Volume Charging Revisited

Costas Courcoubetis
Department of Informatics
Athens University of Economics and Business
Email: courcou@aueb.gr

Antonis Dimakis
Department of Informatics
Athens University of Economics and Business
Email: dimakis@aueb.gr

***Abstract*—We consider fair sharing of network resources over a period of time between offline and interactive users. The former value total transferred volume while the latter have time-dependent valuations for instantaneous bandwidth. Following the work of Kelly on fair bandwidth sharing, we express the system optimum as an equilibrium of traditional congestion control taking place in a fast timescale and a volume charging mechanism in a slow timescale. The problem is very much related to multipath congestion control where paths are unfolded also in the dimension of time, and our algorithms can be regarded as multipath congestion control algorithms operating on two separate timescales. A novel feature of our algorithms is that, contrary to ordinary dual algorithms, users are not overcharged at any time while approaching the equilibrium. This is a sensible economical property not recognized before in the context of congestion control over long timescales. Our results provide the theoretical foundation for designing the 'worse than best-effort' traffic class for handling p2p traffic.**

I. Introduction

Undoubtely one of the essential features of the internet architecture is the ability to accommodate needs of very diverse applications. Connection rates differ by few orders of magnitude, while file transfer sizes vary by more than ten orders of magnitude. Nevertheless this is achieved using only a handful of transport protocols, mainly TCP and its variants, which in essense allocate network bandwidth continuously so as to achieve *instantaneous* fair sharing. Not all applications value instantaneous bandwidth equally though. This is valued more by browsing applications than, say, volume intense applications such as peer-to-peer file sharing. The latter are indifferent to small temporal variations of their bandwidth share, provided the volume downloaded over a long time window is unchanged.

Indeed, this indifference can be exploited to the benefit of browsing users, here called *interactive*, since they would always prefer their downloads finishing sooner rather than later. Since TCP-based congestion control is concerned only with how instantaneous bandwidth is shared, different mechanisms are needed to bridge this efficiency gap. In a way, such mechanisms have already been employed by network operators when placing volume caps for peer-to-peer file sharing traffic during peak hours. This is achieved by middleboxes inside the network performing deep packet inspection []. Since knowledge about time preferences resides on the user side, network operators are not always knowledgable in their traffic shaping decisions. In this paper we favour a user-based approach for sharing bandwidth in time.

We propose new congestion control algorithms to be used by volume intense applications, called *offline* in this paper, while interactive applications can use any TCP-based congestion control algorithm. Using the optimization framework of Kelly [1] we formulate fair sharing in time as optimizing a system objective function. The analysis suggests congestion control algorithms quite distinct from TCP-based ones. In fact, our algorithms operate on two timescales. On the fast one, comparable to a few round-trip times, offline applications will transmit as fast as they can provided congestion, as signalled e.g., by loss rate, delay, or ECN marking, does not exceed a certain threshold. On the slow timescale, comparable to a day or week, applications update their thresholds according to the volume received during a correspondingly long time window. A similar type of flow control for the offline applications, the 'worse than best effort' traffic class, has already been proposed by practitioners for handling p2p traffic over DSL connections.

Kelly in [1] gives an economic interpretation to congestion control. Users continuously adjust their bandwidth share according to congestion signals acting as instantaneous prices so as to maximize their surplus. Eventually, the process converges to a bandwidth allocation that maximizes social welfare. Our algorithms admit a similar interpretation in terms of volume charging. Offline users pay a constant price per unit of bandwidth determined by the value of their threshold, throughout the duration of a long time-period which can be seen as a billing cycle. At the end of each cycle, users adjust the prices they are willing to pay in the next cycle, in response to the difference between the total received volume during the previous cycle and their demand for that price. Interactive users continue to adjust their bandwidth share in the traditional way. Eventually, the bandwidth allocations converge to the social optimum.

Since the algorithm operates at slow timescales, e.g., over days, weeks or even months, it makes sense that an offline user should refrain from downloading if anytime during a billing cycle he has received a volume equal to his demand at the declared price. This behavior is not exhibited by traditional dual algorithms where prices are controlled by excess demand because there the received volume should be able to exceed demand temporarily during transient periods. For the case of a single link we show the existence of a simple algorithm that prevents this type of overcharging at all times. Such a constraint has not been considered before and we believe it is essential for any pricing algorithm operating over long timescales. It is worth noting here that in fast timescales overcharging is not important since transient periods are too short to have any significant impact on performance.

Key et al. [2] consider a similar problem phrased in terms of average bandwidth instead of volume over long time windows. In fact they propose algorithms very similar to our dual algorithm in section V. A key difference in our work is that volume is measured in fixed large windows instead of being averaged continuously as in [2]. This enables us to track downloaded volume explicitly and hence prevent overcharging as described above.

Another feature of our approach that differentiates ordinary congestion control with the one used by the offline users is that the allocations achieved by the latter in the fast timescales are essentially nonunique even at equilibrium. This is because of the congestion level threshold type of algorithm used: there are many allocations of rates among the offline users that achieve the same level of congestion. In contrast, in TCP-based algorithms the bandwidth share is (in general) a function of the loss rate alone, as in the square-root law, and is unique. It is by the operation of the price adjustment in the slow timescale that unique volume allocations over long periods are achieved. To address the issue of nonuniqueness -not present in ordinary congestion control mechanisms [3]- we make use of differential inclusions.

In section II we introduce the model and notation along with the main optimization problems. In section III an example is given illustrating the algorithms and main results. A time decomposition approach is described in section IV. The main results here are Theorems 2 and 4 which parallel the problem decomposition in [1]. In section V we describe two congestion control algorithms for offline users. The first is a dual type of algorithm and the proof given holds for arbitrary networks. The second algorithm prevents overcharging and the proof given is for the case of a single link. Global asymptotic stability and optimality of the equilibria are the subjects of Propositions 4.1,4.2 for the respective algorithms. Finally, we summarize our work in section VI.

