Optimization Flow Control with Estimation Error

Mortada Mehyar Demetri Spanos Steven H. Low
Engineering & Applied Science
California Institute of Technology
{morr, demetri, slow}@caltech.edu

Abstract— **We analyze the effects of price estimation error in a dual-gradient optimization flow control scheme, and characterize the performance of the algorithm in this case. By treating estimation error as inexactness of the gradient, we utilize sufficient conditions for convergence subject to bounded error to characterize the long-term dynamics of the link utilization in terms of a region which the trajectory enters in finite time. We explicitly find bounds for this region under a particular quantization error model, and provide simulation results to verify the predicted behavior of the system. Finally, we analyze the effects of the stepsize on the convergence of the algorithm, and provide analytical and numerical results which suggest a particular choice for this parameter.**

I. Introduction

It has been shown that Internet congestion control protocols (e.g., various versions of TCP) can be interpreted as carrying out a distributed algorithm to solve a network optimization problem (see [7], [9], [8], [10] and references therein). This paper focuses on the dual-gradient algorithm presented in [9]. The structure of the dual problem allows us to solve the optimization problem with each source (link) knowing only the sum of the link prices (source rates) coupled to it. It is shown in [9] that under appropriate conditions the dual-gradient method converges to the primal-dual optimal point, which is the equilibrium of the network.

One practical drawback of this dual-gradient method is the reliance on explicit communication of price information. Schemes such as RED [5], REM [2], and RAM [1] use a congestion-based queue-management protocol which, in the above context, amounts to an implicit price-notification scheme. Although this mechanism is more practical than the explicit transmission of price information, it suffers from various errors inherent in the implicit price-notification. Our aim in this paper is to examine the effects of such errors on the performance of the algorithm presented in [9].

One particular source of error inherent to any physical implementation is the limited information available to individual sources, i.e., the receipt of acknowledgments and the round-trip-time (RTT) for each packet transmitted. The prices (congestion measures) in these two cases are, respectively, loss probability and queueing delay. The exact price is either very hard to estimate (loss probability) or very noisy (queueing delay).

The essential idea in the following analysis is to think of error as inexactness of the gradient. It is known in optimization theory that convergence to a region containing the optimum can still be achieved in the presence of inexact gradient information [3].

II. Basic Model, Assumptions, and Notation

As in [9] and [10], consider a network that consists of a set $L = \{1, \ldots, L\}$ of unidirectional links of capacities c_l, $l \in L$. The network is shared by a set $S = \{1, \ldots, S\}$ of sources. Source s is characterized by four parameters $(L(s), U_s, m_s, M_s)$. The path $L(s) \subseteq L$ is a set of links that source s uses, $U_s : \Re_+ \to \Re$ is a utility function, $m_s \geq 0$ and $M_s < \infty$ are the minimum and maximum transmission rates, respectively, required by source s. Source s attains a utility $U_s(x_s)$ when it transmits at rate x_s that satisfies $m_s \leq x_s \leq M_s$.

We make the following assumptions on the utility functions:

C1: On the interval $I_s = [m_s, M_s]$, the utility functions U_s are increasing, strictly concave, and twice continuously differentiable. For feasibility, assume $\sum_{s \in S(l)} m_s \leq c_l$ for all l.

C2: The curvatures of U_s are bounded away from zero on I_s: $-U_s''(x_s) \geq 1/\overline{\alpha}_s > 0$ for all $x_s \in I_s$.

For each link l let $S(l) = \{s \in S \mid l \in L(s)\}$ be the set of sources that use link l. Define $\overline{L} := \max_{s \in S} |L(s)|$, $\overline{S} := \max_{l \in L} |S(l)|$, and $\overline{\alpha} := \max\{\overline{\alpha}_s,\ s \in S\}$. In words $\overline{L}$ is the length of a longest path used by the sources, $\overline{S}$ is the maximum number of sources sharing any particular link, and $\overline{\alpha}$ is the upper bound on all $-U_s''(x_s)$. We also will make use of the smallest minimum rate, and largest maximum rate, which we denote by $m = \min_s m_s$ and $M = \max_s M_s$.

Our objective is to choose source rates x_s to solve the following optimization problem:

$$\max_{x_s \in I_s} \quad \sum_s U_s(x_s) \tag{1}$$

$$\text{subject to} \quad \sum_{s \in S(l)} x_s \leq c_l, \quad l = 1, \ldots, L. \tag{2}$$

A unique maximizer, called the primal optimal solution, exists since the objective function is strictly concave, and hence continuous, and the feasible solution set is compact.

The dual function can be calculated to be (see [9]):

$$D(p) = \sum_s \left(U_s(x_s(p)) - x_s(p)p^s\right) + \sum_l p_l c_l$$

0-7803-8355-9/04/$20.00 ©2004 IEEE.

where

$$p^s = \sum_{l \in L(s)} p_l, \forall s$$
$$x_s(p) = U_s'^{-1}(\sum_{l \in L(s)} p_l) \quad (3)$$

Notice that (3) defines a source algorithm for the selection of a sending rate as a function of the congestion measure p.

The partial derivatives of the dual function are found to be:

$$\frac{\partial D}{\partial p_l}(p) = c_l - y_l(p)$$

where

$$y_l(p) := \sum_{s \in S(l)} x_s(p)$$

is the aggregate rate on link l. Note that the evaluation of the gradient (which occurs at the links) requires only knowledge of the aggregate rates at each link. This motivates the following link algorithm (which amounts to a gradient-projection method) for the solution of the dual optimization problem:

$$p_l(t+1) = [p_l(t) - \gamma(c_l - y_l(t))]^+ \quad (4)$$

where $[z]^+ = \max\{z, 0\}$.

