, T., "Performance Prediction of Smart Permanent Virtual Circuits in ATM Networks with CACI Comnet," in the Proceedings of the 1998 IEEE International Conference on Systems, Man, and Cybernetics, Intelligent Systems For Humans In A Cyberworld, October 11-14, 1998, Hyatt Regency La Jolla, San Diego, California, pp.3993-3998.

[13] Likhanov N., Tsybakov, B., Georganas, N.D., "Analysis of an ATM Buffer with Self-Similar ("Fractal") Input Traffic," in the Proceedings of the IEEE INFOCOM Conference, 1995, Boston, pp. 985-992.

[14] Benoit 1.1, Trusoft Int'l Inc., www.trusoft-international.com

[15] "Network Health Trend Reports," collected by Tibor Gyires, 1999.

[16] Conversations with Vipul Pabari and Ken Neumeister, CACI Products Company, 1998.