

On the Positive Recurrence of a Markov Chain Describing File Arrivals and Departures in a Congestion-Controlled Network

R. Srikant

Department of Electrical and Computer Engineering
and
Coordinated Science Laboratory
University of Illinois
rsrikant@uiuc.edu

1 Introduction

For networks using Kelly-type resource allocation policies, it has been shown in [2, 1] that a Markov chain describing the number of files in the system is positive recurrent. The model used in these papers assumes a time-scale separation between the congestion control algorithms that facilitate the resource allocation and the file arrival and departure processes. Specifically, it is assumed that the congestion controllers converge instantaneously when viewed at the time scale at which files arrive and depart. In this note, we relax this time-scale separation assumption. A similar result has also been obtained by Lin and Shroff independently in [5].

We assume that files (also known as connections) of type r arrive according to a Poisson process of rate λ_r and a file of type r has a mean file size of $1/\mu_r$. Each file type r is associated with a collection of links known as its route, and we use the notation $l \in r$ to indicate that a link belongs to the route of a file of type r . A file of type r generates data packets according to a (time-inhomogeneous) Poisson process of rate $x_r(t)$, where $x_r(t)$ depends on the congestion price feedback from the network. In this paper, we assume that link congestion prices and source rates are generated according to the dual algorithm described in [3, 6]. The details of this algorithm will be provided in subsequent sections. To facilitate a simple Markov chain description, we assume that packet sizes are exponentially distributed with mean 1. It is easy to relax this assumption and allow phase-type packet size distributions. The capacity of link l in the network is denoted by c_l and we assume the following stability condition:

$$\sum_{r:l \in r} \frac{\lambda_r}{\mu_r} < c_l, \quad \forall l, \quad (1)$$

i.e., we assume that the load on each link is less than the link capacity. As in [8, 2, 1], we further assume that the traffic generated by each user is instantaneously offered to all the links on its route. In other words, there are no delays in the network. In the rest of this note, we show that the network operating under the above model is stable with some additional assumptions.

2 Stability of a Heuristic Fluid Model

Consider the following fluid model for connection arrival and departure dynamics

$$\dot{n}_r = \lambda_r - \mu_r n_r x_r, \quad (2)$$

and link price dynamics

$$\dot{p}_l = y_l - c_l, \quad (3)$$

where the arrival rate at each link is given by

$$y_l = \sum_{s:l \in s} n_s x_s$$

and x_r , the transmission rate of a type r source, satisfies

$$\frac{w_r}{x_r^\alpha} = q_r,$$

where q_r is price of route r and is given by

$$q_r = \sum_{l:l \in r} p_l,$$

and $\alpha > 0$ is some constant that is dictated by the utility function of the users [7]. It is assumed that the positive part of the right-hand side of (3) is used whenever $p_l = 0$. Further, x_r is assumed to be upper-bounded by some constant m_r to ensure that it is well-defined when $q_r = 0$. Consider the Lyapunov function

$$V(\mathbf{n}, \mathbf{p}) = \frac{1}{1 + \alpha} \sum_r K_{1r} n_r^{1+\alpha} + \frac{1}{2} \sum_l K_{2l} p_l^2,$$

where the constants K_{1r} and K_{2l} will be chosen later. The rationale for choosing such a Lyapunov function is as follows: the first term is the Lyapunov function used by Bonald and Massoulié in [1] to prove connection-level stability when the congestion controllers are assumed to converge instantaneously. The second term is a natural Lyapunov function which can be used to prove the stability of (3) if the arrival rates at the links are fixed. Thus, we hope that some linear combination of these two terms would serve as the Lyapunov function for the network. As we will see, this is indeed the case.

Note that

$$\frac{dV}{dt} = \sum_r K_{1r} n_r^\alpha (\lambda_r - \mu_r n_r x_r) + \sum_l K_{2l} p_l (y_l - c_l).$$

From the stability condition, there exists an $\epsilon > 0$ such that

$$\rho_l (1 + \epsilon) \leq c_l,$$

where

$$\rho_l = \sum_{s:l \in s} \rho_s, \quad \rho_s = \frac{\lambda_s}{\mu_s}.$$

Choose such an ϵ , define $\delta = \epsilon/2$ and rewrite \dot{V} as follows:

$$\begin{aligned}
\dot{V} &= -\delta \sum_r K_{1r} n_r^\alpha \lambda_r + \sum_r K_{1r} n_r^\alpha \mu_r (\rho_r(1+\delta) - n_r x_r) + \sum_l K_{2l} p_l (y_l - c_l) \\
&= -\delta \sum_r K_{1r} n_r^\alpha \lambda_r \\
&\quad + \sum_r K_{1r} \mu_r n_r (\rho_r(1+\delta))^\alpha \left(\frac{1}{(\rho_r(1+\delta)/n_r)^\alpha} - \frac{1}{x_r^\alpha} \right) \left(\frac{\rho_r(1+\delta)}{n_r} - x_r \right) \\
&\quad + \sum_r K_{1r} \mu_r n_r \frac{(\rho_r(1+\delta))^\alpha}{x_r^\alpha} \left(\frac{\rho_r(1+\delta)}{n_r} - x_r \right) + \sum_l K_{2l} p_l (y_l - c_l) \\
&\leq -\delta \sum_r K_{1r} n_r^\alpha \lambda_r + \sum_r K_{1r} \mu_r \frac{(\rho_r(1+\delta))^\alpha q_r}{w_r} (\rho_r(1+\delta) - n_r x_r) + \sum_l K_{2l} p_l (y_l - c_l),
\end{aligned}$$

