Congestion pricing for flow control and multipath routing in TCP/IP networks

Fernando Paganini Enrique Mallada
Universidad ORT Uruguay*

Abstract

We consider a TCP/IP style network, in which end-systems control their traffic rates based on congestion feedback, and routing is performed at intermediate nodes on a per-destination basis; extending standard IP, routers are allowed to use multiple outgoing links per destination.

We pose two optimization problems, that generalize and combine those in congestion control and traffic engineering, with variables which are local to either sources or routers. We obtain decentralized algorithms by propagating price variables and using them to control source rates and router traffic splits; we give conditions for convergence of these algorithms to optimal points.

Keywords: congestion control, routing, TCP/IP, optimization, pricing.

1 Introduction

The issue of how a flow network distributes and regulates its traffic is a classical problem, studied initially in transportation networks [13]. In TCP/IP networks, flows are under the control of automatic mechanisms of two types: input packet rates are controlled by the TCP protocol running in end-systems, and packet routing is controlled by the IP protocol running at intermediate routers.

Optimization is relevant to the analysis and design of both types of mechanisms. On the routing side, the *traffic engineering* problem [2, 10] is to select network paths to serve a given "traffic matrix" of demand between nodes, minimizing some measure of congestion: this leads naturally to multicommodity optimization, which could be solved offline and then manually imposed on a network (e.g., via MPLS technology). A more interesting proposition is to seek mechanisms that *adapt* routing to variable traffic conditions; a classic reference from which we will draw here is Gallager [3].

When traffic demands are not fixed but *elastic*, the input rates can themselves be optimized; following Kelly [4], a natural formulation is to maximize the overall *utility* of all traffic sources, subject to fixed routing and network capacity constraints. This economic formulation leads naturally to mechanisms based on *prices* to find decentralized algorithms. There has been substantial success in obtaining provably convergent algorithms, and relating them to those used in current TCP congestion control (see [4, 11, 7]).

Compared to this extensive research on either the supply or the demand sides of the problem, their combination (adapting both routing and source traffic) has been less studied. Some

*Cuareim 1451, Montevideo, Uruguay. Email: {paganini,mallada}@ort.edu.uy.

recent formulations of this problem [4, 12] employ as optimization variables the components of rate for each *end-to-end path* from source to destination. This gives convergent decentralized algorithms, but is not scalable since the number of such paths is exponential.

In this paper we describe a scalable approach to combined congestion control and multipath routing, where only variables with local physical meaning are controlled: source rates, and the traffic split performed at each router among its outgoing links. We will present natural optimization problems, and show convergent decentralized algorithms that result from a combination of congestion pricing and a suitable adaptation of traffic splits. This paper gives an overview of the theory and some initial work on packet implementation of the algorithms. More details on the theory are covered in [9].

2 Problem formulation

We consider a network of nodes, indexed by i, j, connected by a set of directed links $\mathcal{L}$, each denoted by l or by (i, j). The network supports various flows ("commodities") indexed by $k \in \mathcal{K}$, between a source $s(k)$ and a destination $d(k)$, following possibly multiple paths.

We introduce the following variables for each k: x^k, external flow in packets per second entering the network at the source; y_l^k, flow through link l; x_i^k, total flow coming into node i. The total flow on link l is denoted by y_l. We have the following basic relationships:

$$x^k_{s(k)} = x^k; \qquad x^k_j = \sum_{(i,j)\in\mathcal{L}} y^k_{i,j},\ j \neq s(k); \qquad x^k_i = \sum_{(i,j)\in\mathcal{L}} y^k_{i,j},\ \ i \neq d(k); \qquad y_l = \sum_k y^k_l. \quad (1)$$

Following [4], we will associate with each commodity k an increasing, concave utility function $U_k(x^k)$ that specifies the traffic's demand for rate. We formulate the following multipath counterpart of the "system problem" in [4].