It is shown in [9] that if

$$0 < \gamma < \frac{2}{\bar{a}\bar{L}\bar{S}} \quad (5)$$

this algorithm will drive the rates to the optimum of the primal problem (1–2), and the prices to the associated Lagrange multipliers.

III. Price Estimation Error as Inexact Gradient

As described in the introduction, the above algorithm requires exact communication of aggregate price information to individual sources. This requirement is impractical on a real network because it cannot be implemented in the TCP layer. This has motivated price-based congestion notification schemes such as REM and RAM, which allow implicit communication of price information through packet dropping or marking via ECN [4]. These implicit schemes, though more practical than explicit communication, suffer from inherent error in price notification.

The convergence analysis in [9] provides no mechanism for understanding the effects of inaccurate price information on the performance of the algorithm. In particular, it is unclear that convergence should persist in the case that the source algorithm does not set individual sending rates to the exact rates corresponding to the current price. Such an occurrence is inevitable in the case of inexact price communication.

It is evident from (4) that the price update is dependent on the rate update. Thus, an erroneous rate update will result in a corresponding error in the price update. In particular, the direction of the price update will, in general, not be along the gradient direction, but along some perturbed direction. *Thus, the effect of inexact price estimation at the sources amounts to an inexact calculation of the gradient at the links.*

The advantage of the inexact gradient viewpoint is that it allows us to embed the phenomenon of price estimation error in the optimization flow control framework. We will show that in the presence of error, the above algorithm will still converge to a region around the optimum, under a slight modification of the stepsize bound.

IV. Attraction under Inexact Gradient

In this section we will characterize the steady-state dynamics of link utilization in terms of an *attraction region*, which we define as follows:

Definition: A set $A_l \subset \Re_+$ is called an attraction region for link l if there exists an integer N such that for all initial conditions (source rates and link prices), $y_l(n) \in A_l$ for some n less than N.

We remark on two important subtleties. First, this definition does not require that the trajectory remain within the attraction region after entering. It is thus not required to be an invariant set. Second, this *does* imply that if the trajectory ever leaves the attraction region, it will return to the attraction region within N steps.

We will show that as long as the relative error is bounded, the optimization flow control scheme in [9] will still converge in the sense that it will drive the link utilization to an attraction region. The core of the following argument is that reduction of the dual function can still be achieved in the presence of inexact gradient calculation.

At each time t, the lth component of the exact gradient is given by

$$g_l(t) = c_l - \sum_{s \in S(l)} U_s'^{-1}(p(t))$$

Let $v(t)$ be the estimation error, and define the estimated price by $\tilde{p}(t) = p(t) + v(t)$. Hence, the rate update is $x_s(t) = U_s'^{-1}(\tilde{p}(t))$. Thus, the inexact gradient link l actually uses is

$$\tilde{g}_l(t) = c_l - \sum_{s \in S(l)} x_s(t) = c_l - \sum_{s \in S(l)} U_s'^{-1}(\tilde{p}(t))$$

The error in the lth component of the gradient is therefore bounded by

$$\begin{aligned} |\tilde{g}_l(t) - g_l(t)| &= \left| \sum_{s \in S(l)} [U_s'^{-1}(p(t)) - U_s'^{-1}(\tilde{p}(t))] \right| \\ &\leq \sum_{s \in S(l)} \left| [U_s'^{-1}(p(t)) - U_s'^{-1}(\tilde{p}(t))] \right| \\ &\leq \sum_{s \in S(l)} \alpha_s |v_s(t)| \end{aligned}$$

where $1/\alpha_s$ is the lower bound on the curvature of $U_s(x)$ (see [9]) and therefore α_s is a global Lipschitz constant for $U_s'^{-1}(p(t))$ by the Mean Value Theorem.

The following is a sufficient condition which guarantees that the inexact gradient will still be in a descent direction:

$$\sum_{s\in S(l)} \alpha_s |v_s(t)| \le \eta |c_l - y_l(t)|, \forall l. \quad (6)$$

where $0 \le \eta < 1$ can be thought of as the relative error. This condition simply ensures that the error is not large enough to completely negate the gradient, and so the dual function can still be reduced in the direction of the inexact gradient.

When inequality (6) is not satisfied, no conclusion can be drawn, as the above condition is merely sufficient for convergence. Nonetheless, we can show that the region where (6) fails, i.e., where the following holds

$$\sum_{s\in S(l)} \alpha_s |v_s(t)| > \eta |c_l - y_l(t)|, \text{for some } l \quad (7)$$

contains an attraction region.

Theorem 1: The solution set of (7) is an attraction region, provided

$$0 < \gamma < \frac{2}{\bar{\alpha}\bar{L}\bar{S}}(1-\eta) \quad (8)$$

Proof. The condition in the definition of an attraction region will be verified in two steps:

a) Choice of Stepsize: Since $\nabla D = g$ (the exact gradient) is Lipschitz with a Lipschitz constant $\bar{\alpha}\bar{L}\bar{S}$ [9], the Descent Lemma ([3] proposition A.24) implies

$$D(p - \gamma\tilde{g}) \le D(p) - \gamma\langle g, \tilde{g}\rangle + \frac{\bar{\alpha}\bar{L}\bar{S}}{2}\gamma^2\|\tilde{g}\|^2 \quad (9)$$

where $\langle g, \tilde{g}\rangle$ is the Euclidean inner product and $\|g\|$ is the Euclidean norm. Then we see

$$0 < \gamma < \frac{2}{\bar{\alpha}\bar{L}\bar{S}}\frac{\langle g, \tilde{g}\rangle}{\|\tilde{g}\|^2}$$

guarantees that the change in the dual function

$$\Delta D := D(p - \gamma\tilde{g}) - D(p)$$

is strictly less than 0. Therefore when (6) holds, the best bound on γ that guarantees descent is the solution of the following optimization problem:

$$\begin{array}{ll} \min_g & \frac{2}{\bar{\alpha}\bar{L}\bar{S}}\frac{\langle g, \tilde{g}\rangle}{\|\tilde{g}\|^2} \\ \text{subject to} & \|\tilde{g}_l - g_l\| \le \eta\|\tilde{g}_l\|, \forall l \end{array}$$

When g and $\tilde{g}$ are two dimensional (i.e., when there are two links in the network), the feasible region is the shaded area illustrated in Figure 1.