II. The model

Consider discrete time over time periods $t = 1, \dots, T$. Let J be the set of links, with each $j \in J$ having capacity $C_j > 0$. Consider the set of routes R and the incidence matrix $A = (A_{jr}, j \in J, r \in R)$ with $A_{jr} = 1$ or 0 according to whether route r uses link j. We define two sets of users:

1) *Offline* users, by which we model file transfers invoked by a peer-to-peer file sharing application. Let $\mathcal{O}$ be the set of offline users.
2) *Interactive* users, a model for a series of sessions initiated by a human user through a web browser. Let $\mathcal{I}$ be the set of interactive users.

For each user $r \in \mathcal{O} \cup \mathcal{I}$ let $x^r = (x_t^r, t = 1, \dots, T)$ denote the bandwidth allocated to r over time period $\{1, \dots, T\}$, and $u_r : \mathbb{R}_+^T \to \mathbb{R}$ be his utility function. If $r \in \mathcal{O}$ then $u_r(x^r) = U_r(x_1^r + \dots + x_T^r)$, while for $r \in \mathcal{I}$ we have $u_r(x^r) = \sum_{t=1}^T U_r(t, x_t^r)$. The functions $U_r : \mathbb{R}_+ \to \mathbb{R}$ for $r \in \mathcal{O}$, as well as $U_r(t, \cdot) : \mathbb{R}_+ \to \mathbb{R}$ for $r \in \mathcal{I}, t = 1, \dots, T$ unless they are identically zero, are assumed to be increasing, strictly concave and continuously differentiable, properties inherited by $u_r : \mathbb{R}_+^T \to \mathbb{R}$.

The problem we are interested to solve is:

$$\begin{aligned} \text{SYSTEM} : \max & \sum_{r \in \mathcal{O} \cup \mathcal{I}} u_r(x^r) \\ & \text{such that } \sum_{j \in J} A_{jr} x_t^r \le C_j,\ \forall j \in J,\ t = 1, \dots, T \\ & \text{over } x = (x_t^r, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \dots, T) \ge 0 \,. \end{aligned}$$

As in [1] we will see that solutions of the above problem are characterized in terms of fair bandwidth sharing. For weight w^r associated with each flow $r \in \mathcal{O} \cup \mathcal{I}$ let the NETWORK problem be

$$\begin{aligned} \text{NETWORK}(w^r, r \in \mathcal{O} \cup \mathcal{I}) : \max & \sum_{r \in \mathcal{O} \cup \mathcal{I}} w^r \log x_r \\ & \text{such that } \sum_{j \in J} A_{jr} x_r \le C_j,\ \forall j \in J \,, \\ & \text{over } (x_r, r \in \mathcal{O} \cup \mathcal{I}) \ge 0 \,. \end{aligned}$$

Each time instant t, users change their weights in response to congestion or price signals $(\lambda_t^j, j \in J) =: \lambda_t$ posted by links. If user r knew the entire sequence $\lambda := (\lambda_t, t = 1, \dots, T)$ then he would choose weight w_t^r at each t such that he solves:

$$\begin{aligned} \text{USER}_r(\lambda) : \max u_r & \left(\left(\frac{w_t^r}{\sum_{j \in J} A_{jr} \lambda_t^j}, t = 1, \dots, T \right) \right) - \sum_{t=1}^T w_t^r \\ & \text{over } (w_t^r, t = 1, \dots, T) \ge 0 \,. \end{aligned}$$

For future reference we define also the following problem for scalar $p \ge 0$:

$$\begin{aligned} \text{V-USER}_r(p) : \max u_r(x^r) - p \sum_{t=1}^T x_t^r \\ \text{over } (x_t^r, t = 1, \dots, T) \ge 0 \,, \end{aligned}$$

and let $y^r(p)$ be the point where the optimum is achieved.

This problem corresponds to picking the optimal bandwidth allocation if volume is charged at p per unit of volume.

To motivate the results in section IV we consider an example first.

III. Example

Consider the two-link network in Fig. 1(a) consisting of three users; one offline $\mathcal{O} = \{1\}$ using both links, and two interactive users $\mathcal{I} = \{2, 3\}$ with each one using a different link. We consider two periods, i.e., $T = 2$.

The SYSTEM problem is:

$$\begin{aligned} \max U_1(x_1^1 + x_2^1) + \sum_{r \in \mathcal{I}} \sum_{t=1}^2 U_r(t, x_t^r) \qquad (1) \\ \text{such that } x_t^1 + x_t^2 \le C_1 \,, \quad x_t^1 + x_t^3 \le C_2,\ t = 1, 2. \end{aligned}$$

Kelly in [1] has interpreted solutions to such problems as equilibria of congestion control algorithms. The problem

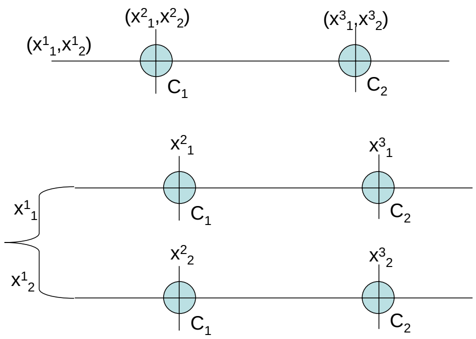

Fig. 1. (a) original network, multi-period problem. (b) alternative network, single-period problem.