Problem 1 (SYSTEM) *Maximize* $\sum_k U_k(x^k)$*, subject to link capacity constraints* $y_l \leq c_l$*, and flow balance constraints (1).*

A second problem can be formulated by replacing capacity constraints with (increasing, convex) barrier functions $\phi_l(y_l)$ that specify the congestion cost at the link, and optimize the *aggregate surplus*:

Problem 2 (BARRIER) *Maximize* $S := \sum_k U_k(x^k) - \sum_l \phi_l(y_l)$ *subject to flow balance constraints (1).*

While these are both convex programs, the challenge is to find *decentralized* solutions that can be embedded in network sources and routers, respecting the information constraints of the Internet. We work with TCP-like sources that control transmitted rate based on a simple congestion feedback signal, and IP-like routers that make routing decisions based on packet destination only. The main generalization we allow here with respect to standard IP is the use of multiple outgoing links with control over the traffic split. Other recent work which applies similar optimization problems to wireless networks can be found in [1, 14].

3 Decentralized control via pricing

Our solutions to the above optimization problems will be based on congestion pricing. The basic price p_l is a scalar variable that measures the congestion state of each link $l \in \mathcal{L}$;

we assume p_l depends only the total traffic y_l; there is no "service differentiation" between commodities. Price information must be propagated through the network and used to control source input rates, and traffic splits at routers.

The latter are specified as in [3] by variables $\alpha_{i,j}^d \geq 0$, $\sum_{(i,j)\in\mathcal{L}} \alpha_{i,j}^d = 1$ that control the fraction of incoming traffic of commodity k that node i sends through link $l = (i,j)$. We thus constrain our system dynamics by

$$y_{i,j}^k = \alpha_{i,j}^{d(k)} x_i^k, \qquad (i,j) \in \mathcal{L}. \tag{2}$$

The key to a scalable solution is the ability to summarize in a simple variable the congestion state of a portion of the network, using current routing patterns. We define, for this purpose, the node prices q_i^d, $i \in \mathcal{N}$, each representing the average price of sending packets from node i to destination d. Node prices must thus satisfy the recursion

$$q_d^d = 0, \qquad q_i^d = \sum_{(i,j)\in\mathcal{L}} \alpha_{i,j}^d [p_{i,j} + q_j^d], \quad i \neq d. \tag{3}$$

Given link prices $p_{i,j}$, it follows through similar arguments as those in [3] that the above equations have unique solutions for q_i^d, provided that the split ratios α^d have a path from every node to the destination. At the source node of commodity k, the node price $q^k := q_{s(k)}^{d(k)}$ summarizes the congestion cost of the network.

Based on these price signals, we must design dynamic methods to control source rates and router splits, in such a way that the resulting dynamics converges to the desired optimal points. We present these dynamic laws next. Before proceeding, we state a relationship (see [9]) between the flow and price quantities defined so far:

$$x^k q^k = \sum_{l\in\mathcal{L}} y_l^k p_l \quad \text{for each } k, \text{ and} \quad \sum_k x^k q^k = \sum_{l\in\mathcal{L}} y_l p_l. \tag{4}$$

Control of source rates and link prices

In the recent literature on congestion control for fixed, single-path routing, a number of algorithms have been proposed to control the generation of link prices and the response of sources to their aggregate price. We focus on two basic choices, and describe them in continuous time; we will use the notation

$$[w]_z^+ := \begin{cases} w, & \text{if } w > 0 \text{ or } z > 0; \\ 0 & \text{otherwise.} \end{cases}$$

In *primal* algorithms, targeted to the BARRIER problem, link prices are chosen statically to reflect the marginal congestion cost, and sources follow dynamically a subgradient in the direction of maximal utility.

Algorithm 1 (Primal, [4])

$$\dot{x}^k = \kappa(x^k)[U_k'(x^k) - q^k]_{x^k}^+, \qquad \kappa(x^k) > 0, \tag{5}$$

$$p_l = \phi_l'(y_l). \tag{6}$$

Dual algorithms are targeted to the SYSTEM problem, and approach it by taking the Lagrangian with respect to the capacity constraints,

$$L(\alpha, p, x) = \sum_k U_k(x^k) + \sum_l p_l(c_l - y_l) = \sum_k [U_k(x^k) - q^k x^k] + \sum_l p_l c_l.$$

Here we have invoked (4). The algorithm below has link prices following a subgradient of the Lagrangian dual, and sources choosing the rate that instantaneously maximizes $U_k(x^k) - q^k x^k$.

