

# Closed-Loop Congestion Control Protocol Design for Elastic Traffic

Omar Ait-Hellal  
Computer Science Department,  
Brooklyn Campus, Long Island University  
1 University plaza, Brooklyn, New York 11201-8423  
E-mail : Omar.Ait-Hellal@liu.edu  
Tel: 1 718 488 1272, FAX: 1 718 780 4145

## Abstract

*In the present paper, we outline few characteristics that every viable congestion control protocol for elastic traffic should satisfy. We provide and outline ways by which these characteristics can be measured and verified. We'll focus in our illustrations on protocols designed for ABR (Available Bit Rate) traffic; nonetheless transmission protocols designed for the Internet (IP) can be easily evaluated using our methods and verifications.*

**Keywords:** *Congestion control, closed-loop protocols, stability, convergence, fairness.*

## 1 introduction

Many Protocols for elastic traffic have been proposed and argued to be stable, convergent and fair. Usually these claims are validated by extensive well chosen simulations showing the excellent sides of the protocol. Very often, a simple analytic model scheme, as those we'll present later, could show an important flaw in the protocol, eg. non convergent protocol, unfairness in common situations, or unstable behavior for some parameters.

In the present paper we propose a general method to evaluate analytically closed-loop congestion control protocols. The sending rate, in such protocols, is adjusted to buffer occupancy or to some feedback (eg. fair share) from the intermediate nodes or from the destination. Stability (in terms of oscillations), Fairness and Full Utilization are considered. The later represent the basic and essential criteria every closed-loop congestion control protocol should satisfy.

Few examples where at least one of the above criteria is not satisfied are described and analyzed using an analytic method and simulations confirming our analytical results are given. The examples chosen in this paper are important algorithms that seem to have had significance in terms of implementations, standardization or discussions.

We should mention that there has been some related research on the stability of congestion control in networks with delayed information [6, 9, 13, 14, 15] yet in the context of the Internet using tools from delayed differential or difference equations. That suggests that the methods presented in this paper can be easily applied to the Internet (IP).

## 2 What a closed-loop protocol should satisfy?

In the sequel, we shall use the following simple model where multiple sources, sending to the same destination, share a single bottleneck node (as depicted in the figure (1)).

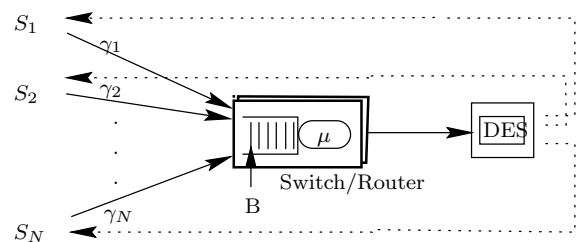


Figure 1: The model with multiple connections sharing a same bottleneck link

Without loss of generality, We will consider only the case where the rate is dictated by the intermediate node(s) (explicit rate); The buffer occupancy may well be a control parameter when the rate is computed. In the case of ABR for instance, intermediate switches compute the fair share for every source, then the computed fair share is conveyed back to the sources through special RM (Resource Management) cells. Upon reception of a backward RM cell (i.e. an RM cell sent by the source itself and made a round-trip back to the source), the source updates its rate to the ER (Explicit Rate) value found in the cell (i.e.  $ACR(t) = ER_i(t - \beta_i)$ ). The fair share attributed by the switch to a given source  $i$  ( $ER_i(t)$ ) is most often (if not always) computed based on the number of connections ( $N$ ) using the same link, the cumulative input rate of the ABR traffic ( $\sum_{i=1}^N CCR_i(t - \gamma_i)$ ), the input rate of connection  $i$  ( $CCR_i(t - \gamma_i)$ ), the output buffer size at shared output ( $y(t)$ ) and the estimated available bandwidth for the ABR traffic ( $\mu$ ), that we assume constant.  $\beta_i$  is the time it takes a cell from the bottleneck (queue) to reach source  $i$ , and  $\gamma_i$  is the time it takes a cell from source  $i$  to reach the bottleneck (queue) ( $\tau_i \triangleq \beta_i + \gamma_i$ ).