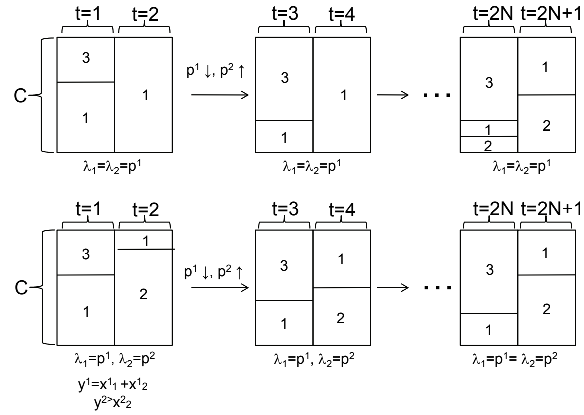

Fig. 2. (up) dual type price adjustment (down) price adjustment without overcharging.

above can be thought of as a single-period problem for an alternative network making it thus amenable for analysis through Kelly's framework. Such a network is depicted in Fig. 1(b). We are interested in finding congestion control equilibria that solve (1). Following the framework laid down by Kelly yields the USER$_1$ problem:

$$\max_{z\geq 0} U_1\left(\frac{z}{\min_{t=1,2}\lambda_t^1+\lambda_t^2}\right)-z\,, \tag{2}$$

where λ_t^j is the price per unit of bandwidth at link j at time t. The solution of (2) has economic significance as optimal willingness-to-pay of user 1 for total volume if link prices were known to him. However, it does not possess any congestion control significance, i.e., does not yield an optimal weight w_t^1 in response to current link price $\lambda_t^j, j=1,2$ at each time $t=1,2$. Moreover, the optimum of (2) depends on link prices of both periods $t=1,2$. This is not the case for the interactive users which solve

$$\max_{z\geq 0} U_r\left(t,\frac{z}{\lambda_t^1+\lambda_t^2}\right)-z\,, r=2,3,\ t=1,2\,, \tag{3}$$

a problem which has a direct interpretation as the equilibrium of a congestion control algorithm (see [1]).

Albeit not able to provide a congestion control characterization of (1) directly, we do this for a parameterized approximation to (1). For scalar parameter $p\geq 0$ consider

$$\begin{aligned}&\max p(x_1^1+x_2^1)+\sum_{r\in\mathcal{I}}\sum_{t=1}^{2}U_r(t,x_t^r)\\&\text{such that } x_t^1+x_t^2\leq C_1\,,\quad x_t^1+x_t^3\leq C_2\,, t=1,2\,.\end{aligned} \tag{4}$$

Notice that the optimal allocation (x_1^1,x_2^1) according to (1) satisfies $\partial U_1(x_1^1+x_2^1)/\partial x_1^1=\partial U_1(x_1^1+x_2^1)/\partial x_2^1$. This is because the offline user is indifferent in transferring his desired amount in any of the two time periods as long as he is utilizing the cheapest periods. Hence, by picking $p=U_1'(x_1^1+x_2^1)$ (6) obtains the same solutions as (1). The problem of course, is that $U_1'(x_1^1+x_2^1)$ is still not known ahead of time. However, the dependence between $t=1,2$ is now captured in a single parameter p.

Fixing some arbitrary $p\geq 0$ the problem (6) separates for $t=1$ and $t=2$, and applying Kelly's decomposition for (6) yields the following congestion control equilibria at each time: User 1 solves,

$$\max_{z_t\geq 0}\frac{p}{\lambda_t^1+\lambda_t^2}z_t-z_t\,, t=1,2\,. \tag{5}$$

These optima are not unique, but coupled with interactive users' congestion control (3) give unique equilibrium weight parameters for each $t=1,2$. The intuition behind (5) is the following: at the fast timescale user 1 may see that $\lambda_t^1+\lambda_t^2<p$. (5) indicates that he should increase weight z_t, hence the transmitted rate. But this will drive $\lambda_t^1+\lambda_t^2$ up until the point that is equal to p. Notice that (5) is not a prescription for a congestion control algorithm as (3); it merely characterizes equilibria. A possible primal algorithm achieving (5) at equilibrium, in the case of a single link with a single offline user is $\hat{z}=p/\lambda(z)-1$, where $\lambda(z)$ is link price resulting from using weight z.

Now if $\lambda_t^1+\lambda_t^2>p$, i.e., congestion is high, then $z_t=0$, i.e., the offline user will not send any data during period t. Thus, at the end of the two periods, user 1 will have paid $p(x_1^1+x_2^1)$ in total. That is, p expresses his willingness-to-pay for units of volume. For any such p, the allocations resulting from the corresponding congestion controls will solve (6). At the end of the two periods, user 1 can increase or decrease p according to whether his excess demand $y^1(p)-(x_1^1+x_2^1)$ is >0 or <0. As we show in section V a continuous time version of the price adjustment above, converges to the correct p so SYSTEM is solved at equilibrium.

Let us give an example of how this price adjustment works. Consider the case of a single link with $\mathcal{O}=\{1,2\}$, $\mathcal{I}=\{3\}$ and $T=2$. Also assume $U_3(2n,\cdot)\equiv 0, n=1,2,\ldots$, i.e., the interactive user is not present during even periods. The upper part of figure 2 depicts the bandwidth alloted to each used for the first N billing cycles (each consisting of $T=2$ periods) using the dual type algorithm mentioned above. Analogously to (6), the fast timescale equilibrium allocation arising in the

k-th cycle are determined by now solving

$$\max \sum_{r=1,2} p^r (x^r_{2k} + x^r_{2k+1}) + \sum_{t=2k,2k+1} U_3(t, x^3_t) \quad \text{such that } x^1_t + x^2_t + x^3_t \le C_1 \ , t = 2k, 2k+1 \ . \tag{6}$$

It is assumed that $p^1 > p^2$ during the first cycle, so user 2 will not download anything during the first cycle. This is because user 1 sends at a high enough rate so the link price becomes p^1. User 2 chooses not to send at that congestion level since $p^2 < p^1$. User 1 competes with the interactive user only for the bandwidth of period $t = 1$, while in $t = 2$ he has no competitor at this congestion level. Thus, most likely the volume downloaded by user 1 exceeds by far his demand $y^1(p^1)$ since the received volume during $t = 2$ does not depend on p^1 as long as $p^1 > p^2$. Now at the end of the first cycle, user 2 will increase p^2 while user 1 will decrease p^1 (provided $y^1(p^1) < x^1_1 + x^1_2$). At $t = 3$ user 1 takes a smaller portion of capacity, while he grabs the entire capacity at $t = 4$ as long as p^1 is still greater than p^2. This pattern will continue to occur in subsequent cycles until $p^2 = p^1$. At this point, (6) supports multiple allocations between the offline users. This is because any fixed congestion level can be caused by multiple allocations. Hence, the allocation that prevails is determined by other factors, such as packet level dynamics. This nonuniqueness is not an artefact of the model but an essential feature of the linear decomposition (6). As we saw above, it is precisely this linear structure that permits us to decompose SYSTEM into separate problems for each time period and hence obtain simple congestion control algorithms. In the next section we comment more on the necessity of this linarity.