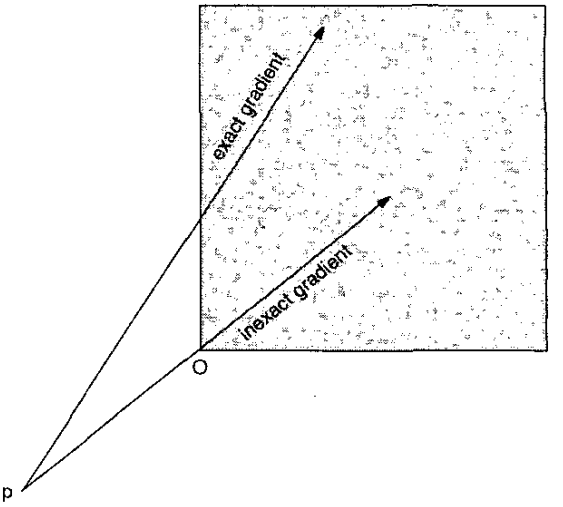

Fig. 1. The exact gradient lies within the shaded region. Notice that the region is rectangular because of the componentwise error bound.

Since $\langle g, \tilde{g}\rangle = \sum_l g_l\tilde{g}_l$, it is easy to see that the minimum occurs at point O where

$$\langle g, \tilde{g}\rangle = (1-\eta)\|\tilde{g}\|^2 \quad (10)$$

and therefore the minimum bound for γ that guarantees descent is $\frac{2}{\bar{\alpha}\bar{L}\bar{S}}(1-\eta)$.

b) Entry in Finite Steps: From (9) we see that the minimal decrease of the dual function in each step is

$$\begin{aligned} \Delta D &\le \frac{1}{2}\bar{\alpha}\bar{L}\bar{S}\|\tilde{g}\|^2\gamma\left(\gamma - \frac{\langle g, \tilde{g}\rangle}{\|\tilde{g}\|^2}\frac{2}{\bar{\alpha}\bar{L}\bar{S}}\right) \\ &\le \frac{1}{2}\bar{\alpha}\bar{L}\bar{S}\|\tilde{g}\|^2\gamma\left[\gamma - (1-\eta)\frac{2}{\bar{\alpha}\bar{L}\bar{S}}\right] \\ &\le \frac{1}{2}\bar{\alpha}\bar{L}\bar{S}\|\tilde{g}\|^2\left[\gamma^2 - (1-\eta)\frac{2\gamma}{\bar{\alpha}\bar{L}\bar{S}}\right] \end{aligned} \quad (11)$$

where $\|\tilde{g}\|^2$ is *strictly* positive since (6) holds and $|v(t)|$ is in general not trivially zero. Therefore as long as $0 < \gamma < \frac{2}{\bar{\alpha}\bar{L}\bar{S}}(1-\eta)$, the dual function is decreased by a *finite* amount in each iteration. Now since the primal problem (1–2) is, by hypothesis, feasible, the dual function is lower bounded [3]. Therefore the inequality (6) must fail after a *finite* number of steps, or it would contradict the fact that the dual function is lower bounded. In other words, (7) must hold after a finite number of steps, i.e., the trajectory of $y_l(t) = \sum_{s\in S(l)} x_s(t)$ enters the solution set of (7) after a finite number of steps. ■

Some comments on the relationship between η and γ are now in order. First, note that larger values of η will result in smaller solution sets for (7). Of course, in order to guarantee that this is an attraction region, (8) must be satisfied. So, a larger η corresponds to a tighter attraction region, but demands a smaller γ. Conversely, given a choice of γ, (8) constrains the maximal η consistent with the above analysis, and hence the smallest obtainable attraction region.

The above proof demonstrates that convergence can be guaranteed with *any* γ satisfying (8), but like the convergence proof in [9], it does not suggest a criterion for selecting γ within this range. As is usual in iterative optimization algorithms, there is a tradeoff between taking larger steps at each iteration (i.e., selecting a large γ) and ensuring that the (inexact) gradient remains a good predictor of local function behavior (i.e., selecting a small γ).

It turns out that we can obtain a satisfactory optimality result which strongly suggests choosing

$$\gamma_{opt} := \frac{(1-\eta)}{\bar{\alpha}\bar{L}\bar{S}}$$

which happens to be half of the bound imposed by the convergence criterion.

Theorem 2: The choice of stepsize γ_{opt} has the following properties:

a) Worst-case optimality: It is the worst-case optimal γ, in the sense that it maximizes the Lipschitz-bounded progress in the dual function at each iteration.

b) Superiority to smaller γ: At each iteration, γ_{opt} generates a new price with a smaller value of the dual function than any smaller choice of γ.

