****

ATM Forum Document Number: ATM Forum/94-0692

****

Title: Congestion Control With Explicit Rate Indication

****

Abstract:

In a connectionless network, the intermediate systems (such as switches or routers) have very little information about active sources and so they cannot communicate directly to the source its share of the available resources. Therefore, the sources have to increase or decrease their load based on simple indication, such as overload or underload (using a binary feedback) from the switches. In a connection-oriented network, on the other hand, the switches have a lot more information and so there is no need for sources to do a long search using limited (one-bit) feedback.

A scheme that lets sources quickly adapt to the correct rates is described.

****

Source:

Anna Charny
Digital Equipment Corp
550 King St.
Littleton, MA 01460-1289
Phone: 508-486-5862
Email: Charny@nac.enet.dec.com

David D. Clark
Laboratory for Computer Science
Massachusetts Institute of Technology
545 Tech Square (NE43)
Cambridge, MA 02139
Phone: 617-253-6003
Email: DDC@lcs.mit.edu

Raj Jain
The Ohio State University
Department of CIS
Columbus, OH 43210

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

****

Date: July 1994

****

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

****

is not a binding proposal on the part of any of the contributing companies. 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.

****

INTRODUCTION

In the past decade, the DECbit scheme has become very popular due primarily to its simplicity. The scheme requires the switches to monitor their load and set a bit in the packets if overloaded. The end-systems monitor the sequence of bits returned from the network and adjust their load up or down.

The original DECbit scheme was implemented in DECnet Phase V for window based flow control. ISO transport and connectionless network layer protocol (CLNP) standards contain mechanisms to allow sources and networks to adjust the load using binary feedback. Several rate-based versions of the scheme have been developed and are used in frame-relay and fast packet networks. The Explicit Forward Congestion Notification (EFCN) schemes envisioned for ATM networks are also based on the eventual adoption of some binary feedback scheme for ATM networks. In fact, the rate-based scheme currently being considered by the traffic management group is based on concepts originating from the DECbit scheme work.

The DECbit scheme was designed to solve congestion problems for connectionless networks. The problem for connection-oriented networks is simpler. In particular, the switches have a lot more information about the flows and can provide more information than was possible in connectionless environment of DECbit.

The scheme described in this contribution has been designed specially for the connection-oriented environment. In particular, the switches provide an explicit feedback about the rate at which the source should send its data traffic. The scheme as described here is based on the Master's Thesis work of Anna Charny at MIT and was originally formulated in the context of packet switching. However, it can be easily adapted for the use in ATM networks.

OVERVIEW OF THE SCHEME

In the original packet-switching version of the scheme the data packets contain control information used for congestion management. In the ATM environment special resource management cells can be used for this purpose. In what follows we refer to these cells as "control cells".

Control cells contain two special fields. The first field is one bit long and is called "Reduced bit". The second field is several bits long and contains a rate estimate, which will be referred to as "stamped rate." Each switch monitors its traffic and calculates its available capacity per flow. This quantity is called the "advertised rate".

The source puts its current rate estimate in the stamped-rate field and clears the reduced-bit. Initially, the stamped rate is simply the desired rate of the source. The switches compare the "stamped rate" with their "advertised rate".

If the "stamped rate" is higher than or equal to the "advertised

rate", the stamped rate is reduced to the "advertised rate" and the reduced-bit is set. If the stamped rate is less than the advertised rate, the switch does not change the fields of the control cell.

When the control cell reaches the destination, the "stamped rate" contains the minimum of the source's rate estimate (at the time the control cell was sent) and the best rate that the flow is allowed to have by the switches in its route. The destination sends the control cell back to the source. After a full round trip, the setting of the reduced bit indicates whether the flow is constrained along the path. That is, if the reduced bit is set, the rate is limited by some switch in the path and cannot be increased. In this case the source adjusts its rate to the "stamped rate" of the control cell. If the reduced bit is clear, the source can increase its rate estimate and as a result its "stamped rate" is increased to its desired value.