The problem with the above dual price adjustment algorithm is that $y^1(p^1) < x^1_{2k} + x^1_{2k+1}$ might continue to occur over many cycles before $p^1 = p^2$, as explained above. In the context of economics, user 1 would not want to download any volume in excess of his demand since that would have been a rational decision only at a lower price. Since price is being adjusted over very long timescales, a dual algorithm is not desirable. In section V we propose another algorithm that places the constraint that the downloaded volume within a cycle cannot exceed the demand. After cycles for which the volume is less than demand, price is increased according to their difference as in the dual algorithm. If during a cycle, the constraint placed by demand is met, the price to be used at the next cycle is decreased by a fixed amount. The behavior of this algorithm is depicted in the lower part of Fig. 2. During $t = 2$ the volume constraint of user 1 is met and user 2 will grab the rest of the capacity since user 1 will not compete until the end of this cycle. The latter will decrease his price p^1 and will be able to download higher volume during the second cycle. Eventually, the same equilibrium with the dual algorithm is reached.

In the next section we formalize these ideas.

IV. Time decomposition approach

In this section we describe a time decomposition approach for solving SYSTEM. We do this by defining a different problem V-SYSTEM(p), which for an appropriate value of $p = (p^r, r \in \mathcal{O})$ has the same solution as SYSTEM. But the solution of V-SYSTEM can be done at each time t independently, i.e., it decouples in the fast timescale. By following the traditional approach by Kelly, V-SYSTEM suggests congestion control algorithms that have its optimum as their equilibrium. There is an algorithm for the fast timescale to be used by all users and another for the slow timescale to be used by the offline users. Unfortunately, these congestion control algorithms are not practical to implement for economic reasons. Although they reach the right equilibrium, during the transient phase they may result in huge budget expenses on behalf of the offline users. Therefore we need to define a carefully altered version of V-SYSTEM that avoids these pitfalls.

As in section III consider the problem

$$\text{V-SYSTEM}(p) : \max \sum_{r \in \mathcal{O}} p^r \sum_{t=1}^{T} x^r_t + \sum_{r \in \mathcal{I}} \sum_{t=1}^{T} U_r(t, x^r_t) \quad \text{such that } \sum_{j \in J} A_{jr} x^r_t \le C_j, \ j \in J, \ t = 1, \ldots, T \ , \quad \text{over } (x^r_t, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T) \ . \tag{7}$$

Since V-SYSTEM decomposes into separate problems for each time t, we can apply Kelly's decomposition, i.e., Theorem 2 in [1], to each such period to obtain the following theorem that suggests the decomposition of the problem into user congestion control actions and network price calculations.

Theorem 1. *For any* $(p^r, r \in \mathcal{O}) \in \mathbb{R}^{\mathcal{O}}_+$ *there exist* $x := (x^r_t, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T)$, $(w^r_t, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T)$, $(\lambda^j_t, j \in J, t = 1, \ldots, T)$ *such that*

1) *For each* $r \in \mathcal{O}, t = 1, \ldots, T$, w^r_t *solves*
$$\max_{z \ge 0} \frac{p^r}{\sum_j A_{jr} \lambda^j_t} z - z \ . \tag{8}$$
2) *For each* $r \in \mathcal{I}, t = 1, \ldots, T$, w^r_t *solves*
$$\max_{z \ge 0} U_r \left(t, \frac{z}{\sum_j A_{jr} \lambda^j_t} \right) - z \ . \tag{9}$$
3) *For each* $t = 1, \ldots, T$, $(x^r_t, r \in \mathcal{O} \cup \mathcal{I})$ *solves NETWORK*$((w^r_t, r \in \mathcal{O} \cup \mathcal{I}))$.
4) $w^r_t = x^r_t \sum_j A_{jr} \lambda^j_t$, *for each* $r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T$.

Moreover, under the above conditions x *solves V-SYSTEM*(p).

Proof: It follows as a direct consequence of Theorem 2 in [1]. ■

In the next theorem we prove existence of parameters p^r such that SYSTEM is optimized by solving V-SYSTEM(p) for this particular p.

Theorem 2. *There exist* $p := (p^r, r \in \mathcal{O}) \in \mathbb{R}^{\mathcal{O}}_+$, $x := (x^r_t, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T)$, $w := (w^r_t, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T)$, $\lambda := (\lambda^j_t, j \in J, t = 1, \ldots, T)$ *such that*

1) x, w, λ *satisfy Theorem 1 for* p. *In particular,* x *solves V-SYSTEM*(p).

2) *For each* $r \in \mathcal{O}$, $\sum_{t=1}^{T} x_t^r$ *solves V-USER*$_r(p^r)$.
Under the above conditions x *solves SYSTEM.*

Proof: Consider the following problem which is equivalent to SYSTEM:

$$\begin{aligned} &\max \sum_{r\in\mathcal{O}} U_r(y^r) + \sum_{r\in\mathcal{I}} u_r(x^r) \\ &\text{such that } \sum_{t=1}^{T} x_t^r = y^r,\ \forall r \in \mathcal{O}, \\ &\qquad \sum_{j\in J} A_{jr} x_t^r \le C_j,\ \forall j \in J,\ t = 1,\dots,T\ , \\ &\text{over } (x_t^r, r \in \mathcal{O}\cup\mathcal{I}, t = 1,\dots,T), (y^r, r \in \mathcal{O})\ . \end{aligned} \tag{10}$$

