# CONGESTION CONTROL TECHNIQUES FOR PACKET NETWORKS

Simon S. Lam

Department of Computer Sciences
University of Texas at Austin

Abstract

Congestion control mechanisms are needed for store-and-forward packet networks to maintain a high level of throughput. The basic requirements of network congestion control and some specific control techniques are examined. The use of input buffer (IB) limits for network congestion control is introduced. The rationale for the effectiveness of IB limits is discussed. Strategies for the design of IB limits have been investigated using both queueing analysis and simulation experiments. Some of our preliminary findings are presented.

## INTRODUCTION

Consider a store-and-forward packet communication network of N nodes. Each node is connected to a multiplicity of packet sources and sinks. A source generates data packets (at some rate) which may be destined for any sink in the network. Routes followed by packets may be fixed or adaptive. If a packet arrives at the node connected to its destination sink, the packet will be consumed in a finite amount of time. Each node must hold on to a packet that it has accepted until the packet has been successfully consumed or forwarded to the next node on its route, at which time its buffer can then be freed. Each node has a finite number of packet buffers. When all these buffers in a node are in use, packets offered by its sources or forwarded to it by its neighbors are discarded; packets being forwarded by its neighbors will, of course, be retransmitted later. The network throughput is defined to be the aggregate rate at which packets are consumed under the assumption of statistical equilibrium conditions.

Store-and-forward packet networks without flow control mechanisms have been shown to exhibit the throughput-load relationship illustrated in Figure 1 [KAHN 72, DAVI 72, PRIC 77, GIES 76, KLEI 78]. A characteristic, typical of many contention systems, is that as the offered load is increased from zero, the network throughput increases to a maximum and then turns down and decreases sharply to a low value (possibly zero). Realistically, when a store-and-forward network is congested, some "processes" may be blocked, and data packets may be lost or delayed due to a lack of resources [OPDE 74]. In either case, work is not conserved; hence, the degradation in network throughput.

Degradation in network throughput is often caused by deadlocks [KAHN 72, DAVI 72, OPDE 74]. However, networks which are deadlock-free may still be degraded in the sense that the throughput, though nonzero, is relatively low [GIES 76]. Hence, flow control mechanisms are needed to prevent throughput degradation whether or not a network can be formally proved to be deadlock-free.

From now on, a network is said to be <u>congested</u> when it operates in the region of negative slope in Fig. 1.

## NETWORK VERSUS END-TO-END CONTROL

By network congestion control we mean any mechanism with the primary objective of preventing the network from operating in the congested region for any significant period of time. Typically, most networks are based upon the concept of virtual channels (or connections, sessions etc.) and are end-to-end flow-controlled between pairs of sources and sinks. Examples of end-to-end controls are SNA pacing [IBM 75], RFNM in the ARPANET [OPDE 74], and various window mechanisms [POUZ 73, CERF 74]. An important function of such end-to-end controls is synchronization of the source input rate to the sink acceptance rate. All of them work by limiting the number of packets permitted in a virtual channel. Suppose $L_i$ is the maximum number of packets in virtual channel i and the network has a total of K virtual channels. The maximum number of packets permitted to enter the network is thus

$$N_{max} = L_1 + L_2 + \ldots + L_K.$$

The fact than $N_{max}$ is bounded does not imply that network congestion control is not necessary. In fact, one of the motivations for a store-and-forward network in the first place is that data traffic sources are typically bursty [LAM 78]. In other words, virtual channels require actual transmission capacity only intermittently with a small duty cycle. If, for example, a network is operated such that $N_{max}$ is at point B in Figure 1, it is obvious that network congestion control is not necessary. However, due to the bursty traffic, the average utilization of the network will be very low (such as at point A). It is therefore desirable for store-and-forward networks to operate on the principle of <u>overcommitment</u> such that $N_{max}$ is far to the right (such as at point C) in Figure 1 and through averaging, the network utilization is at point B with a correspondingly high throughput. An immediate consequence is that network congestion control is now necessary to prevent the network operating point from going over the peak of the curve as a result of statistical fluctuations.