The behavior of the system described above can be easily put into the following equations ( $x_i(t) \triangleq CCR_i(t)$ ):

$$\left\{ \begin{array}{l} x_i(t) = f(x_1(t - \beta_1 - \gamma_1), \dots, \\ x_N(t - \beta_N - \gamma_N), y(t - \beta_i), N, \mu, B) \\ \dot{y}(t) = \sum_{i=1}^N x_i(t - \gamma_i) - \mu \end{array} \right. \quad (1)$$

**Remark 2.1** *Protocols that don't use the buffer occupancy as a control parameter are not stable, if they aim to fully (100%) use the available bandwidth [3](eg. ERICA).*

## 2.1 Stability

For equivalent sources, from the above equations it is always possible to express  $y(t)$  as a first order delayed differential equation. The obtained equation can be always put in the following form:

$$\begin{aligned} a \ddot{H}(t) &+ b \dot{H}(t - \tau) + c \dot{H}(t) \\ &+ mH(t - \tau) + nH(t) \\ &= G(\dot{H}(t - \tau), \dot{H}(t), H(t - \tau), H(t)) \end{aligned} \quad (2)$$

where  $H(t) \triangleq y(t) - \bar{y}$ ,  $G(0, 0, 0, 0) = 0$  and  $H(t) = 0$  (equivalent to  $y(t) = \bar{y}$ ) is a trivial solution. That goes to say that if there is stability, then the queue length will tend to stabilize at size  $\bar{y}$ . ( $\bar{y}$  in general is the value of  $y(t)$  for which  $\dot{y}(t) = 0$ ). Two kind of analysis are then possible:

The first is to check whether we could have  $\bar{y} > B$ , where  $B$  is the maximum buffer size at the bottleneck node. If there are parameters for which this situation can happen, then it is obvious that losses will always occur because the buffer occupancy (queue length) will try to stabilize at a size exceeding the maximum buffer size. Therefore the protocol can't be stable for those parameters satisfying the inequality  $\bar{y} > B$ .

The second analysis is to use the theory of stability of solutions to delayed differential equations [1, 5, 12]. If the solution  $y(t) = \bar{y}$  is uniformly asymptotically stable, then whatever the initial parameters are, The queue length will stabilize at  $\bar{y}$  with minimal oscillations at the beginning. An extensive study using this method (the two analysis) for a number of protocols can be found in [1] where the stability conditions for these algorithms are established.

In the sequel, we shall consider only algorithms where the sending rate of each source becomes constant (roughly constant) for a relatively long period of transmission, if the total available bandwidth is constant (the rate of each source stabilizes, ultimately with few oscillations at the beginning).

## 2.2 Full Utilization

In order to achieve full utilization, the rate attributed to each source should be an increasing function if the whole bandwidth is not fully used. This requirement is needed to make sure that in the case of existing restricted sources (sources that are restricted by other nodes or cannot send at their Allowed Cell Rate (ACR)),

the unused bandwidth by the latter is shared among other unrestricted sources if such sources exist.

Assume that the sending rate of each source can be written in the form of equation (1). Then, the necessary (since the rate of each source converges to a constant value) and sufficient condition to ensure convergence is the following:

**Lemma 2.1** *Under the above hypothesis, a closed-loop congestion control protocol based on explicit rate is convergent (i.e. guarantees the full utilization of the bandwidth), if and only if*

$$\begin{aligned} & \forall \mu_{max} \leq \mu, \exists t_0 : \\ & \text{If } \sum_{k=1, k \neq i}^N x_k(t) = \mu_{min} = \mu - \mu_{max} \quad \forall t, \\ & \text{Then, } x_i(t) \geq \mu_{max}, \forall t > t_0 \end{aligned} \quad (3)$$

**Proof:** The proof is trivial: It is easy to verify that if the above condition is met, then the maximum utilization is guaranteed. Indeed, we'll have  $\forall t > t_0, \sum_{k=1}^N x_k(t - \gamma_k) \geq \mu$ . On the other hand (the other way of the equivalence), if  $\exists \mu_{max} \leq \mu : \forall t x_i(t) < \mu_{max}$ , then the full utilization is not guaranteed. Indeed, let's consider the model described previously, and assume that  $N - 1$  sources are restricted, thus  $\sum_{k=1, k \neq i}^N x_k(t) = \mu - \mu_{max}$ , and only source  $i$  is not a restricted source. Because  $x_i(t) < \mu_{max} \quad \forall t$ , it follows that  $\sum_{k=1}^N x_k(t) < \mu, \quad \forall t$ , therefore the protocol is not convergent. ■