where the last line follows from the fact that $1/x_r^\alpha$ is a decreasing function and $q_r = w_r/x_r^\alpha$. Now, choose $K_{2l} = 1$ and $K_{1r} \mu_r (\rho_r(1+\delta))^\alpha / w_r = 1$. Then,

$$\dot{V} \leq -\delta \sum_r K_{1r} n_r^\alpha \lambda_r + \sum_r q_r (\rho_r(1+\delta) - n_r x_r) + \sum_l p_l (y_l - c_l).$$

Noting that

$$\sum_r q_r \rho_r = \sum_l p_l \rho_l \text{ and } \sum_r q_r n_r x_r = \sum_l p_l y_l,$$

we have

$$\dot{V} \leq -\delta \sum_r K_{1r} n_r^\alpha \lambda_r - \delta \sum_l p_l \rho_l.$$

In the above derivation, we have assumed that $n_r > 0$ and we have ignored the upper-bound m_r on source r 's rate. One way to complete the proof is to carefully justify these modeling assumptions and further show that (2)-(3) are the correct fluid limit under an appropriate scaling such as the one in [4]. Alternatively, we can uniformize the countable-state, continuous-time Markov chain to obtain a countable-state, discrete-time Markov chain and use the Foster-Lyapunov criterion as in [9, 10]. We will use the second approach in the next section.

3 Positive Recurrence of the Markov Chain

We only outline the key steps in establishing positive recurrence since the details are nearly identical to the analysis in [9, 10]. For the stochastic model, we assume that link price is simply the (virtual) queue length at the link queue, where the arrival process into the queue is Poisson with a time-varying mean $\sum_{r:l \in r} n_r(t) x_r(t)$. With this set-up, the process $(\mathbf{n}(t), \mathbf{p}(t))$ can be described by a continuous-time Markov chain. We first perform uniformization to convert this to a discrete-time chain. However, we have to make one more assumption to do this. We assume that $n_r(t) x_r(t)$ is upper-bounded by some (arbitrarily large, but finite) constant M . In other words, when $n_r > 0$,

$$x_r = \min \left(\left(\frac{w_r}{q_r} \right)^{1/\alpha}, \frac{M}{n_r} \right). \quad (4)$$

This ensures that the transition rates in the Markov chain are bounded, thus allowing us to uniformize to obtain a discrete-time chain. Now, following [10], it is easy to show that the expected drift of the Lyapunov function used in the previous section is negative when the system state (\mathbf{n}, \mathbf{p}) is sufficiently far away from the origin.

The assumption (4) may seem artificial, but in most realistic networks, the access rates into the network are bounded and hence such rate limitations at the edges are perhaps not unreasonable to assume. As has been pointed out in [5], without such a restriction, stability can only be established for $\alpha \geq 1$. With an upper-bound on the total rate on route r , positive recurrence can be established for all $\alpha > 0$.

References

- [1] T. Bonald and L. Massoulié. Impact of fairness on Internet performance. In *Proceedings of ACM Sigmetrics*, 2001.
- [2] G. de Veciana, T.-J. Lee, and T. Konstantopoulos. Stability and performance analysis of networks supporting elastic services. *IEEE/ACM Transactions on Networking*, 9(1):2–14, February 2001.
- [3] F. P. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.
- [4] F.P. Kelly and R. J. Williams. Fluid model for a network operating under a fair bandwidth-sharing policy. *Annals of Applied Probability*, 14:1055–1083, 2004.
- [5] X. Lin and N. Shroff. On the stability region of congestion control. In *Proceedings of the Allerton Conference on Communications, Control and Computing*, 2004. To appear.
- [6] S. H. Low and D. E. Lapsley. Optimization flow control, I: Basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, pages 861–875, December 1999.
- [7] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, October 2000.
- [8] J. Roberts and L. Massoulié. Bandwidth sharing and admission control for elastic traffic. In *Proc. ITC specialists seminar*, Yokohama, Japan, 1998.
- [9] R. Srikant. *The Mathematics of Internet Congestion Control*. Birkhauser, 2004.
- [10] R. Srikant. Models and methods for analyzing Internet congestion control algorithms. In *Advances in Communication Control Networks in the series Lecture Notes in Control and Information Sciences (LCNCIS)*, C.T. Abdallah, J. Chiasson and S. Tarbouriech (eds.). Springer, 2004. To appear.