## NETWORK CONGESTION CONTROL TECHNIQUES

A network congestion control mechanism must be capable of: (1) detection of network congestion, and (2) shutting off input into the network according to some rule.

The isarithmic technique proposed by DAVIES [DAVI 72] and studied by Price [PRIC 77] does the above functions by limiting the number of packets permitted to enter the network. This is accomplished by circulating a fixed number of "containers" in the network. A newly generated packet will be accepted by a network node only if it can get hold of an empty container. A second technique is to control the window size of each virtual channel as a function of network load [KERM 77]. Both techniques are attractive in theory but are quite difficult to implement in practice because individual network nodes do not have fresh and accurate information about the rest of the network.

A third technique that we have studied, assumes some knowledge about the pattern of traffic flows in the network. It attempts to control network inputs by differentiating between input and transit traffic at each node and imposing a limit on the fraction of buffers in a node's buffer pool that input traffic can occupy. This fraction will be referred to as the <u>input buffer limit</u>. Note that transit traffic can occupy all buffers in the buffer pool. In times of extreme congestion, input traffic may be shut out by transit traffic but not vice-versa; a desirable property. The advantage of discriminating against input traffic was first noticed by Price [PRIC 77]. He observed that if one or two buffers are dedicated to transit traffic, the network throughput can be much improved. A similar idea was also suggested by Chou and Gerla [CHOU 76]. This idea, however, is most clearly demonstrated and investigated in the GMD simulation studies [GIES 76]. In addition, they have also shown that if the buffer pool is structured into nested subsets of buffers and packets are assigned to these subsets according to the number of hops they have covered, then it can be proved that store-and-forward deadlocks of the type described in [KAHN 72, OPDE 74] can be avoided.

## DESIGN STRATEGIES FOR IB LIMITS

Although the GMD studies explored the use of input buffer (IB) limits for congestion control, it was not known how to design such limits. We have studied design strategies for IB limits using both queueing analysis and simulation experiments. Our objective is to achieve the maximum network throughput as well as to provide some safety margin for errors and uncertainties in our traffic flow assumptions.

We employed an analytic model based upon our earlier work on the modeling of store-and-forward networks [LAM 76] and an extended class of queueing networks [LAM 77]. Using the analytic model, we were able to study the tradeoffs among network offered load, nodal buffer capacity and IB limit. From these results, we found that the key to designing effective IB limits hinges upon a very intuitive "capacity law" [LAM 79].

In Fig. 2, we have shown analysis results for a homogeneous network of identical nodes. Fixed routing is used. Let $\sigma_{max}$ be the theoretical maximum network throughput rate assuming infinitely many nodal buffers; $\sigma_{max}$ can be easily calculated from the channel capacities. $\lambda$ is the rate of offered load to the network (i.e. aggregate packet generation rate of all sources). $N_T$ is the number of buffers in a node. An important observation from the analysis results in Fig. 2 is that if $\lambda < \sigma_{max}$, IB limits are not necessary. If $\lambda > \sigma_{max}$, then IB limits (<1 in value) are necessary for congestion control. In fact, under such a heavy loading condition, IB limits are what determines the amount of traffic that can be admitted into the network. The results in Fig. 2 clearly show that there is a critical value of IB limit beyond which the traffic carrying capability of the network is seriously impaired. We shall refer to this critical value as the IB capacity.

The explanation for this critical behavior turns out to be rather intuitive. For each new packet that the network admits into an input buffer, additional buffers are needed elsewhere for its subsequent journey to the destination. Therefore, there is a natural ratio, say $\alpha_0$, of the number of input buffers to the number of total buffers in the network that serves as an upper bound for IB limits.