Proof.

a) From (11), the worst-case progress of the dual function is

$$\frac{\bar{\alpha}\bar{L}\bar{S}}{2}\|\tilde{g}\|^2\left[\gamma^2 - \frac{2\gamma}{\bar{\alpha}\bar{L}\bar{S}}(1-\eta)\right] \qquad (12)$$

A simple calculation shows that the minimum of the quadratic as a function of γ occurs when $\gamma = \gamma_{opt}$.

b) Consider the directional derivative $\frac{d}{d\gamma}D(p-\gamma\tilde{g}) = -\langle\nabla D(p-\gamma\tilde{g}),\tilde{g}\rangle$. The Lipschitz property of ∇D implies:

$$\|\nabla D(p-\gamma\tilde{g}) - \nabla D(p)\| \leq \bar{\alpha}\bar{L}\bar{S}\gamma\|\tilde{g}\| \qquad (13)$$

Therefore the magnitude of the difference between the directional derivatives at p and $p-\gamma\tilde{g}$ is

$$\|\langle\nabla D(p-\gamma\tilde{g}) - \nabla D(p),\tilde{g}\rangle\| \leq \bar{\alpha}\bar{L}\bar{S}\gamma\|\tilde{g}\|^2$$

Here we have used the Cauchy-Schwarz inequality and the Lipschitz bound. This in turn implies:

$$\begin{aligned}\frac{d}{d\gamma}D(p-\gamma\tilde{g}) &\leq -\langle\nabla D(p),\tilde{g}\rangle + \bar{\alpha}\bar{L}\bar{S}\gamma\|\tilde{g}\|^2 \\ &= -\langle g,\tilde{g}\rangle + \bar{\alpha}\bar{L}\bar{S}\gamma\|\tilde{g}\|^2 \\ &\leq \left[-(1-\eta) + \bar{\alpha}\bar{L}\bar{S}\gamma\right]\|\tilde{g}\|^2 \\ &= (\gamma-\gamma_{opt})\bar{\alpha}\bar{L}\bar{S}\|\tilde{g}\|^2\end{aligned}$$

In the second last line we have applied (10), since we are only interested in points where the algorithm has not driven the system to the attraction region and can hence provably decrease the dual function at each iteration.

Thus, whenever γ is chosen to be smaller than γ_{opt}, the derivative of the dual function with respect to γ is strictly negative. This implies that γ_{opt} achieves greater decrease in the dual function at each iteration than any smaller γ.

■

See section VIII for discussion and numerical verifications of these properties.

With these two theorems in hand, we are ready to understand the behavior of the optimization flow control algorithm in specific cases of error models.

V. Constant Error

Here we analyze the explicit price-notification scenario, in which links directly notify sources of the current price by encoding this information in a packet (as in the experiments of [9]). Even this explicit communication scheme suffers from the inherent error of finite information transmission, i.e., that the packets which convey the price information contain finitely many bits.

Suppose the range of price the users respond to is between 0 and 1 which is the case when the congestion measure is loss probability. Suppose also that the explicit price feedback has n bits of data. Then the error has a constant bound $|v(t)| \leq 2^{-n}$ and from (7) the inequality for an attraction region for each link l is thus given by

$$\sum_{s\in S(l)} \alpha_s 2^{-n} > \eta\,|c_l - y_l(t)|$$

Figure 2 illustrates the form of these attraction regions when there are two links.

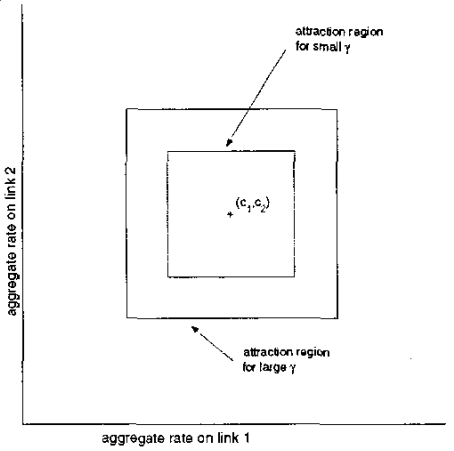

Fig. 2. Attraction regions corresponding to different γ in the constant error case.

VI. Quantization Error of Marking

When the congestion measure p is loss probability, e.g., when routers implement RED or REM, during each RTT the source sends out one window size w of packets and has to estimate p by observing how many packets are dropped or marked. The fraction of packets lost is an instantaneous estimator of p and is subject to two kinds of errors: quantization and probabilistic fluctuation. For example if $w = 4$, then $\tilde{p} \in \{0, 0.25, 0.5, 0.75, 1\}$. So, if the actual price occurs at some intermediate value, say $p = .3$, the closest one could estimate would be $\tilde{p} = .25$. We call this an error resulting from

quantization. Further, due to the probabilistic nature of the dropping scheme, we could get (albeit with lower probability), say $\tilde{p} = 1$ as the estimate of p and incur a larger error in the effective gradient. We call this the fluctuation error.

This quantization error is inevitable under such a price estimation scheme. Naturally, in a probabilistic dropping scheme, one would also suffer from the fluctuation error, but in this section we analyze the effects of quantization in isolation, as it is the minimal error that could occur.

The quantization error at each time, $|v_s(t)|$, will be bounded by $\frac{1}{2d_s x_s(t)}$, where d_s is the RTT of user s which is assumed to be constant. Therefore the condition (7) becomes

$$\sum_{s \in S(l)} \frac{\alpha_s}{2\eta d_s x_s(t)} > \left| c_l - \sum_{s \in S(l)} x_s(t) \right|, \forall l. \tag{14}$$

We will apply this to derive the attraction regions for various cases.