An example of non convergent protocols is the the "pure-rate matching" [4, 3]. This is in fact a "pure rate matching" corrected by the queue length, which we will refer to simply as "pure rate matching". In this protocol, the rate is updated as follows: Assume the number of active connections at the bottleneck node is  $N$  and the available bandwidth is  $\mu$ , then the rate (fair share) attributed to each source is given by:

$$\begin{aligned} ER_i(t) &= (1 - \alpha)x_i(t - \gamma_i) + \alpha \frac{\mu}{N} \\ &\quad + \eta(Q_H - y(t)) \\ &= x_i(t + \beta_i) \end{aligned}$$

where  $y(t)$  is the queue length at time  $t$  and  $Q_H$  is given threshold smaller than  $B$ . It is easy to verify that the maximum rate ( $x_i(t)$ ) allocated to a given source  $i$ , is obtained for  $y(t - \beta_i) = 0$ , thus we have

$$\begin{aligned} x_i(t) &\leq (1 - \alpha)x_i(t - \tau_i) + \alpha \frac{\mu}{N} + \eta Q_H \\ \implies x_i(t) - x_i(t - \tau_i) &\leq \\ &\alpha \left( \frac{\mu}{N} - x_i(t - \tau_i) + \frac{\eta}{\alpha} Q_H \right) \implies \\ x_i(t) &< x_i(t - \tau_i) \text{ if } x_i(t - \tau_i) > \frac{\mu}{N} + \frac{\eta}{\alpha} Q_H \end{aligned}$$

Recursively, we can prove that

$\forall t x_i(t) \leq \frac{\mu}{N} + \frac{\eta}{\alpha} Q_H \triangleq \mu_{max}$ . Therefore,  $\exists \alpha, \beta, N : \mu_{max} \leq \mu$  and  $\forall t x_i(t) \leq \mu_{max}$ . In virtue of Lemma 2.1, it follows that the scheme is not convergent. In other words, if the available bandwidth that could be allocated to an unrestricted source  $i$ , because of existing restricted sources, is larger than  $\frac{\mu}{N} + \frac{\eta}{\alpha} Q_H$ , then the maximum utilization of the link will not be achieved.

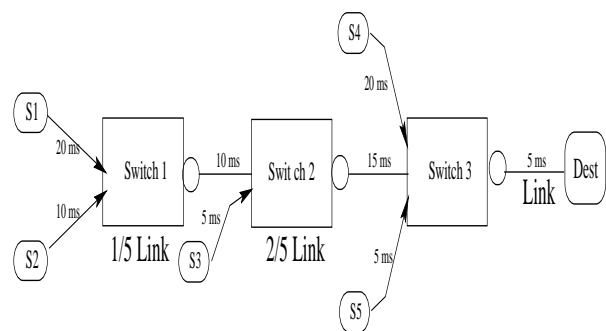


Figure 2: Simulation scenario for convergence and fairness .

Figure 2 illustrates the scenario and network setup we used to verify the convergence and the fairness of the different examples presented in this paper. The scenario is the following: All sources have the same destination. The capacity of the different bottleneck links are  $1/5 * Link = 4000 \text{ cells/sec}$ ,  $2/5 * Link = 8000 \text{ cells/sec}$ , and  $Link = 20000 \text{ cells/sec}$ , respectively at the first, second and third node (switches). Sources 2, 3, 5 start to transmit two

(2) seconds before sources 1 and 4. During the two first seconds of simulation, ideally (if the protocol is convergent) the three sources will be allocated the following rates: S1, S3: 4000 cells/sec and S5: 12000 cells/sec (because S1 and S2 are restricted sources). After two seconds of simulation, we should ideally converge to the following rate allocation: S1, S2: 2000 cells/sec, S3: 4000 cells/sec, and S4, S5: 6000 cells/sec.

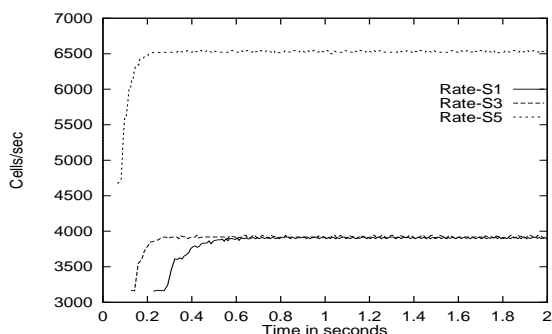


Figure 3: Rate allocation with "pure-rate matching" protocol

Figure 3 shows the rate of each source between time 0 and 2 seconds, when the "pure-rate matching" algorithm is considered. We can see that source 5 transmits at 6500 cells/sec which is far smaller than the bandwidth it could use (12000 cells/sec). The utilization for the 2 seconds of simulation is 72.5% and could be worse had the fair share for source 5 been much higher.