Algorithm 2 (Dual, [6])

$$\dot{p}_l = \gamma_l [y_l - c_l]^+_{p_l}, \tag{7}$$

$$x^k = f_k(q^k) := [(U')^{-1}(q^k)]^+_{x_k}. \tag{8}$$

Route adaptation

We now discuss how to update $\alpha^d_{i,j}$. The intuition is to gradually shift traffic from more expensive to cheaper routes. Specifically, for each destination d and node i the vector (over j) of derivatives $\{\dot{\alpha}^d_{i,j}\}$ should satisfy:

- $\{\dot{\alpha}^d_{i,j}\}$ is a function of the current ratios $\{\alpha^d_{i,j}\}$ and the prices $\{p_{i,j} + q^d_j\}$.
- $\{\dot{\alpha}^d_{i,j}\}$ is negatively correlated with the route prices, and maintains node balance:

$$\sum_{(i,j)\in\mathcal{L}} \dot{\alpha}^d_{i,j}(p_{i,j} + q^d_j) \le 0, \qquad \sum_{(i,j)\in\mathcal{L}} \dot{\alpha}^d_{i,j} = 0. \tag{9}$$

- The inequality in (9) only becomes an equality if $\{\dot{\alpha}^d_{i,j}\} = 0$, and this happens only if for each $(i,j) \in \mathcal{L}$ we have either $q^d_i = p_{i,j} + q^d_j$, or $\alpha^d_{i,j} = 0$ and $q^d_i < p_{i,j} + q^d_j$.

4 Convergence results

We overview here the theoretical results that can be obtained on the global behavior of the system under the above decentralized control algorithms. Proofs are found in [9]. The first result concerns the primal laws:

Theorem 1 *Under (5-6), and the assumptions on the adaptation of α, the system converges globally to a solution of Problem 2.*

The method of proof is to show that under the proposed laws, the surplus S is increasing over system trajectories, and to carefully analyze the conditions under which $\dot{S} \equiv 0$.

The analysis of the dual dynamics under route adaptation is more delicate, since both dynamics contribute in an opposite way to the evolution of the Lagrangian. Indeed while simulation evidence suggests convergence except in some some special degenerate netowrks, we do not have at the time of writing a general convergence result. If, however, routes are adapted much more slowly than prices (a reasonable situation in practice), one can first analyze the dynamics for fixed α's, and show that it converges to the solution of

$$\Psi(\alpha) := \max \sum_k U_k(x^k), \qquad \text{subject to } y_l \le c_l, \text{ flow balance (1) and splitting constraints (2).}$$

Considering now the α dynamics at a slower time-scale, we have the following.

Theorem 2 *For each set of split ratios α, define $\Psi(\alpha)$ as above. Updating α through (9), $\Psi(\alpha)$ converges to its global maximum, the solution to Problem 1.*

5 Implementation

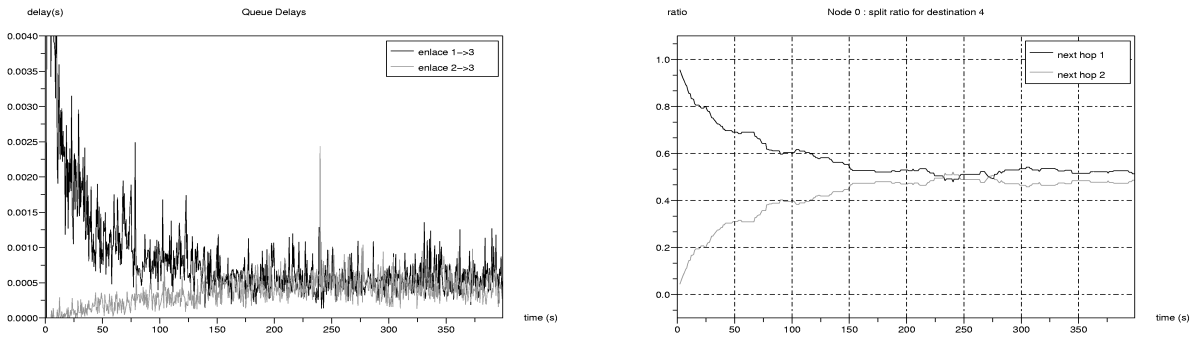
This section describes work in progress towards a packet implementation of these methods, suitable for operation in real TCP/IP networks; the aim is to introduce the smallest amount of changes to current protocols. The algorithms are being implemented in the standard ns-2 simulator [8]. The main features are:

- Formation of node prices. Given link prices generated at each router, the corresponding node prices that satisfy (3) are found iteratively: each node periodically updates q_i^d to the right-hand side of (3), based on announcements of neighboring nodes and its own link prices, and then announces its new price to its neighbors. Under the assumption of continued connectivity, this recursion converges. The message passing is exactly the same as in the well-known distance-vector routing protocol RIP; we have thus modified this module from the ns-2 distribution.