For a homogeneous network with fixed routing and fixed input traffic pattern, the IB capacity is identical for each node and the ratio $\alpha_0$ can be easily determined [LAM 79]. ($\alpha_0$ is equal to 0.344 for our example in Fig. 2.) If IB limits exceeding $\alpha_0$ are used, it will then become possible for input buffers to become completely filled such that the network will not have enough buffers to accommodate the resulting transit traffic! This explains the critical behavior observed in Fig. 2.

Another important observation from Fig. 2 is that if the IB limits are properly designed, the offered load $\lambda$ can go to infinity without degrading the network throughput! Note that the throughput curves in Fig. 2 all have a wide "plateau" around the maximum point, meaning that the IB limit can be made much smaller than the IB capacity without sacrificing much throughput. Thus a large safety margin can be provided for errors in our evaluation of the IB capacity. Errors may arise due to time and statistical flucuations in user traffic as well as uncertainties in our estimate.

The above observations from analysis results are supported by simulation results of a four-node homogeneous networks. In Fig. 3, results from our analysis and two different simulators are shown. The first simulator assumes fixed-length packets. The throughput results from this simulator (consider the curve $\lambda = 2.2$, $N_T=30$) are actually better than the analytic throughput results which assume exponentially distributed packet lengths. The second simulator uses exponentially distributed packet lengths (but no Independence Assumption [KLEI 64]) and give throughput values which are slightly less than the analytic results.

Note that the part of an analytic throughput curve with negative slope is not realizable in simulation. This is because the assumption of equilibrium used in the analysis breaks down under such congested conditions. In practice, the network should not operate in this region.

Another very interesting observation which we have not yet been able to provide a satisfactory explanation is: if retransmissions are given priority over the transmission of other messages, the network behaves much better in terms of a higher throughput and a larger IB capacity.

For a nonhomogeneous network, the network nodes will have different IB capacities. Furthermore, for networks with adaptive routing, the IB capacities will not be fixed. These problems are currently being studied. Some heuristic methods for desiging IB limits have been proposed.

A promising approach is under investigation for networks based upon virtual channels. (We note that the new generation of packet networks are mostly virtual channel networks [ROBE 78]. In fact, the CCITT X.25 international standard for network interface is designed for virtual channel networks.) Virtual channels are individually end-to-end flow controlled via a window mechanism. Knowing the set of virtual channels and their window sizes, we use a queueing network model to obtain estimates of (peak) traffic flows in the network. From these estimates, we have developed various heuristic assignment strategies for designing IB limits.

Simulation results are shown in Fig. 4 for a 7 node (nonhomogeneous) network with 84 virtual channels. Each network node has 30 packet buffers. Each virtual channel uses a fixed route. The window size of each virtual channel is set equal to 3 times the number of hops (links) in its route. The network throughput is plotted versus the ratio $N_I/N_T$, where $N_T$ is the total number of buffers and $N_I$ is the total number of input buffers in the whole network. Two assignment strategies are considered: a uniform assignment strategy which uses the same IB limit at each node and a heuristic assignment strategy that uses IB limits proportional to our estimates of traffic flows.

We found that both assignment strategies achieve approximately the same maximum network throughput. However, the IB capacity of the heuristic assignment strategy is better than that of the uniform assignment strategy. The heuristic assignment strategy also gives rise to a lower value of mean network delay for packets at the same level of network throughput.

However, at small values of $N_I/N_T$, say 0.1 to 0.2 in Fig. 4, we found that the heuristic assignment strategy does not perform well because of the small number of input buffers available in the network; some IB limits become excessively small according to the heuristic assignment strategy thus restricting the network throughput.

The following strategy appears to be a reasonable compromise. Each IB limit consists of two components: (1) the first component is fixed and guarantees a minimum acceptance rate at the node, (2) the second component is designed proportional to network traffic flow estimates.