The dual function obtained by retaining the link constraints is:

$$\begin{aligned} G(p) &= \sum_{r\in\mathcal{O}} \max_{y^r \ge 0} \left[U_r(y^r) - p^r y^r\right] \\ &+ \sum_{t=1}^{T} \max_{(x_t^r): \sum_j A_{jr} x_t^r \le C_j, j\in J} \left[\sum_{r\in\mathcal{O}} p^r x_t^r + \sum_{r\in\mathcal{I}} U_r(t, x_t^r)\right]\ , \end{aligned} \tag{11}$$

where $p := (p^r, r \in \mathcal{O})$ are the (nonnegative) Lagrange multipliers for the equality constraints in (10). The assumptions on utilities imply that strong duality holds, so there exists p such that $G(p)$ attains the optimal value of (10). For this p let $(y^r, r \in \mathcal{O})$, x be the optimum achieving points of the optimization problems in (11). For each $r \in \mathcal{O}$, primal feasibility gives $y^r = \sum_{t=1}^{T} x_t^r$ so this and the leftmost optimization problem in (11) imply that part 2 holds. The rightmost optimization problem in (11) is (7), so Theorem 1 yields $(w_t^r, r \in \mathcal{O}\cup\mathcal{I}), (\lambda_t^j, j \in J)$ satisfying part 1.

Conversely, if condition 1 holds then Theorem 1 implies that x solves the rightmost optimization problem in (11). Condition 2 implies feasibility of the equality constraints in 10 as well as optimality of the leftmost optimization in (11). Hence, Lagrangian sufficiency implies that x solves SYSTEM. ■

Here we need to discuss the implications of Theorem 2 regarding congestion control algorithms. It suggests that in the fast timescale t offline users should solve V-SYSTEM using (8) as their congestion control algorithm starting from some initial choice of p^r, and then slowly adjust p^r (in consecutive periods T) in a way that their resulting total flow within T obtained by the previous procedure solves V-USER$_r$ for the above price.

It is worth noting here that there are not any real alternatives to the linear congestion control of part 1 in Theorem 1 at fast timescales. For example, consider a system with $T = 2, J = \{1\}, \mathcal{O} = \{1\}, \mathcal{I} = \{2\}$, with users executing the following congestion control algorithms at fast timescales:

$$\dot{x}^1(\tau) = \kappa_1 x^1(\tau)(F'(x^1(\tau)) - C(x^1(\tau) + x^2(\tau)))\ , \tag{12}$$
$$\dot{x}^2(\tau) = \kappa_2 x^2(\tau)(U_2'(t, x^2(\tau)) - C(x^1(\tau) + x^2(\tau)))\ , \tag{13}$$

where $x^r(\tau)$ is the instantaneous rate of user r at time τ, and $C(x)$ is a link cost function such as marginal delay when input rate is x. It is easy to see that any equilibrium of (13) necessarily solves (9), while (12) is an alternative to part 1 for some concave function $F(\cdot)$. After a few round-trips, the equilibrium allocations x_t^1, x_t^2 at time t will be such that $F'(x_t^1) = U'_1(t, x_t^2)$. But part 2 of Theorem 2 implies that if $0 < x_1^1 < x_1^1$ then $\lambda_1^1 = \lambda_2^1$, and so $F'(x_1^1) = F'(x_2^1)$. In turn this means that F is linear in $[x_1^1, x_2^1]$. Thus at equilibrium, the congestion control algorithms of the offline users must be in effect linear.

Unfortunately, the congestion control algorithms suggested by Theorem 2 are not economically realistic. This is because during transient phases (8) may lead an offline user to get all the bandwidth of the network resulting in unreasonably high charges in each billing period T. This motivates a careful modification of V-SYSTEM to include a constraint on the maximum download offline users are permitted to do during a billing period. More specifically, if such a user operates its congestion control assuming a price p^r, its maximum volume during T should not exceed his demand $y^r(p^r)$ for the given price p^r. In analogy to V-SYSTEM(p) consider the following sequence of problems indexed by $t = 1,\dots,T$:

$$\begin{aligned} \text{V-SYSTEM}(p,t) : &\max \sum_{r\in\mathcal{O}} p^r z^r + \sum_{r\in\mathcal{I}} U_r(t, z^r) \\ &\text{such that } \sum_{r\in\mathcal{O}\cup\mathcal{I}} A_{jr} z^r \le C_j\ , \quad j \in J \\ &\qquad \sum_{s=1}^{t-1} x_s^r + z^r \le y^r\ , \quad r \in \mathcal{O} \\ &\text{over } (z^r, r \in \mathcal{O}\cup\mathcal{I}) \ge 0\ , \end{aligned} \tag{14}$$

where $(x_s^r, r \in \mathcal{O}\cup\mathcal{I})$ is any solution to V-SYSTEM(p, s) for $s < t$. We say that $(x_t^r, r \in \mathcal{O}\cup\mathcal{I}, t = 1,\dots,T)$ solves V-SYSTEM$(p,\cdot)$ if for each $t = 1,\dots,T$ $(x_t^r, r \in \mathcal{O}\cup\mathcal{I})$ solves V-SYSTEM(p, t).

This leads to Theorem 3, a constrained version of Theorem 1.

Theorem 3 leads to the specific congestion control algorithms discussed in Section V.

