

Dynamic Analysis of Congested TCP Networks¹

Yossi Chait^a, C.V. Hollot^b, Vishal Misra^c, Huaizhong Han^b and Yoram Halevi^d

^aMIE Department, University of Massachusetts Amherst, MA 01003

^bECE Department, University of Massachusetts Amherst, MA 01003

^cDepartment of Computer Science, Columbia University, New York, NY 10027-7003

^dFaculty of Mechanical Engineering, Technion - I.I.T., Haifa 32000, Israel

Abstract

The objective of this paper is to study the small-signal stability of congested TCP networks. The starting point for this analysis is a fluid model describing the interaction of TCP-controlled traffic with congested routers. Assuming integral action in the active queue management scheme, we compute the network's equilibrium point about which we determine the network's small-signal dynamics. Specializing to networks comprised of m links and $m + 1$ sources, we study closed-loop stability, study network scaling, and evaluate robustness to variations in network parameters.

1 Introduction

Recently, there has been a focus on analyzing the dynamic behavior of congestion control in the Internet; e.g., see the tutorial article [1] and the articles cited therein. For the most part, this work has been limited to transmission control protocol (TCP) networks having a single congested link. A departure was taken in [1], where a network of multiple sources and links were considered; however, in the treatment, the authors sidestepped the existing TCP and proposed an alternative protocol, having possibly advantageous scalability properties. In this paper, we content ourselves with the status quo¹, thereby acknowledging the entrenchment of TCP, and analyze dynamic behavior beyond the case of a single congested link. By way of introduction to more complex networks, we first recall from [3] the fluid-flow model for N homogeneous TCP-controlled sources interacting with a single congested link. These TCP-controlled sources adjust their aggregate packet flow r as a function of the probability of packet mark

¹This work is supported in part by the National Science Foundation under Grant CMS-9800612 and by DARPA under Contract DOD F30602-00-0554.

²We restrict our attention to the additive-increase, multiplicative decrease (AIMD) behavior of TCP, ignoring timeouts and slow start. Also, we assume that routers communicate congestion using the explicit congestion notification (ECN) mechanism; e.g., see [2].

p as described by:

$$\begin{aligned}\dot{W} &= \frac{1}{R} - \frac{W^2}{2R} p_R \\ r &= \frac{N}{R} W\end{aligned}$$

where W is the TCP window size, R is the round-trip time and where $p_R(t) \triangleq p(t-R)$ is the delayed probability. The round-trip time is composed of both queuing and transport delays $R = \frac{q}{C} + T_p$ where C is the link capacity, T_p the propagation delay and where the queue length q of the congested router satisfies

$$\dot{q} = \begin{cases} -C + r, & q > 0 \\ [-C + r]^+, & q = 0. \end{cases}$$

The router description is completed by introducing an AQM rule that relates a link or router quantity, such as the queue length q , to the marking probability p . Some examples include drop tail, random early discard (RED), random exponential marking (REM) and PI control, which all react to queue length. The block diagram in Figure 1 puts these dynamics together to emphasize the feedback nature of TCP/AQM.

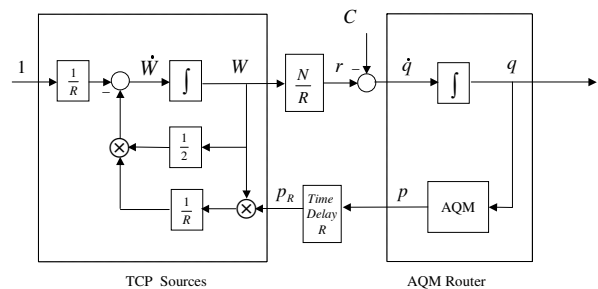


Figure 1: AQM signals TCP sources to change sending rate in response to router queue length.