Note that the value of the "stamped rate" the source writes in the control cell does not have to reflect the actual transmission rate at all times. In particular, if the desired value is unknown or very large, the actual transmission rate is not increased when the reduced bit is clear, while the "stamped rate" is set to the large value. If the reduced bit is set, however, the actual transmissionrate can be adjusted to be the same as the "stamped rate". When all flows stabilize to their optimal rates the reduced bit will always be set when a contol packet returns to the source. The advertised rate of the switches varies as new flows are started or the old flows are closed. However, at all times, the switches attempt to give all available capacity fairly to all flows. The response is fast since the sources come to know the rate of the tightest bottleneck in one round-trip delay. If there is only one bottleneck (or several bottlenecks with identical per flow capacities), one roundtrip is sufficient for all sessions to obtain their optimal rates. If there are several bottlenecks of different bottleneck capacity, then additional roundrips can be required for sessions sharing the "higher-bandwidth" bottlenecks to recapture the capacity unused by sessions constrained by "tighter" bottlenecks.

The switches maintain a list of all of its flows and their last seen stamped rates. All flows whose stamped rate is higher than the switch's advertised rate are considered "overloading flows." Similarly, flows with stamped rate below the advertised rate are called "underloading flows." The underloading flows are bottlenecked at some other switch and, therefore, cannot use additional capacity at this switch even if available.

The capacity unused by the underloading flows is divided equally among the overloading flows. Thus, the advertised rate of the flows is calculated as follows:

$$\text{Advertized rate} = \frac{\text{total capacity-sum of bw of underloading flows}}{\text{total no. of flows - no. of underloading flows}}$$

More specifically, upon receipt of a new control cell with the "stamped rate" below the current "advertised rate", the switch preliminary marks the flow as underloading. Then, the advertized rate is calculated as shown above. However, it can be the case that after this calculation some flows that were previously underloading with respect to the old advertised rate can become overloading with respect to the new advertized rate. In this case these flows are re-marked as overflowing and the advertised rate is recalculated once more. It can be shown that the second

recalculation is sufficient to ensure that any flow marlked as underloading before the second reacalculation remains underloading with respect to the newly calculated advertised rate.

Some sources cannot use additional bandwidth for some reason. For example, if the source itself is a bottleneck. In that case, it starts the packet off with the "reduced bit" set. In this case, subsequent switches in the path cannot increase the rate but they can still decrease the rate.

PROPERTIES OF THE SCHEME:

1. There are no oscillations. The sources get to the fair share of the tightest bottleneck rate in their route within one roundtrip delay and then converge to the final optimum rate in a bounded and predictable time. If there is only one bottleneck in the network, one roundtrip delay is sufficient to obtain the final optimal rates. Once the optimal rate is reached, the sources stay at that rate until bandwidth supply or demand changes.

2. The initial rate does not have a significant impact. The sources can start at their link rate. The overload will last for only one round trip. In a bit-based scheme, several roundtrips are typically required to reduce to the right bottleneck rate even for a one-bottleneck network, since the information provided to the sources is a binary value, and the sources have to be conservative in their action to increase or decrease.

3. Each computation of advertised rate may require more than one iteration. In particular, if an additional bit per flow is used by the switches to retain memory of whether a flow was previously considered underloading, then at most two recomputations are required to obtain correct value of the advertised rate. However, if the engineering choice is made not to maintain the bit per flow to retain the memory of wheteher a particular flow was previously underloading, then the correct advertised rate can still be calculated. In this case several iterations may be needed. The latter approach is similar to the calculation of fair allocation of the Selective Binary Feedback Scheme presented in [SBF].

4. Policing misbehaved sources is easy. Since the rate feedback is explicit, the entry gateways can monitor the stamped values and enforce for admission control.

5. The scheme is particularly suited for connection-oriented networks, when there is only one route for a particular flow. However, the scheme can also be used if the route is allowed to change, as long as these changes do not occur often.

6. The scheme is shown to converge to correct optimal rates from any initial conditions on the flow rates and the state of the switches. This makes the scheme extremely robust in the peresense of any past errors, control cell loss and dynamic changes in network load.

SIMULATION RESULTS

In the forum presentation, we will present simulation results for a number of cases. All these results are also described in the MIT Technical report [Charny94]. The report is available by sending email to charny@nac.enet.dec.com or by requesting it from

the MIT library.

REFERENCES:

[Charny94] Anna Charny, "An Algorithm for Rate Allocation in a Cell-Switching Network with Feedback", MIT TR-601, May 1994.

[SBF] K.K. Ramakrishnan, Raj Jain, Dah-Ming Chiu. "Congestion Avoidance in Computer Networks With a Connectionless Layer. Part IV: A Selective Binary Feedback Scheme for General Topologies Methodology," DEC-TR-510, Digital Equipment Corporation. 1987.