Raj Jain is now at Washington University in Saint Louis, jain@cse.wustl.edu http://www.cse.wustl.edu/~jain/

The presentation of this contribution at the ATM Forum is sponsored by NASA.

Date: October 1995, Honolulu

The Ohio State University

Distribution: ATM Forum Technical Working Group Members (Traffic Management)

Notice: This contribution has been prepared to assist the ATM Forum. It is offered to the Forum as a basis for discussion and is not a binding proposal on the part of any of the contributing organizations. The statements are subject to change in form and content after further study. Specifically, the contributors reserve the right to add to, amend or modify the statements contained herein.

Most of the data traffic on computer networks is expected to be bursty. The clients generally send a request to the server. The server responds and then the clients waits for a some time before sending the next request. This type of traffic is very different from infinite source traffic that has been studied widely in the traffic management group. Our goal in this study was to see how bursty sources behave under current TM 4 rules.

We define a new "closed loop" bursty traffic model. The model consists of cycles and in each cycle a source sends a set of requests and receives a set of responses from the destination. The next cycle begins after all the responses of the previous cycle have been received and an inter cycle time has elapsed. There is a gap between successive requests. It is called inter-request time. The request contains a bunch of cells sent back to back by the source at PCR. The adapter controls the output rate to ACR, however. The model as presented above represents Web traffic, transaction oriented traffic, or request-response traffic. The model is "closed loop" in the sense that the rate at which cycles (and hence requests) are generated depends upon the responsivity of the network. If the network is congested, the response take longer time to come back and the sources do not generate new requests until the previous ones have been responded. This is more realistic than an "open loop" model where bursts are generated at a fixed rate regardless of the congestion in the network.

We must point out that the packet train model [1] used previously to model bursty behaviour is an open loop model. This model has the source sending out cars (having back to back packets) with an intercar time and trains having an inter-train time. But, the source does not wait for any response from the destination.

There has been a tremendous growth in client-server traffic (like NFS request/responses), WWW traffic (using applications like xmosaic and netscape) and interactive gaming traffic. Our closed loop model is directed to capture such traffic patterns and their aggregates.

Our main reason for examining bursty sources was to see if there is throughput degradation due to idle periods between bursts. The interaction of such idle periods with the timeout mechanisms like TOF and Xrm needs to be studied. During idle periods, a source may be treated as a low rate source and the effects of ICR, TCR and rescheduling are pronounced.

In addition to the usual metrics of ACR, queue lengths, utilization, Cells received per VC at destinations we use two additional metrics to characterize the performance of bursty sources. These are: Burst Throughput and Average Burst Throughput. The new metrics capture the response during the burst periods ignoring the idle periods.

# cells in burst

sigma\_i(# cells in burst\_i)
Avg Burst Throughput = -----sigma\_i( exit\_time(last\_cell\_of\_burst\_i) - exit\_time(first\_cell\_of\_burst\_i) )

Note: The avg burst throughput is not the same as 1/n \* sigma (burst\_throughput\_i)

The parameters of the model are the request size, the inter-request time, the response time (latency from the time the request is received till the corresponding response is sent), the response size, the cycle size (number of requests per cycle), and the inter cycle time.

Note that the inter response time (not a parameter) is a function of the inter-request size and network congestion. The time between the start of cycle\_i and cycle\_(i+1) is also a function of the inter request times, the network RTT and the network congestion. Interesting situations arise as the request and response sizes (in terms of time) are in the range of RTT. Initially we consider simple cases where cycle size is 1.

We will present simulation results at the Forum meeting.

[1] R. Jain and S. Routhier, ``Packet Trains - Measurement and a new model for computer network trafic,'' {\m IEEE Journal of Selected Areas in Communications,} Vol. SAC-4, No. 6, September 1986, pp. 986-995. Reprinted in Amit Bhargava, Ed., ``Integrated Broadband

Networks'' Artech House, Norwood, MA, 1990.