For more general networks, we again appeal to the model in [3] and consider a network comprised of n aggregate TCP flows and m congested routers. To describe the interconnection between sources and routers

we let vector V_j denote the ordered set of links traversed by the j th flow r_j . These paths can be represented by a binary *routing matrix* $A = \{a_{ij}\}$ where $a_{ij} = 1$ if the i th link handles the r_j flow; i.e., $i \in V_j$. We assume that A is full rank and has no zero column, and write

$$\hat{r} \triangleq Ar \quad (1)$$

where \hat{r} denotes the link flows. Conversely, with p_i denoting the marking probability at the i -th congested router, the marking probability at the j th TCP source is

$$\hat{p}_j = 1 - \prod_{i \in V_j} (1 - a_{ji} p_i).$$

If $p_i \ll 1$, this simplifies to

$$\hat{p} = A^T p. \quad (2)$$

The queue length q_i of each of the $i = 1, 2, \dots, m$ routers satisfies

$$\dot{q}_i = \begin{cases} \hat{r}_i - C_i, & q_i > 0 \\ [\hat{r}_i - C_i]^+, & q_i = 0 \end{cases} \quad (3)$$

where $C = \{C_i\}$ is the vector of link capacities. Conversely, each of the $j = 1, 2, \dots, n$ sources controls N_j packet-flows, whose aggregate rate r_j is described by the window-dynamic:

$$\begin{aligned} \dot{W}_j &= \frac{1}{R_j} - \frac{W_j^2}{2R_j} \hat{p}_{jR} \\ r_j &= \frac{N_j}{R_j} W_j \end{aligned} \quad (4)$$

where W_j is the window length and where the round-trip time R_j satisfies

$$R_j = T_{p_j} + \sum_{i \in V_j} \frac{q_i}{C_i} \quad (5)$$

with T_{p_j} the propagation delay. The delayed marking \hat{p}_{jR} is defined by $\hat{p}_{jR}(t) = \hat{p}_j(t - R_j)$ where we account for the aggregated round-trip time delay at the sources. Finally, an AQM marking rule, possibly dynamic in nature, relates the queue lengths q to the marking probabilities p , thus tying the source and router dynamics together as illustrated in the feedback interconnection of Figure 2.

The objective of this paper is to study the small-signal stability of the TCP/AQM network dynamic described by (1) – (5). This first requires computation of the network's operating point which is described in the next section, and, leads to a general description of the network's small-signal dynamics. Subsequently, we will specialize the analysis to a network comprised of m links, $m + 1$ sources and common parameters N_j , C_i and R_j . This allows us to study closed-loop stability, study the impact of network scale, and to evaluate robustness to variations in these network parameters.

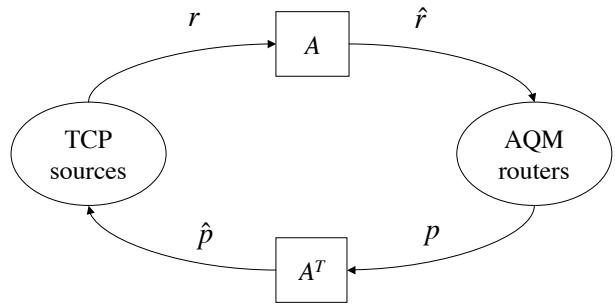


Figure 2: A network of n TCP sources interacting with m links.

2 Linearization

In this section we linearize the open-loop dynamics of the congested network (1) – (5) about its operating point.

2.1 Operating point

The operating point (r_0, p_0, q_0) of the network is obtained by setting $\dot{q}_i = 0$ and $\dot{W}_j = 0$ in (3) and (4) to yield the constraints

$$C \geq Ar \quad (6)$$

and

$$\Lambda \begin{bmatrix} \frac{1}{r_1^2} \\ \vdots \\ \frac{1}{r_n^2} \end{bmatrix} = A^T p \quad (7)$$

where

$$\Lambda \triangleq \text{diag} \left\{ 2 \frac{N_j^2}{R_j^2} \right\}.$$

Solutions $(r, p, q) = (r_0, p_0, q_0)$ to the above define the network's operating points, and, in this paper, we fix the operational queue level q_0 , anticipating the use of integral action in the AQM control law. Thus, for network parameters (N, C, R) , we seek feasible solutions (r_0, p_0) to (6) and (7) which are characterized by the following result:

Proposition (See Appendix A for proof): *Given network data (N, C, R) , the operating point (r_0, p_0) of the congested network dynamics (1) – (5) is unique, where $p = p_0$ maximizes the concave function*

$$z(p) \triangleq 2 \sum_{j=1}^n \sqrt{(\Lambda A^T p)_j} - p^T C \quad (8)$$

and $r = r_0$ verifies (7). The operating point is feasible if $r_{0_j} > 0$ and $p_{0_i} \in [0, 1)$.