Theorem 3. *For any* $(p^r, r \in \mathcal{O}) \in \mathbb{R}_+^{\mathcal{O}}$ *there exist* $x := (x_t^r, r \in \mathcal{O}\cup\mathcal{I}, t = 1,\dots,T)$, $(w_t^r, r \in \mathcal{O}\cup\mathcal{I}, t = 1,\dots,T)$, $(\lambda_t^j, j \in J, t = 1,\dots,T)$ *such that*

1) *For each* $r \in \mathcal{O}, t = 1,\dots,T$, x_t^r *solves*

$$\max p^r z - \sum_j A_{jr} \lambda_t^j z\ ,$$

over $z \ge 0$ *such that,*

$$\sum_{s=1}^{t-1} x_s^r + z \le y^r(p^r)\ . \tag{15}$$

2) *For each* $r \in \mathcal{I}, t = 1,\dots,T$, w_t^r *solves*

$$\max_{z\ge 0} U_r\left(t, \frac{z}{\sum_j A_{jr}\lambda_t^j}\right) - z\ . \tag{16}$$

3) *For each* $t = 1, \ldots, T$, $(x_t^r, r \in \mathcal{O} \cup \mathcal{I})$ *solves NETWORK*$((w_t^r, r \in \mathcal{O} \cup \mathcal{I}))$ *with the additional constraints*

$$x_t^r \leq y^r(p^r) - \sum_{s=1}^{t-1} x_s^r\,, \quad \forall r \in \mathcal{O}\,.$$

4) *For any* $r \in \mathcal{I}$, $w_t^r = x_t^r \sum_j A_{jr}\lambda_t^j$ *while for* $r \in \mathcal{O}$,

$$w_t^r = \begin{cases} x_t^r \sum_j A_{jr}\lambda_t^j\,, & \text{if } \sum_{s=1}^t x_s^r < y^r(p^r) \\ x_t^r p^r\,, & \text{if } \sum_{s=1}^t x_s^r = y^r(p^r), x_t^r > 0\,, \\ 0\,, & \text{if } \sum_{s=1}^t x_s^r = y^r(p^r), x_t^r = 0 \end{cases} \tag{17}$$

for each $t = 1, \ldots, T$.

Moreover, under the above conditions x *solves V-SYSTEM*$(p, \cdot)$.

Proof: Let $(x_t^r, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T)$ be a solution to V-SYSTEM$(p, \cdot)$. By strong duality there exist nonnegative Lagrange multipliers $(\lambda_t^j, j \in J), (q_t^r, r \in \mathcal{O} \cup \mathcal{I})$ satisfying

$$p^r \leq \sum_j A_{jr}\lambda_t^j + q_t^r\,, \quad \forall r \in \mathcal{O} \tag{18}$$

for which,

$$\sum_{r\in\mathcal{O}} \max_{z^r \geq 0} \left(p^r z^r - \sum_j A_{jr}\lambda_t^j z^r - q_t^r z^r \right) + \sum_{r\in\mathcal{I}} \max_{z^r \geq 0} \left[U_r(t, z^r) - \sum_j A_{jr}\lambda_t^j z^r \right] + \sum_j \lambda_t^j C_j + \sum_{r\in\mathcal{O}} q_t^r y^r \tag{19}$$

attains the optimum value of V-SYSTEM(p, t) and $z^r = x_t^r, r \in \mathcal{O} \cup \mathcal{I}$ solve the optimization problems in (19). Also complementary slackness gives

$$\lambda_t^j \left(C_j - \sum_j A_{jr} x_t^r \right) = 0\,, \quad \forall j \in J \tag{20}$$

$$q_t^r \left(y^r - \sum_{s=1}^t x_s^r \right) = 0\,, \quad \forall r \in \mathcal{O}\,. \tag{21}$$

Now define $w_t^r := x_t^r \sum_j A_{jr}\lambda_t^j$ for $r \in \mathcal{I}$, and $w_t^r := x_t^r \left(\sum_j A_{jr}\lambda_t^j + q_t^r \right)$ for $r \in \mathcal{O}$. It is easy to see that (16) holds. To show (15) notice that x_t^r, q_t^r are a primal-dual optimal pair for (15). Moreover, the last observation and (21) imply (17).

It remains to establish part 3. Let $r \in \mathcal{O}$ be such that $w_t^r = 0$. Then, (17) suggests $x_t^r = 0$ when $\sum_{s=1}^t x_s^r = y^r$. If $\sum_{s=1}^t x_s^r < y^r$, (21) gives $q_t^r = 0$ which together with (18) imply $x_t^r = 0$. A similar argument shows that $x_t^r = 0$ if $w_t^r = 0, r \in \mathcal{I}$. Now for $r \in \mathcal{O}$ with $w_t^r > 0$, (17) implies $x_t^r > 0$ and that

$$\frac{w_t^r}{x_t^r} = \sum_j A_{jr}\lambda_t^j + q_t^r\,,$$

must hold. For the same reasons we get $w_t^r / x_t^r = \sum_j A_{jr}\lambda_t^j$ if $r \in \mathcal{I}$. But the last two conditions along with complementary slackness (20)-(21) and dual feasibility (18) establish optimality of $(x_t^r, r \in \mathcal{O} \cup \mathcal{I})$ for NETWORK$(w_t^r, r \in \mathcal{O} \cup \mathcal{I})$.

The converse follows by noting that optimality conditions (15)-(16), primal and dual deasibility (18), and complementary slackness (20)-(21) are satisfied by primal and dual variables x_t, λ_t, q_t for each $t = 1, \ldots, T$. ■

The next theorem is in analogy to Theorem 2, and states that applying the constrained congestion control (for the offline users) in the fast timescales and adapting p^r to solve V-USER$_r$ in the slow time scales solves SYSTEM in equlibrium. As before, the congestion control of the interactive users remains the same as in the classic formulation by Kelly.

Theorem 4. *There exist* $p := (p^r, r \in \mathcal{O}) \in \mathbb{R}_+^{\mathcal{O}}$, $x := (x_t^r, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T)$, $w := (w_t^r, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T)$, $\lambda := (\lambda_t^j, j \in J, t = 1, \ldots, T)$ *such that*

1) x, w, λ *satisfy Theorem 3 for* p, *with all Lagrange multipliers associated with* (15) *equal to zero for all* $t = 1, \ldots, T, r \in \mathcal{O}$.
2) *For each* $r \in \mathcal{O}$, $\sum_{t=1}^T x_t^r$ *solves V-USER*$_r(p^r)$.

Under the above conditions x *solves SYSTEM.*

Proof: Let p, x, w, λ be as in Theorem 2. First note that part 2 follows by the same theorem. Now notice that x satisfies parts 1,3 of Theorem 3 without the volume constraints. But these do hold by part 2 of this theorem. Thus, x is optimal *a fortiori* in the constrained case too. This establishes part 1.