### 2.3 Fairness

A sufficient condition for fairness using the model described in figure 1 can be stated as follows:

A protocol is fair if

$$\begin{aligned} \sum_{k=1}^N x_k(t - \gamma_k) &= \mu \\ \implies x_i(t + \beta_i) - x_j(t + \beta_j) &= 0 \quad \forall t \\ \text{or } |x_i(t + \beta_i) - x_j(t + \beta_j)| & < |x_i(t - \gamma_i) - x_j(t - \gamma_j)| \quad \forall i \neq j \end{aligned} \quad (4)$$

This means that the rate allocated to two distinct connections ( $i, j$ ) is the same or tends to

be the same after some time, since the difference between the two rates allocated to these connections currently is smaller than the difference of their sending rates measured at the switch (incoming rate to the switch).

Moreover, a protocol is not fair (does not achieve fairness) if

$$\begin{aligned} \exists i, j : \sum_{k=1}^N x_k(t) &= \mu \\ \text{and } |x_i(t) - x_j(t)| &\geq C^{st} > 0 \quad \forall t \end{aligned}$$

The following is an example of unfair algorithm (first version of ERICA protocol [8]):

$$ER_i(t) = \max \left( \frac{1}{N}, \frac{x_i(t - \gamma_i)}{\sum_{k=1}^N x_k(t - \gamma_k)} \right) \mu_1$$

where  $\mu_1 \triangleq \mu * \text{Target\_Utilization}$

ERICA targets to use only a fraction of the available bandwidth (90-95%). Therefore, we assume that the available bandwidth seen by ERICA is  $\mu_1$  instead of  $\mu$ . When few connections are restricted and can't send at their fair share, the unrestricted connections should send at rates higher than  $\mu_1/N$ . Assume that there are at least two unrestricted connections  $i$  and  $j$ :

$$x_i(t) = \frac{x_i(t - \tau_i)}{\sum_{k=1}^N x_k(t - \beta_i - \gamma_k)} \mu_1$$

Therefore, we have

$$\begin{aligned} |x_i(t) - x_j(t)| &= \left| \frac{x_i(t - \tau_i)}{\sum_{k=1}^N x_k(t - \beta_i - \gamma_k)} \right. \\ &\quad \left. - \frac{x_j(t - \tau_j)}{\sum_{k=1}^N x_k(t - \beta_j - \gamma_k)} \right| \mu_1 \end{aligned} \quad (5)$$

From (5), we have:

$$\begin{aligned} \sum_{k=1}^N x_k(t) &= \mu_1 \quad \forall t \\ \implies |x_i(t) - x_j(t)| &= |x_i(t - \tau_i) - x_j(t - \tau_j)| \end{aligned}$$

It means that there exist allocations where the whole available bandwidth is allocated while the sources could have different rates, therefore fairness is not guaranteed not even for equivalent sources ( $\tau_i = \tau_j$ ).

Note that the fairness in this scheme is often achieved, when there are unrestricted connections and the sources start to transmit relatively at the same time or at rates close to the fair share. However, when a source starts to transmit while others are in their “steady state”, we can have situations where the full utilization is achieved ( $\sum_{k=1}^N x_k(t - \beta_j - \gamma_k) = \mu \forall t, j$ ) while the sources are allocated different rates.

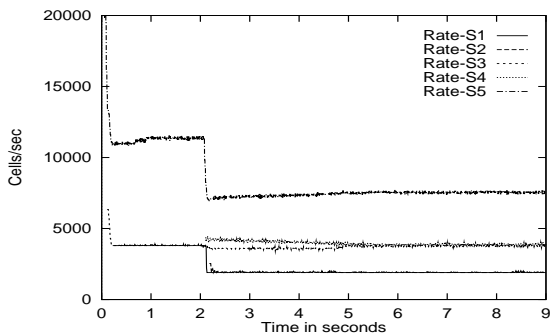


Figure 4: Rates in cells/sec when using ERICA.

To illustrate that, we run a simulation using the scenario depicted in Figure 2 where the ERICA algorithm is used targeting 95% utilization (19000 cells/sec). Figure 4 shows the rates achieved by the five sources. We can see that the available bandwidth is fully used, but the sources did not get their fair share. Indeed, source 5 was allocated a rate of 7500 cells/sec, and sources 3 and 4 have received less than 3800 cells/sec each. Source 4 which is not restricted in this example received less than half what source 6 has received while they should have received the same rate had the protocol been fair. As noted before, the fairness of this protocol depends heavily on the initial rates. The closer the initial rates to the optimal shares the more fair the algorithm is and inversely.