Now, we consider linearization about this operating point.

2.2 Small-signal dynamics

Linearizing (1) – (5) about the operating point satisfying (6) and (7) gives

$$\begin{bmatrix} \delta \dot{W} \\ \delta \dot{q} \end{bmatrix} = \begin{bmatrix} -D_{11} & 0 \\ AD_{21} & -AD_{22L}A^TD_{22R} \end{bmatrix} \begin{bmatrix} \delta W \\ \delta q \end{bmatrix} + \begin{bmatrix} -D_B A^T \\ 0 \end{bmatrix} \delta p \quad (9)$$

where

$$\begin{aligned} D_{11} &= \text{diag} \left\{ \frac{2N_i}{r_{0i}R_i^2} \right\}; & D_{21} &= \text{diag} \left\{ \frac{N_i}{R_i} \right\}; \\ D_{22L} &= \text{diag} \left\{ \frac{r_{0i}}{R_i} \right\}; & D_{22R} &= \text{diag} \left\{ \frac{1}{C_j} \right\}; \\ D_B &= \text{diag} \left\{ \frac{R_i r_{0i}^2}{2N_i^2} \right\}. \end{aligned}$$

In the next section we analyze these linear dynamics for the special case of m links and $m + 1$ flows.

3 Analysis of a Network with m Links and $m + 1$ Sources

The previous proposition and linearization (9) allows one to *practically* analyze the small-signal behavior of arbitrarily-sized networks. In an effort to explicitly relate such behavior to network parameters, we presently consider the case $n = m + 1$; i.e., when there is one more source than link, and assume a routing matrix with structure:

$$A = \begin{bmatrix} & 1 & & \\ & 1 & & \\ I & & \ddots & \\ & & & 1 \end{bmatrix}.$$

In this case, each of the first m sources passes through a distinct router, with the $m + 1$ th source visiting each of the m routers. Furthermore, we consider $C_j = c$, $R_i = \rho$ and $N_i = \eta$ for all j and i .

3.1 Aggregate loss and throughput

This network's operating point is found from the proposition where the maximization of (8) gives

$$p_{0i} = \frac{2\eta^2 (1 + \sqrt{m})^2}{\rho^2 c^2 m}$$

for all $i = 1, 2, \dots, m$, and, from (7):

$$r_{0j} = \begin{cases} \frac{c\sqrt{m}}{1+\sqrt{m}} & j = 1, 2, \dots, m \\ \frac{c}{1+\sqrt{m}} & j = m + 1. \end{cases}$$

The aggregate network loss and throughput is then

$$\sum_{j=1}^{m+1} \hat{p}_{0j} = \frac{4\eta^2}{\rho^2 c^2} (1 + \sqrt{m})^2;$$

$$\sum_{j=1}^{m+1} r_{0j} = c \frac{1 + m\sqrt{m}}{1 + \sqrt{m}}.$$

In comparison, recall the loss and throughput for a single-source, single-link case (see [4]):

$$p_0 = \frac{4\eta^2}{\rho^2 c^2}; \quad r_0 = c.$$

3.2 Dynamic analysis

We now analyse the feedback nature of TCP/AQM depicted in Figure 2 using the linearized dynamics (9). Let $\delta q(s) = P(s)\delta p(s)$ be a transfer function representation for this dynamic and consider the block-diagram representation of TCP/AQM in Figure 3 where the AQM rule is $k(s)I$ which, in addition to being decentralized, implements the same dynamic $\delta p_i(s) = k(s)q_i(s)$ at each congested router. A computation

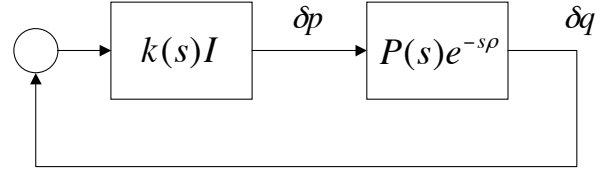


Figure 3: Block diagram representation of a TCP/AQM network where $k(s)I$ denotes the AQM rule.