To show the converse, note that x satisfies parts 1,3 of Theorem 3 without the volume constraints, so it must satisfy parts 1,3 of Theorem 1 as well. By Theorem 2 this implies that x solves SYSTEM. ■

In the next section we give two stable algorithms for updating p^r. However, convergence is not smooth for x. This is because of the discontinuity of solutions in part 1 of Theorem 1 as p^r varies. For this reason in section V we provide another algorithm which achieves convergence of x.

V. Algorithms

Theorem 2 gives a characterization of the optimum achieving points of SYSTEM as equilibria of congestion control algorithms in fast and slow timescales. In the fast timescales the second term of the RHS in (11) is maximized for a given value of the Lagrange multiplier p^r using the user algorithms in (8),(9). In slow timescales p^r should be varied in order to induce constraint satisfaction in (11), i.e., to match $y^r(p^r)$ with $\sum_t x_t^r$. Here we give a user-based algorithm that converges to the correct p^r simultaneously for all $r \in \mathcal{O}$.

We consider a scenario where the discrete time τ moves now in slots corresponding to cycles of T slots of the fast timescale t. Such a slow timescale sees changes of parameters p^r that occur over different billing periods T, and in each such period the value of p^r remains constant. Here, p^r is to be interpreted as a willingness-to-pay parameter for units of volume received by user r. After the end of a billing cycle, users increase (or decrease) p^r if the received volume is smaller (or larger) of their optimal demand for such price p^r per unit of volume. This gives rise to the following algorithm in continuous time τ given by a differential inclusion:

$$\text{Algorithm I: } \left(\frac{dp^r_\tau}{d\tau}, r \in \mathcal{O}\right) \in \left\{ \left(\kappa_r \left(y^r(p^r_\tau) - \sum_{t=1}^{T} x^r_t \right), r \in \mathcal{O} \right) \middle| (x^r_t, r \in \mathcal{O}, t = 1, \ldots, T) \text{ solves V-SYSTEM}(p_\tau) \right\}, \quad \tau \in \mathbb{R}_+ , \quad (22)$$

where κ_r is a positive constant. Notice that $\sum_{t=1}^{T} x^r_t$ is the volume received at the current billing cycle and $y^r(p^r)$ is the optimum achieving point of V-USER$_r(p^r)$.

For the map between price p and the set on the RHS of (22) it is not hard to show that it satisfies the conditions (see, e.g., []), of existence of solutions to the above differential inclusion i.e., absolutely continuous trajectories $(p_\tau, \tau \geq 0)$ for which (22) is true τ-a.e.

We will not need any specific solution concept of (22) nor be concerned with uniqueness of solutions, because we are primarily interested in stability of any solution and in properties of equilibria.

Proposition 4.1. *Algorithm I is globally asymptotically stable. At the unique equilibrium $p = (p^r, r \in \mathcal{O})$, any solution of V-SYSTEM(p) that makes both sides of* (22) *equal, solves SYSTEM as well.*

Proof: Let G be as in (11) and consider any absolutely continuous trajectory $(p_\tau, \tau \geq 0)$ of algorithm I for some arbitrary initial condition. First notice that we can assume that for each $r \in \mathcal{O}$, $p^r_\tau \geq \underline{p}^r > 0$ holds for some $\underline{p}^r$. This is because one can pick $\underline{p}^r$ such that $y^r(p^r) - \sum_t x^r_t(p)$ is greater than some constant independent of $p^s, s \in \mathcal{O} \setminus \{r\}$. By Lemma 4.1 G is Lipschitz continuous so $\tau \mapsto G(p_\tau)$ is absolutely continuous. Now fix any τ for which both p_τ and $G(p_\tau)$ are differentiable at τ. (Points without this property are of zero Lebesgue measure.) The time derivative of $G(p_\tau)$ is given by

$$\frac{d}{d\tau} G(p_\tau) = \langle g, \dot{p}_\tau \rangle , \quad \forall g \in \partial G(p_\tau) , \qquad (23)$$

where $\partial G(p_\tau)$ is the subdifferential of G at p_τ where we have made use of Lemma 4.2. Using the fact that G is a pointwise maximum of smooth functions (see Lemma 4.1 below) by Theorem 2.1 in [4] we have that $(-y^r(p^r_\tau) + \sum_t x^r_t(p_\tau), r \in \mathcal{O}) \in \partial G(p_\tau)$ holds, where $(x^r_t(p_\tau), r \in \mathcal{O}, t = 1, \ldots)$ is any solution of V-SYSTEM(p_τ) satisfying $\dot{p}^r_\tau = -y^r(p^r_\tau) + \sum_t x^r_t(p_\tau)$. Hence using this subgradient in (23) yields,

$$\frac{d}{d\tau} G(p_\tau) = -\sum_{r \in \mathcal{O}} \left(y^r(p^r_\tau) - \sum_t x^r_t(p_\tau) \right)^2 < 0 ,$$

except at $p := (p^r, r \in \mathcal{O})$ for which the inequality becomes equality. By Theorem 2, $x(p)$ solves SYSTEM. Also, note that p is the unique price vector with this property since the objective of (10) is strictly concave with respect to $(y^r, r \in \mathcal{O})$.