## CONCLUSION

We discussed the need for congestion control mechanisms in store-and-forward packet communication networks. The rationale for the effectiveness of IB limits as a congestion control mechanism was examined. Design strategies for IB limits have been investigated using both queueing analysis and simulation experiments. Some of our preliminary findings are presented.

## ACKNOWLEDGEMENT

This work was supported by National Science Foundation under Grant No. ENG78-01803.

## REFERENCES

CERF 74 Cerf, V. and R. Kahn, "A protocol for Packet Network Intercommunication," IEEE Trans. on Commun., Vol. COM-22, May 1974.

CHOU 76 Chou, W. and M. Gerla, "A Unified Flow and Congestion Control Model for Packet Networks," Proc. Third International Conf. on Computer Communication, Toronto, August 1976.

DAVI 72 Davis, D. W., "The Control of Congestion in Packet Switching Networks," IEEE Trans. on Commun., Vol. COM-20, June 1972.

GIES 76 Giessler, A., J. Haenle, A. Koenig and E. Pade, "Packet Networks with Deadlock-Free Buffer Allocation--An Investigation by Simulation," GMD Report, Darmstadt, Germany, 1976.

IBM 75 IBM Corp., Systems Network Architecture General Information, GA27-3102-0, January 1975.

KAHN 72 Kahn, R. and W. Crowther, "Flow Control in a Resource-Sharing Computer Network," IEEE Trans. on Commun., Vol. COM-20, June 1972.

KERM 77 Kermani, P., "Switching and Flow Control Techniques in Computer Communication Networks," Ph.D. Dissertation, Computer Science Department, UCLA, 1977.

KLEI 64 Kleinrock, L., Communication Nets: Stochastic Message Flow and Delay, McGraw Hill, New York, 1964.

KLEI 78 Kleinrock, L., "On Flow Control in Computer Networks," Conf. Rec. ICC'78, Toronto, June 1978.

LAM 76 Lam, S. S., "Store-and-Forward Buffer Requirements in a Packet Switching Network," IEEE Trans. on Commun., Vol. COM-24, April 1976.

LAM 77 Lam, S. S., "Queueing Networks with Population Size Constraints," IBM Journal of Res. and Develop., Vol. 21, July 1977.

LAM 78 Lam, S. S., "A New Measure for Characterizing Data Traffic," IEEE Trans. on Commun., Vol. COM-26, Jan. 1978.

LAM 79 Lam, S. S. and M. Reiser, "Congestion Control of Store-and-Forward Networks by Input Buffer Limits -- An Analysis," IEEE Trans. on Commun., Jan. 1979.

OPDE 74 Opderbeck, H. and L. Kleinrock, "The Influence of Control Procedures on the Performance of Packet-Switched Networks," Proc. National Telecommunications Conference, San Diego, California, Dec. 1974.

POUZ 73 Pouzin, L., "Presentation and Major Design Aspects of the CYCLADES Computer Network," Data Networks: Analysis and Design, Third Data Communications Symposium, St. Petersburg, Florida, Nov. 1973.

PRIC 77 Price, W. L., "Data Network Simulation Experiments at the National Physical Laboratory 1968-1976," Computer Networks, Vol. 1, 1977.

ROBE 78 Roberts, L. G., "The Evolution of Packet Switching," Proc. IEEE, Vol. 66, Nov. 1978.

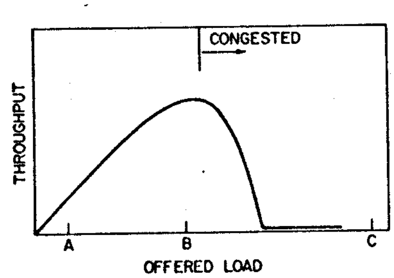


Fig. 1. Throughput versus offered load.