Single-Source-Single-Link

In this simple case an analytical characterization of an attraction region on the link will be derived. The condition (7) reduces to

$$\frac{\alpha}{2\eta d x(t)} > |c - x(t)| \tag{15}$$

Therefore when $x \geq c$, it becomes $c > x - \frac{\alpha}{2\eta d x}$ and when $x < c$, it becomes $c < x + \frac{\alpha}{2\eta d x}$. In the first case there is always the solution $c \leq x < \frac{c}{2} + \frac{1}{2}\sqrt{c^2 + \frac{2\alpha}{\eta d}}$. In the second case the inequality is solvable if and only if $c > \sqrt{\frac{2\alpha}{\eta d}}$ and the solution to the inequality is $0 \leq x < \frac{c}{2} - \frac{1}{2}\sqrt{c^2 - \frac{2\alpha}{\eta d}}$ and $\frac{c}{2} + \frac{1}{2}\sqrt{c^2 - \frac{2\alpha}{\eta d}} < x < c$. This is illustrated in Figure 3. Notice that the solution set includes a "small window region" where the extremely small rate can result in very large error which overcomes the gradient information. However, it is not difficult to show (by realizing that persistence in this region will drive the price to zero) that the trajectory will eventually escape this region. Thus, the part of the solution set containing capacity (as shown) is indeed an attraction region and is given by

$$\left(\frac{c}{2} + \frac{1}{2}\sqrt{c^2 - \frac{2\alpha}{\eta d}}, \frac{c}{2} + \frac{1}{2}\sqrt{c^2 + \frac{2\alpha}{\eta d}} \right) \tag{16}$$

Intuitively, the quantity $\sqrt{\frac{2\alpha}{\eta d}}$ is a measure of the aggressiveness of the protocol. If the capacity is greater than this quantity, the analysis shows that the attraction region is given by (16). However, when the capacity is not greater than this threshold, the attraction region loses the lower bound and becomes $[0, \frac{c}{2} + \frac{1}{2}\sqrt{c^2 + \frac{2\alpha}{\eta d}})$, which suggests that the system can oscillate all the way from 0 to $\frac{c}{2} + \frac{1}{2}\sqrt{c^2 + \frac{2\alpha}{\eta d}}$. This situation is illustrated in Figure 4.

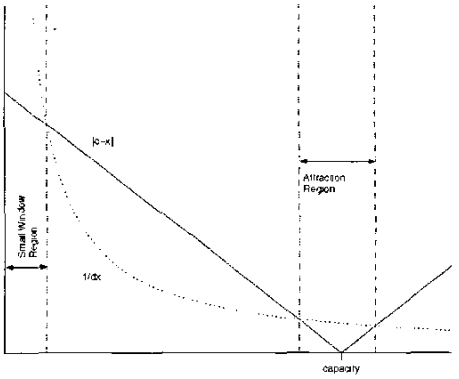

Fig. 3. Illustration of the attraction region.

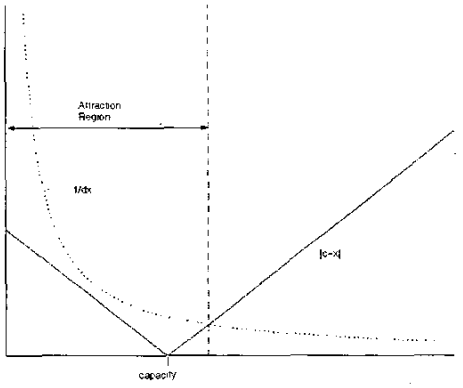

Fig. 4. Attraction region with capacity below threshold.

Multiple-Source-Single-Link

Suppose there are N sources sharing one link. Then (7) becomes

$$\sum_{s=1}^{N} \frac{\alpha_s}{2\eta d_s x_s(t)} > \left| c - \sum_{s=1}^{N} x_s(t) \right| \tag{17}$$

Let $b = \frac{1}{2\eta}\sqrt{\sum_{s=1}^{N} \left(\frac{\alpha_s}{d_s}\right)^2}$ then the left hand side of (17) becomes b multiplied by a convex combination of all the $\frac{1}{x_s}$ and, through simple calculus, is bounded by $\frac{(M+m)^2}{4Mm}\frac{N}{y(t)}$. Therefore (7) implies

$$\frac{(M+m)^2}{4Mm}\frac{bN}{y(t)} > |c - y(t)| \tag{18}$$

Using the single-source-single-link results, the bounds for an attraction region for $y(t)$ are thus

$$\frac{c}{2} + \frac{1}{2}\sqrt{c^2 \pm bN\frac{(M+m)^2}{4Mm}}$$

where, as in the single-source-single-link case, whenever the quantity in the square root is negative, the lower bound is replaced by 0. These results will be verified by the simulations in section VII.

Multiple-Source-Multiple-Link

When there are multiple links, we can obtain a region (using the same manipulations from the previous section) for each link and the bounds are

$$\frac{c}{2} + \frac{1}{2}\sqrt{c^2 \pm b_l |S(l)|^2 \frac{(M(l)+m(l))^2}{4M(l)m(l)}}$$

where $m(l) = \min_{s(l)} m_s$, $M(l) = \max_{s(l)} M_s$, and $|S(l)|$ is the number of sources sharing link l. Unfortunately, the proof of Theorem 1 only allows us to guarantee that *one* of these regions will be an attraction region. This is a serious theoretical drawback, but the simulation results in the upcoming section suggest that these regions actually still characterize the dynamics of multiple links quite well.

VII. Simulation Results

In this section, we provide numerical results which corroborate the previous analysis. Throughout, we set $\eta = 0.5$.

We first simulate the single-source-single-link case. The utility function is chosen to be $U(x) = -\frac{x^2}{2(M-m)} + \frac{Mx}{M-m}$ where $M = 100$, $m = 1$, and we set $d = 1$. Note that this function is strictly concave, and satisfies $0 \leq U'(x) \leq 1$. Further, the constant α is given by $M - m$. The simulations implement the synchronous version of the protocol proposed in [9], with the addition of the deterministic quantization error discussed prior. At each time t the estimated price $\tilde{p}(t)$ is given by the point closest to $p(t)$ in the set $\{0, \frac{1}{x}, \frac{2}{x}, ...\frac{x-1}{x}, 1\}$.