Following the steps described in this section about how to evaluate fairness, we can easily prove that the “pure rate matching” protocol described in the previous section is indeed fair.

**Remark 2.2** *If a congestion control protocol is fair (satisfies (4)) and convergent (satisfies (3)), then it is “MAX-MIN” fair [10]. Indeed, if it is fair then any two unrestricted connec-*

*tions will have the same share, therefore all unrestricted connections will have the same share. Moreover, if the algorithm is convergent, then the entire bandwidth will be used should there be few unrestricted connections. Because the algorithm is fair then the bandwidth not used by restricted sources is going to be fairly shared among unrestricted ones.*

### 3 Conclusion

In this paper we presented a simple yet efficient method for evaluating protocols based on explicit rate and/or buffer occupancy and can be applied to congestion protocols in any type of network be it ATM IP or anything else, as long as the information used to compute the sending rate is derived from the available bandwidth and/or buffer occupancy at intermediate nodes. We focused mainly on ABR service in ATM networks for it is the service for which thousands of protocols were designed, and where the fair share (Explicit Rate) rate is clearly computed using an estimate of the available bandwidth and buffer occupancy.

We focused on the three characteristics of stability, convergence and fairness because they are the most important measures of robustness that every protocol has to satisfy. We used few examples to show how to apply our conditions in practice, then by means of simulation we validated our hypothesis.

More and more congestion control experts argue that Max-Min fairness should not be considered the ideal of fairness. Indeed, one can argue that connections crossing more hops use more resources than connections traversing only few routers. Therefore, the bandwidth should be allocated in proportion to some criteria such as round trip time, number of crossed hops etc... [10]. In general we would like to see more bandwidth being allocated to connections using less network resources. TCP for instance achieves proportional fairness based on the round trip time, because it favors connections with small round trip time over connections with large round trip time. This is the topic we would like to explore in the near future.

## References

- [1] O. Ait-Hellal, E. Altman, "Stability of ABR congestion control using the theory of delayed differential equations", *International Journal of Systems Science*, Vol. 34, Numbers 10-11, pp. 575-584, 2003.
- [2] O. Ait-Hellal, G. Leduc, "TCP Vegas-Like algorithm for layered multicast transmission", in *Proc. International Conference On Communications in Computing (CIC'04)* pp. 58-64, Las Vegas, Nevada, June 21-24, 2004.
- [3] O. Ait-Hellal, E. Altman, T. Basar, "Rate based flow control with bandwidth information", *European Trans. on Telecom.*, pp. 55-66, 1997.
- [4] E. Altman, F. Baccelli, J. C. Bolot, "Discrete-time analysis of adaptive rate control mechanisms", *High Speed Networks and their performance*, H. G. Perros and Y. Viniotis Eds., North Holland, pp. 121-140, 1994.
- [5] R. Bellman, K. L. Cooke, "Differential-Difference Equations", *Academic press*, 1963.
- [6] C.V. Hollot, V. Misra, D. Towsley, and W.B. Gong, "A control theoretic analysis of RED", *Proceedings of IEEE Infocom*, Ankoroge, Alaska, 2001.
- [7] Van Jacobson. "Congestion avoidance and control". *ACM SIGCOMM 88*, pages 273-288, 1988.
- [8] R. Jain, S. Kalyanaraman, R. Goyal, S. Fahmy, R. Viswanathan, "ERICA switch algorithm: a complete description", *ATM Forum/96-1172*, 1996.
- [9] R. Johari, D. Tan, "End-to-End congestion control for the Internet: delays and stability", *IEEE/ACM Transactions on Networking*, Vol. 9, 818-832, 2001.
- [10] F. P. Kelly, A.K. Maulloo, D.K.H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability", *Journal of the Operational Research Society*, 49, pp. 237-252, 1998.
- [11] S. Keshav, "REAL: A network simulator", *Comp. Sci. Dept., UC Berkeley, Technical Report 88/472*, 1988.
- [12] Y. Kuang, "Delay differential equations, with applications in population dynamics", *Academic press*, 1993.
- [13] L. Massoulié, "Stability of distributed congestion control with heterogenous feedback delays", *IEEE Trans. on Automatic Control*, 47, 895-902, 2002.
- [14] F. Paganini, J. Doyle, S. Low, "Scalable laws for stable network congestion control", *Proceedings of IEEE conference on decision & control*, 2001.
- [15] G. Vinnicombe, "On the stability of networks operating TCP-like congestion control", in *IFAC'02*, 2002.