gives the $2m + 1$ poles of $P(s)$ to be:

$$\lambda_k = \begin{cases} -\frac{2\eta}{\rho^2 c} \left(\frac{\sqrt{m}}{1+\sqrt{m}} \right) & k = 1, 2, \dots, m \\ -\frac{2\eta}{\rho^2 c} \left(\frac{1}{1+\sqrt{m}} \right) & k = m + 1 \\ -\frac{1}{\rho} \left(\frac{\sqrt{m}}{1+\sqrt{m}} \right) & k = m + 2, m + 3, \dots, 2m \\ -\frac{1}{\rho} \left(\frac{m+\sqrt{m}}{1+\sqrt{m}} \right) & k = 2m + 1. \end{cases}$$

which compares to the pole-pair for the single-link, single-source case:

$$-\frac{2\eta}{\rho^2 c}; \quad -\frac{1}{\rho}.$$

An explicit representation of $P(s)$ in terms of network parameters is

$$P = -\ell(w + 1) \left[I + \frac{w}{w + 1} E \right]$$

where

$$\ell(s) = \frac{c^2 m}{2\eta(1+\sqrt{m})^2} \frac{1}{\left(s + \frac{2\eta(1+\sqrt{m})}{\rho^2 c \sqrt{m}} \right) \left(s + \frac{\sqrt{m}}{\rho(1+\sqrt{m})} \right)},$$

$$w(s) = -\frac{1}{\rho(1+\sqrt{m})} \frac{1}{s + \frac{\sqrt{m}}{\rho}} + \frac{1}{m} \frac{\left(s + \frac{\sqrt{m}}{\rho(1+\sqrt{m})} \right) \left(s + \frac{2\eta(1+\sqrt{m})}{\rho^2 c \sqrt{m}} \right)}{\left(s + \frac{\sqrt{m}}{\rho} \right) \left(s + \frac{2\eta(1+\sqrt{m})}{\rho^2 c} \right)}$$

and

$$E = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & \cdots & 0 \end{bmatrix}.$$

The closed-loop characteristic function is

$$\begin{aligned} \phi &= (1 + \tilde{\phi}) \det(I + \tilde{T}E) \\ &= (1 + \tilde{\phi}) \prod_{i=1}^m (1 + \tilde{T}\lambda_i[E]) \\ &= (1 + \tilde{\phi}) (1 + (m-1)\tilde{T})(1 - \tilde{T})^{m-1} \end{aligned} \quad (10)$$

where

$$\tilde{\phi}(s) = \ell(s)(w(s) + 1)k(s)e^{-s\rho}$$

and

$$\tilde{T}(s) \triangleq \frac{\ell(s)w(s)k(s)e^{-s\rho}}{1 + \ell(s)(w(s) + 1)k(s)e^{-s\rho}}.$$

Now we consider some examples using specific network parameters.

3.3 Numerical example

Consider nominal values of $\eta = 60$ flows, $\rho = 250$ ms and $c = 3750$ packets/sec. In the case of a single congested link, the PI AQM controller

$$k(s) = 3.71 \times 10^{-5} \frac{s}{0.499s + 1}$$

stabilizes the network and provides 30 degrees of phase margin; see [4] for details. With this PI controller fixed, we plot stability boundaries for variations in network parameters (η, ρ, c) in Figure 4. Parameters be-

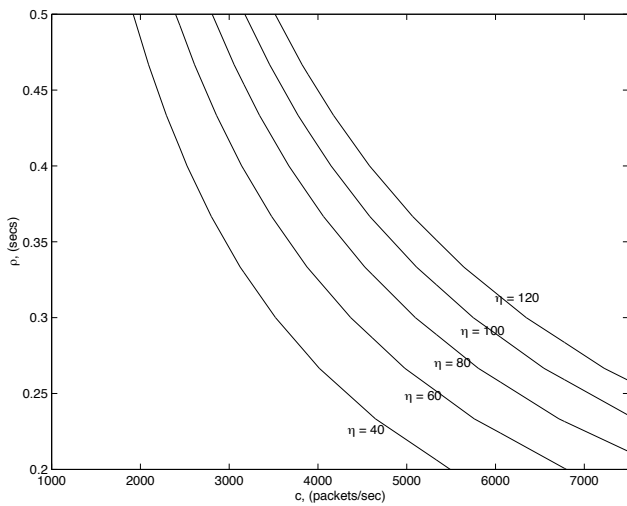


Figure 4: Stability boundaries for a single-link network under PI AQM. Parameters below curves correspond to stable networks.