The simulation results in Figure 5 verify the predicted bounds on the attraction region. The capacity is set to 50. The results corroborate the above analysis; the system converges to within the predicted bounds, and oscillates unpredictably thereafter.

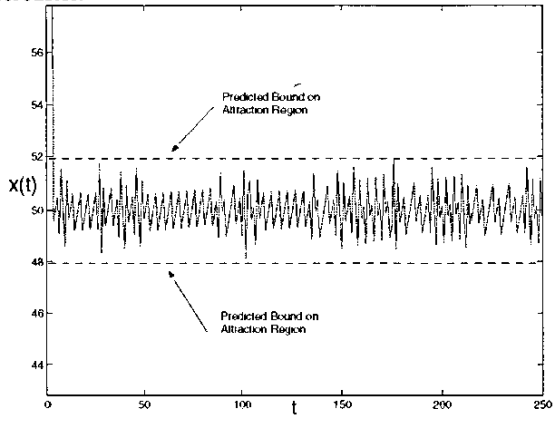

Fig. 5. When capacity is above threshold, the attraction region is bounded both from below and from above.

Figure 6 shows the behavior of the system when the capacity is below the threshold. Here, the attraction region is no longer bounded from below, and indeed the rate drops all the way to the minimum rate from time to time. Notice that there are large one-step excursions from the attraction region but the trajectory always returns to it.

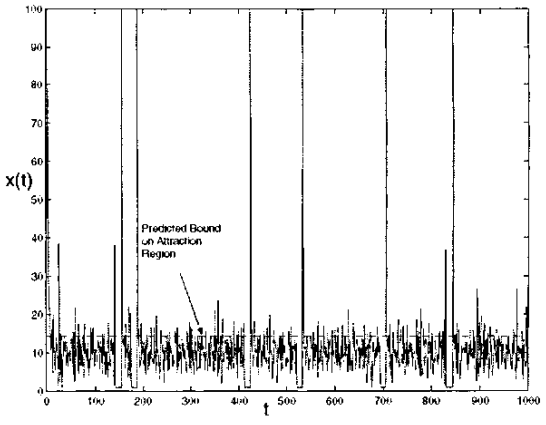

Fig. 6. When capacity is below threshold, the attraction region is no longer lower bounded and the rate oscillates wildly.

Figure 7 is a "phase-plane" diagram showing several trajectories of the system, and their convergence to a set that resembles a parallelogram. The choice of initial conditions of p_l and x_s does not affect the inexact gradient argument and therefore Theorem 1 implies that all such trajectories enter the attraction region.

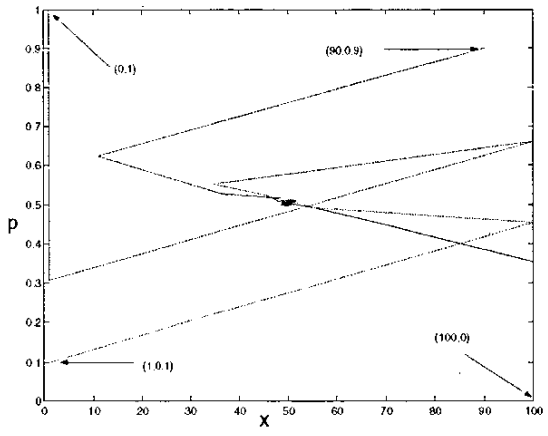

Fig. 7. Convergence with different initial points in the x-p plane.

Figure 8 shows the results from a two-source-one-link simulation. The capacity is set to 70 and the utility functions are such that $m_1 = m_2 = 10$, $M_1 = 50$, and $M_2 = 100$. The dashed lines show the predicted bounds from the above analysis. While we have not derived any bounds on the individual source rates, the simulation suggests that the variation in the individual rates is bounded by the variation in the aggregate rate.

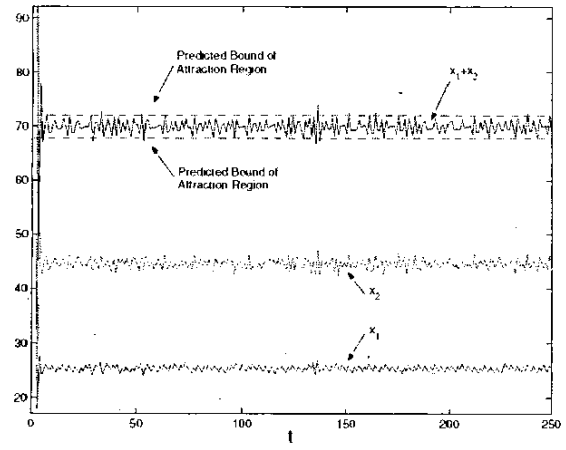

Fig. 8. Bounds on the attraction region in the multiple-source-single-link case.

Finally we present simulation results for a multiple-source-multiple-link network, illustrated in Figure 9, where link 1 is shared by x_1 and x_3 and link 2 is shared by x_2 and x_3.

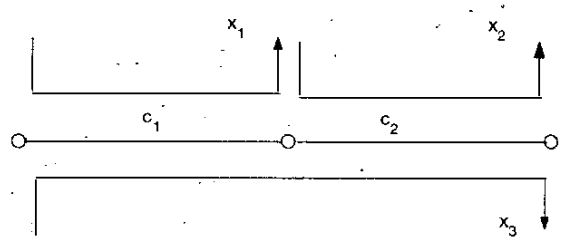

Fig. 9. Three-source-two-link Network.