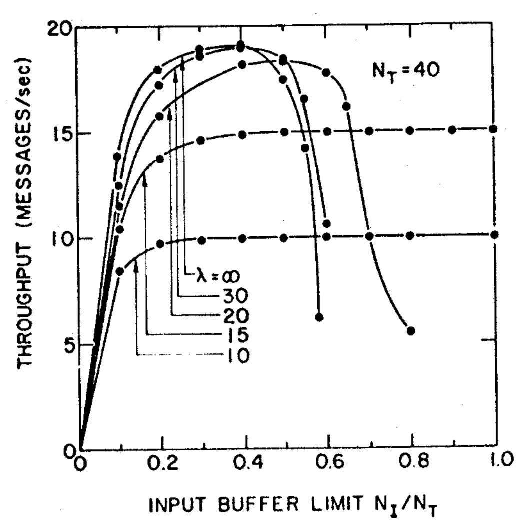


Fig. 2. Throughput versus IB limit.

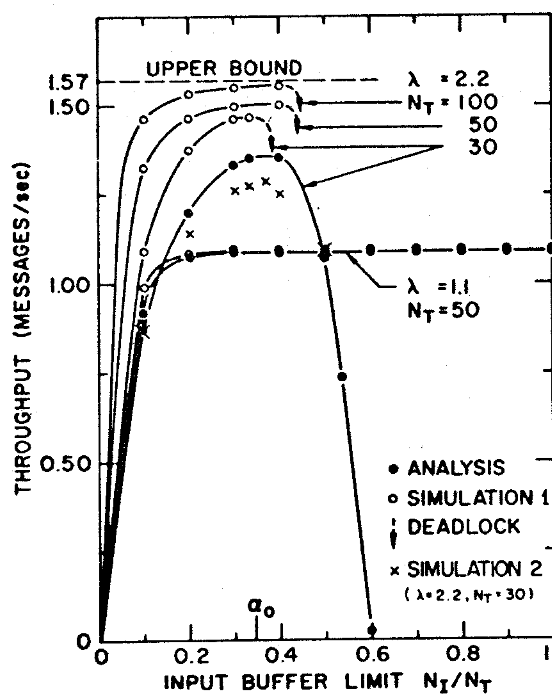


Fig. 3. Analysis and simulation results (four-node homogeneous network).

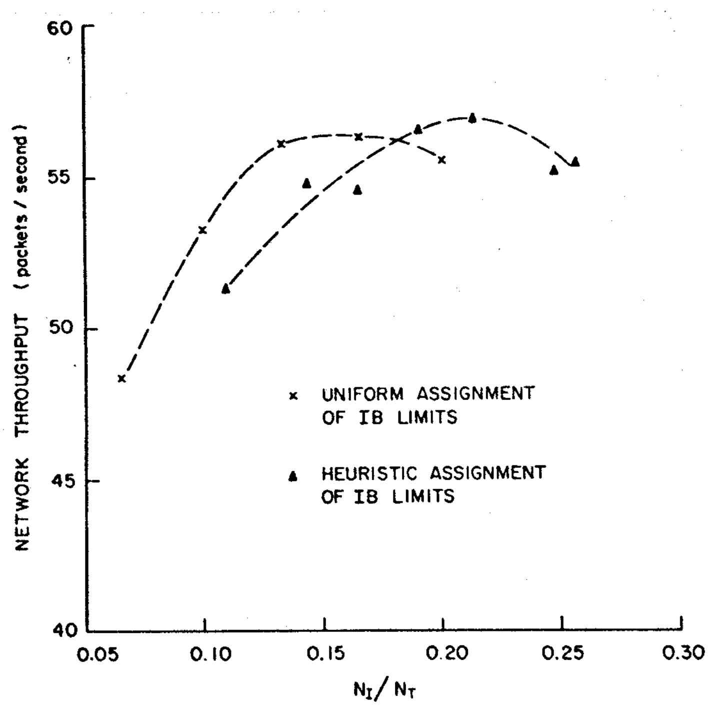


Fig. 4. Simulation results (seven-node nonhomogeneous network).