low the curves correspond to closed-loop stability while

parameters above correspond unstable networks. Stability margins decrease with increased link capacity, increased round-trip time, or decreased number of TCP flows. It is interesting to compare these stability margins to the case when this same PI controller is applied, in a decentralized way, to the m routers in our more complex network. To do this, we first consider a congested 2-link, 3-source network; i.e., $m = 2$. The Nyquist plots of $\tilde{\phi}$, $(m-1)\tilde{T}$ and $-\tilde{T}$ in Figure 5 show that $\phi(s)$ in (10) has only stable zeros, and, that the network is stable. With this decentralized PI controller fixed, and designed on the basis of nominal parameters $(\eta, \rho, c) = (60, 0.25, 3750)$, we now plot stability boundaries for variations in (η, ρ, c) as shown in in Figure 6. This 2-link network appears to be more robust

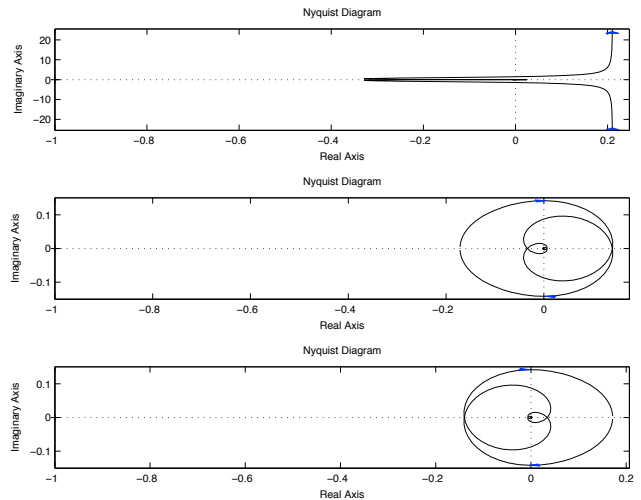


Figure 5: The Nyquist plots of $\tilde{\phi}$, $(m-1)\tilde{T}$ and $-\tilde{T}$ show that $\phi(s)$ in (10) is stable when $m = 2$ links.

than the single-link case. For example, from Figure 4, the single link network is unstable for the parameters $(\eta, \rho, c) = (60, 0.3, 4000)$, while, in Figure 6, the 2-link, 3-source network is stable. Let's explore the consequences for a larger network by repeating the analysis for a 20-link, 21-source network as shown in Figures 7 and 8. Figure 7 shows that this larger network is stable. Figure 8 indicates that this 20-link network is more robust than the single-link network, but has compressed stability regions when compared to the 2-link, 3-source network.

4 Conclusions

In this paper, we have analyzed the dynamics of a multiple link, multiple flow TCP network. Our work departs from the existing body of work in analyzing the *multiple* congested link scenario. Using a simplified fluid flow model to describe the network, our analysis reveals several interesting observations. Under the assumption of integral action by controllers at the con-

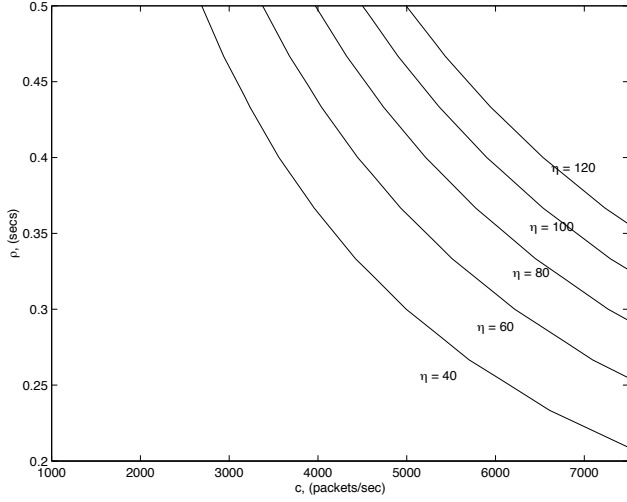


Figure 6: Stability boundaries for a 2-link, 3-source network under PI AQM. Parameters below curves correspond to stable networks.