We set $m_1 = m_2 = m_3 = 3$, $M_1 = M_2 = M_3 = 100$, $c_1 = 90$, $c_2 = 60$, and illustrate the attraction regions on each link in Figure 10. We also present the individual source rates in Figure 11.

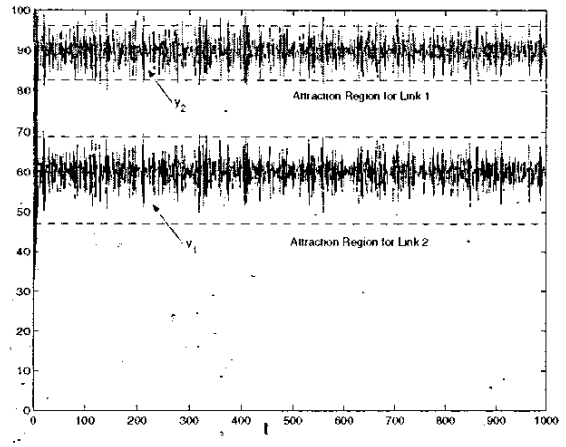

Fig. 10. Bounds on the attraction regions in the multiple-source-multiple-link case.

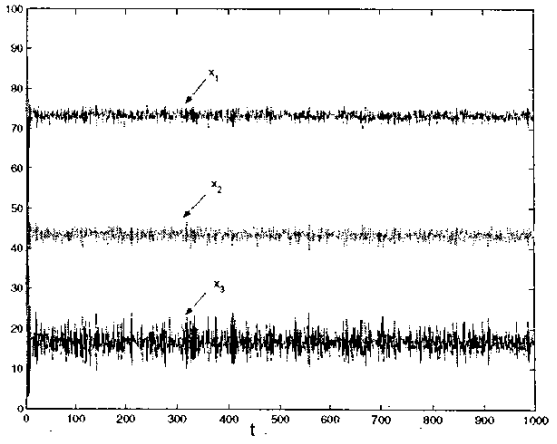

Fig. 11. Dynamics of individual source rates.

VIII. Choice of the Stepsize γ

The properties guaranteed by Theorem 2 strongly suggest the use of γ_{opt} in the absence of any additional information about the optimization problem (1–2). We can conclusively say that no γ below this value should be chosen, and the worst-case analysis suggests that we do not choose a larger γ if we are interested in provable convergence behavior.

In this section, we present simulation results which explore the performance of different choices of γ, and verify the aforementioned properties.

The scenario in [9] (i.e., with no estimation error) is a special case of our formulation, in which (6) always holds with $\eta = 0$. Our stepsize analysis thus suggests that we take precisely half of the bound presented in [9]. The single-source-single-link simulation with a quadratic utility function in Figure 12 verifies the optimality properties of γ_{opt}.

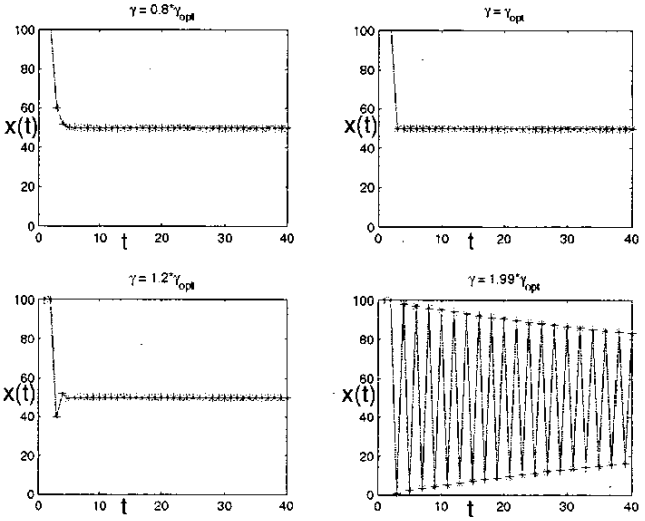

Fig. 12. Convergence behavior with different choices of γ without error: γ_{opt} is clearly superior to other choices in this case.

Note first that, as predicted, all choices of γ below γ_{opt} are outperformed by γ_{opt}. Further note that all larger values are also outperformed, although this is not guaranteed in general.