- Extended routing tables. For each destination address, we store multiple outgoing links and their split ratios. A pseudo-random generator and comparators determine packet forwarding decisions.

- Communication of the price to the sources, and congestion control. An explicit communication from the edge router is a significant departure from current practice, and might be too slow to control a TCP source, given the rate of convergence of routing updates. We thus propose to use prices the source can directly estimate: in particular, an attractive option is the queueing delay, used in recently proposed TCP variants [5]. Other possibilities are discussed in [9].

We present some preliminary results for the simple symmetric network topology shown in Figure 2. For the moment, we focus on routing and use standard TCP for the congestion control portion. There are two bottlenecks, depicted by thin links, and a single source. Initially, the router at node 0 sends all traffic through the top route. As time progresses, congestion is detected in the top bottleneck and our multipath routing algorithm begins to use the bottom link as well, eventually settling for an even split of traffic. Figure 1 depicts the evolution of bottleneck queueing delays (our link price variables) and the split ratios at node 0.



(a) Bottleneck queueing delays

(b) Split ratios at node 0

Figure 1: ns-2 simulation results

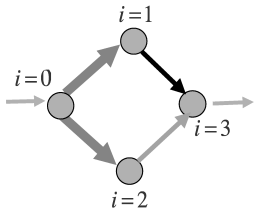


Figure 2: Simulated network topology

6 Conclusions and future work

We have formulated optimization problems that maximize aggregate utility or surplus of a network under multipath routing and congestion control, and proposed decentralized algorithms that provably converge to the global optima, under certain time-scale assumptions. Open questions remain on the behavior of our algorithms in the absence of such restrictions.

From a practical perspective, the algorithms involve moderate extensions of currently available protocols; we presented initial experimental results performed on the ns-2 packet simulator. We will further develop this implementation in future work.

References

[1] L. Chen, S. H. Low, M. Chiang and J.C. Doyle, "Cross-layer congestion control, routing and scheduling design in ad-hoc wireless networks", to appear in IEEE INFOCOM 2006.

[2] B. Fortz, M. Thorup,"Internet Traffic Engineering by Optimizing OSPF Weights", IEEE Infocom 2000.

[3] R. G. Gallager, "A minimum delay routing algorithm using distributed computation", *IEEE Trans. on Communicactions*, Vol Com-25 (1), pp. 73-85, 1977.

[4] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability", *Jour. Oper. Res. Society*, vol. 49(3), pp 237-252, 1998.

[5] C. Jin, D. X. Wei and S. H. Low, "FAST TCP: motivation, architecture, algorithms, performance"; *IEEE Infocom*, March 2004

[6] S. H. Low and D. E. Lapsley, "Optimization flow control, I: basic algorithm and convergence", *IEEE/ACM Transactions on Networking*, vol.7, no.6,pp861-874, December 1999.

[7] S. Low, F. Paganini, J. Doyle, "Internet Congestion Control", *IEEE Control Systems Mag.*, Feb 2002.

[8] ns-2 network simulator(ver 2) , URL: http://www.isi.edu/nsnam/ns/.

[9] F. Paganini, "Congestion control with adaptive multipath routing based on optimization", *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, Mar 2006.

[10] A. Sridharan, R.Guerin, C. Diot. "Achieving Near-Optimal Traffic Engineering Solutions for Current OSPF/ISIS Networks". IEEE/ACM Transactions on Networking. March 2005

[11] R. Srikant, *The Mathematics of Internet Congestion Control,* Birkhauser, 2004.

[12] T. Voice, "A global stability result for primal-dual congestion control algorithms with routing", *ACM Sigcomm CCR, 2004.*

[13] J. G. Wardrop, "Some theoretical aspects of road traffic research", *Proc. Inst. of Civil Engineers*, pt II., vol 1, 325-378, 1952.

[14] Y. Xi and E. Yeh, "Node-Based Distributed Optimal Control of Wireless Networks", *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, Mar 2006.