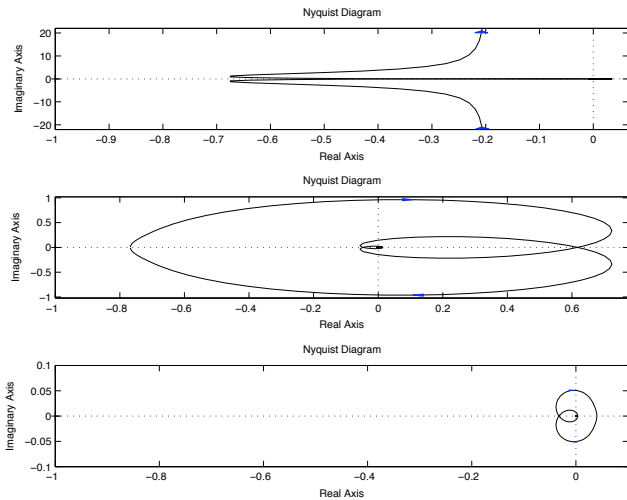


Figure 7: The Nyquist plots of $\tilde{\phi}$, $(m-1)\tilde{T}$ and $-\tilde{T}$ show that $\phi(s)$ in (10) is stable for a 20-link, 21-source network.

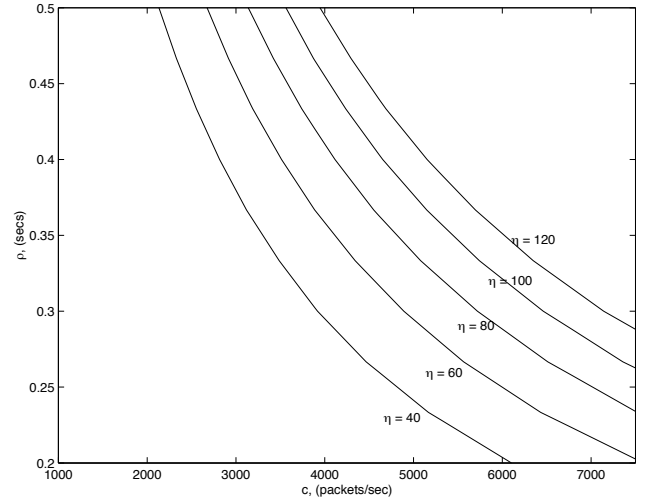


Figure 8: Stability boundaries for a 20-link, 21-source network under PI AQM. Parameters below curves correspond to stable networks.

gested link, we cast the rate assignment problem in this network as an optimization problem and demonstrate that there exists an equilibria which is unique. Further, we give feasibility conditions on achieving the equilibria. Next, focusing our attention on a specialized network with m congested links and $m+1$ flows, we have studied the closed loop properties of the system. One surprising observation is that the controller appears to be more robust under the multiple congested link setting as compared to the single congested link case.

While our work yields several insights, there are a number of avenues for future work. There are a number of TCP details that we do not model, e.g. slow start and time out. Also, our model considers the case where marks are the feedback mechanism, as opposed to packet drops, which is the prevailing scenario on the Internet. Finally, extending our study to the more general case of m links and n flows is of deep interest.

Acknowledgement The authors would like to thank Professor Arkadi Nemirovski for pointing to us the simplification in the dual optimization problem.