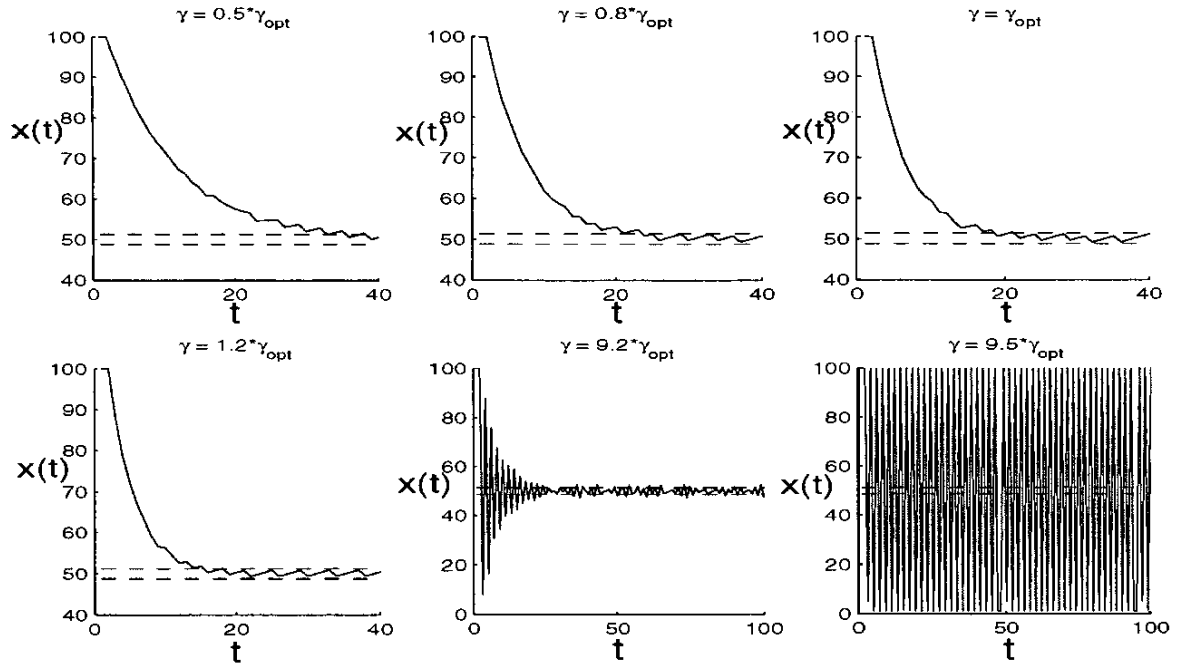

Fig. 13. Convergence behavior with different choices of γ when quantization error is present: Notice that since $\eta = 0.8$, γ_{opt} is $\frac{1}{10}$ the allowable upper bound on γ in the no-error case.

This global superiority holds because, in the case of quadratic utility functions without error, the Lipschitz bounds utilized in the analysis are in fact tight.

The simulations also show that, as γ approaches the bound imposed by the convergence criterion, increasingly large oscillations ensue. It can in fact be shown that the bound imposed by the convergence condition, while only sufficient in general, is also necessary in this case.

We now turn our attention to the case with quantization error as in the simulations of section VII. As shown in Figure 13, we compare the performance of γ_{opt} against that of a range of other choices above and below it. Here the attraction region (denoted by the dashed lines in each subgraph) corresponds to $\eta = 0.8$.

Once again, we verify the prediction that γ_{opt} outperforms all lower values of γ. However, we note in this case that some larger values of γ outperform γ_{opt}. Also note that the transition to instability occurs significantly above the bound we provide for provable convergence in the presence of error. Nonetheless, it is still below the bound presented in [9].

IX. Conclusion and Future Work

We have described an inexact-gradient approach to analyze the effects of price estimation error on the optimization flow control scheme in [9]. We have examined specific error models and derived attraction regions that characterize the oscillation of link utilization. In particular, the attraction region seems to have characterized the oscillation induced by quantization error very well. The results hold for any network topology but unfortunately the dual function argument we use in the proof of theorem 1 can only guarantee that one of the link rates has an attraction region given by (7). However, this seems to be the most general result one can get since the analysis does not require any assumptions on what the error can be. We hope to extend Theorem 1 (possibly with additional assumptions on the error process) to cover all links in the multiple-link case in the future.

For more general sources of stochastic error, we do not expect this analysis to provide such tight bounds on the attraction region, because we have taken a purely deterministic worst-case approach. Indeed, the analysis is most appropriate for error sources with relatively flat probability distributions, as heavy-tail distributions would necessitate a highly conservative bound. For example, the aforementioned probabilistic fluctuation error would result in essentially trivial bounds since the worst-case error could in principle wipe out any useful information about the real price. However, we have seen in our simulations that if an empirical bound (e.g. three times the standard deviation) is taken then the attraction region characterizes the fluctuation error case very nicely as well.

In finding sufficient conditions for convergence subject to error, we have derived a modified bound on the stepsize presented in [9], which shows a linear tradeoff between allowable stepsize and allowable (relative) error. In attempting to find a non-heuristic choice for this stepsize parameter, we have provided a candidate which demonstrates some optimality properties. We have categorically eliminated all smaller stepsizes as sub-optimal, and also shown that this choice gives

0-7803-8355-9/04/$20.00 ©2004 IEEE.

the best provable reduction in the dual function within the scope of our analysis.

X. Acknowledgement

We thank the reviewers for pointing out some errors and giving us useful suggestions.

References

[1] Micah Adler, Jin-Yi Cai, Jonathan K. Shapiro, and Don Towsley, "Estimation of Congestion Price Using Probabilistic Packet Marking", *Proceedings of INFOCOM*, 2003

[2] Sanjeewa Athuraliya, Victor H. Li, Steven H. Low, and Qinghe Yin, "REM: active queue management", *IEEE Network*, 15(3):48–53, May/June 2001. Extended version in *Proceedings of ITC17*, Salvador, Brazil, September 2001.

[3] D. Bertsekas, *Nonlinear Programming*, Athena Scientific 1995

[4] S. Floyd, "TCP and Explicit Congestion Notification", *ACM Computer Communication Review*, 24(5), October 1994

[5] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance", *IEEE/ACM Trans. on Networking*, 1(4):397–413, August 1993.

[6] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1952.

[7] Frank P. Kelly, Aman Maulloo, and David Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability", *Journal of Operations Research Society*, 49(3):237–252, March 1998.

[8] S. Kunniyur and R. Srikant, End-to-end congestion control schemes: utility functions..random losses and ECN marks, *Proceedings of IEEE Infocom*, March 2000.

[9] S.H. Low and D.E. Lapsley, "Optimization Flow Control, I: Basic Algorithm and Convergence", *IEEE/ACM Transactions on Networking* 7(6):861-75, Dec. 1999

[10] S.H. Low, "A Duality Model of TCP and Queue Management Algorithms", To appear in *IEEE/ACM Trans. on Networking* October 2003