Appendix A

Consider the following optimization problem²

$$\min_{r \geq 0} \sum_{j=1}^n \frac{2N_j^2}{R_j^2 r_j}$$

$$\text{subject to } C - Ar \geq 0$$

²See also [1] where the equilibrium structure of networks is related to utility optimization.

where the cost is strictly convex over a closed and convex constraint set. Let r^* be the global minimizer. Then, there exists Lagrange multipliers $\lambda = \lambda^*$ minimizing the adjoined function $L(r^*, \lambda)$ where

$$L(r, \lambda) \triangleq \sum_{j=1}^n \frac{2N_j^2}{R_j^2 r_j} + \lambda^T (Ar - C)$$

and

$$\lambda_i^* \begin{cases} \geq 0, & i = 1, 2, \dots, m \\ = 0, & i \notin V_c \end{cases}$$

where

$$V_c \triangleq \{i : (Ar^*)_i - C_i = 0, i = 1, \dots, m\}.$$

Conditions for (r, λ) to yield a stationary value of L are

$$\nabla_r L = - \begin{bmatrix} \frac{2N_1^2}{R_1^2 r_1^2} \\ \vdots \\ \frac{2N_n^2}{R_n^2 r_n^2} \end{bmatrix} + A^T \lambda = 0;$$

$$\nabla_\lambda L_i = (Ar - C)_i = 0; \quad j \in V_c$$

which are precisely (6) and (7). As noted in [1], the role of the Lagrange multipliers is played by the network's loss probabilities p .

Computational complexity can be reduced by considering the *dual optimization* problem

$$\begin{aligned} \max \quad & z(p) \\ \text{subject to} \quad & p \geq 0 \end{aligned} \quad (11)$$

where

$$z(p) \triangleq \min_{r > 0} L(r, p).$$

The stationary condition $\nabla_r L = 0$ gives (7) from which

$$r_j = \frac{\sqrt{2}N_j}{R_j \sqrt{(A^T p)_j}}, \quad j \in V_c. \quad (12)$$

The full-rank assumption on A assures these r_j to be well-defined. Using (7), we substitute for $A^T p$ in $L(r, p)$ to obtain

$$\begin{aligned} z(p) &= \sum_{j=1}^n \frac{2N_j^2}{R_j^2 r_j} + \left[\frac{2N_1^2}{R_1^2 r_1^2} \cdots \frac{2N_n^2}{R_n^2 r_n^2} \right] r - p^T C \\ &= \sum_{j=1}^n \frac{4N_j}{R_j^2 r_j} - p^T C. \end{aligned}$$

Substituting (12) into this gives

$$z(p) = \sum_{j=1}^n \frac{2\sqrt{2}N_j}{R_j} \sqrt{(A^T p)_j} - p^T C$$

which is (8). This proves the proposition.

The innovation here is that the transmission rates r do not appear in $z(p)$, in contrast to the standard dual formulation followed in [1]. The Lagrangian for the dual problem is

$$\tilde{L}(p, \mu) = \sum_{j=1}^n \frac{2\sqrt{2}N_j}{R_j} \sqrt{(A^T p)_j} - p^T C + \mu^T p$$

with gradients

$$\frac{\partial \tilde{L}}{\partial p_j} = \sum_{j=1}^n a_{ij} r_j - C_i + \mu_i.$$

For any congested router, $p_i > 0$; hence, $\mu_i = 0$ and we obtain

$$\nabla_\lambda L_i = (Ar - C)_i = 0; \quad j \in V_c$$

which brings us back to our original set of network equations (6) and (7). The advantage in the dual problem is that we have at most $2m$ equations which is computationally attractive when $m \ll n$.

References

- [1] S. Low, F. Paganini, and J.C. Doyle, "Internet Congestion Control," *IEEE Control Systems Magazine*, Vol. 22, no. 1, pp. 28-43, 2002.
- [2] K.K. Ramakrishnan, S. Floyd. "A Proposal to add Explicit Congestion Notification (ECN) to IP," RFC 2481, Jan. 1999.
- [3] V. Misra, W. B. Gong, and D. Towsley, "Fluid-based Analysis of a Network of AQM Routers Supporting TCP Flows with an Application to RED," in *Proceedings of ACM/SIGCOMM*, 2000.
- [4] C. V. Hollot, V. Misra, D. Towsley, and W. B. Gong, "On Designing Improved Controllers for AQM Routers Supporting TCP Flows." in *Proceedings of IEEE INFOCOM 2001* a longer version of this paper is to appear in a special issue of *IEEE Transactions on Automatic Control* on "Systems and Control Methods for Communication Networks."