It remains to show that $p_\tau \to p$ as $\tau \uparrow +\infty$ from any initial p_0. Notice that there exist positive $(\delta^r_t, r \in \mathcal{O}, t = 1, \ldots, T)$ which satisfy $\sum_j A_{jr} \delta^r_t < C_j, \forall j \in J$ and $G(a) \geq \sum_{r \in \mathcal{O}} a^r \sum_t \delta^r_t$ for any $a \in \mathbb{R}^{\mathcal{O}}_+$. Hence G has compact sublevel sets and in particular this is true for $D := \{a \in \mathbb{R}^{\mathcal{O}}_+ | G(a) \leq G(p_0)\}$. For any $\epsilon > 0$ let $B(p, \epsilon)$ be the open ball with radius ϵ around p, and observe that

$$\sup_{\substack{a \in D \setminus B(p,\epsilon) \\ \hat{x} \text{ solves V-SYSTEM}(a)}} -\sum_{r \in \mathcal{O}} \left(y^r(a^r) - \sum_t \hat{x}^r_t \right)^2 < 0 \qquad (24)$$

holds. We prove this claim by contradiction: assume that the supremum is zero. Then there exists sequence $p_n \to \bar{p}$ as $n \to \infty$ for which,

$$\lim_n \sup_{\hat{x} \text{ solves V-SYSTEM}(p_n)} -\sum_{r \in \mathcal{O}} \left(y^r(p^r_n) - \sum_t \hat{x}^r_t \right)^2 = 0 .$$

But using the fact that G is convex and attains its minimum at p we get

$$0 > G(p) - G(p_n) \geq \sum_{r \in \mathcal{O}} \left(-y^r(p^r_n) + \sum_t \hat{x}^r_t(p_n) \right) \cdot (p^r - p^r_n) \underset{n}{\to} 0 ,$$

for a sequence of solutions $\hat{x}^r_t(p_n)$ to V-SYSTEM(p_n), which is a contradiction since $G(p_n) \to G(\bar{p}) > G(p)$ by the uniqueness of optimum p.

Since (24) is true for arbitrary $\epsilon > 0$, p is a limit point of $(p_\tau, \tau \geq 0)$. Moreover, $\lim_\tau G(p_\tau) = G(p)$ using optimality of p and $\dot{G}(p_\tau) \leq 0$ for τ-a.e. If $\bar{p} \neq p$ is also a limit point then $G(\bar{p}) = G(p)$ should hold. But clearly $G(\bar{p}) > G(p)$ so $p_\tau \to p$ holds. ■

Lemma 4.1. *For any positive $\underline{p}^r, r \in \mathcal{O}$, the restriction of G on $\times_{r \in \mathcal{O}} [\underline{p}^r, +\infty)$ is Lipschitz continuous.*

Proof: Observe that G can be represented as $G(p) = \sup_{z \in D} f_z(p)$ for $D = \{(z^r_t, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T) | \sum_j A_{jr} z^r_t \leq C_j, \forall j \in J, t = 1, \ldots, T\}$ and

$$f_z(p) = \sum_{r \in \mathcal{O}} [U_r(y^r(p^r)) - p^r y^r(p^r)] + \sum_{r \in \mathcal{O}} \sum_t p^r z^r_t + \sum_{r \in \mathcal{I}} \sum_t U_r(t, z^r_t) .$$

But $\partial f_z(p)/\partial p^r$ exists at every p and $|\partial f_z(p)/\partial p^r| = |-y^r(p^r) + \sum_t z^r_t|$ which can be bounded above for all $p^r \geq \underline{p}^r$ by a constant independent of $(z^r_t, r \in \mathcal{O} \cup \mathcal{I}, t = 1, \ldots, T)$. ■

Lemma 4.2. *Let $\tau \mapsto p_\tau$ and $\tau \mapsto V(p_\tau)$ be differentiable at p_τ, where V is a convex function. Then,*

$$\frac{d}{d\tau} V(p_\tau) = \langle g, \dot{p}_\tau \rangle , \quad \forall g \in \partial V(p_\tau) .$$

Proof: Let $g \in \partial V(p_\tau)$ be arbitrary. Then,

$$\langle g, \dot p_\tau\rangle \geq \lim_{h\downarrow 0} \frac{V(p_\tau) - V(p_\tau - \dot p_\tau h) + o(h)}{h} = \frac{d}{d\tau}V(p_\tau) = \sup_{\xi\in\partial V(p_\tau)} \langle \xi, \dot p_\tau\rangle \,.$$

Conclude by noting that the expression on the RHS dominates that on the LHS. ∎

As we noted in the last section, the convergence to the correct volume allocation may be slow as slight differences in the p^r parameter across different offline users may make allocations to vary abruptly. Assume that during a cycle of length T the offline users use parameters $p := (p^r, r \in \mathcal{O})$. Then for any offline user r, part 1 of Theorem 3 implies $\sum_{t=1}^T x_t^r(p) \leq y^r(p^r)$. Thus an excess demand type of algorithm, such as Algorithm I, may possess other suboptimal equilibria since the time derivative of p will never be negative. To remedy this we propose Algorithm II which decreases price at an appropriate rate when excess demand is zero.

Consider the following

$$\text{Algorithm II: } \left(\frac{dp_\tau^r}{d\tau}, r\in\mathcal{O}\right) \in \left\{ \left(\frac{y^r(p_\tau^r) - \sum_{t=1}^T x_t^r}{\epsilon^r} - 1, r\in\mathcal{O}\right) \middle| (x_t^r, r\in\mathcal{O}\cup\mathcal{I}, t=1,\ldots,T) \text{ solves V-SYSTEM}(p_\tau,\cdot)\right\}, \quad \tau\in\mathbb{R}_+ \,, \tag{25}$$

where $(\epsilon^r, r \in \mathcal{O})$ are arbitrary positive constants.

Proposition 4.2. *In the case of a single link, i.e., $J = \{1\}$, Algorithm II is globally asymptotically stable. At the unique equilibrium $p = (p^r, r \in \mathcal{O})$, the problem*

$$\begin{aligned} &\max \sum_{r\in\mathcal{O}} U_r(y^r) + \sum_{r\in\mathcal{I}} u_r(x^r) \\ &\text{such that } \sum_{t=1}^T x_t^r + \epsilon^r = y^r,\ \forall r\in\mathcal{O}\,, \\ &\qquad \sum_{r\in\mathcal{O}\cup\mathcal{I}} x_t^r \leq C_1,\ \forall\, t=1,\ldots,T\,, \\ &\text{over } (x_t^r, r\in\mathcal{O}\cup\mathcal{I}, t=1,\ldots,T)\in\mathbb{R}_+^{(\mathcal{O}\cup\mathcal{I})\times\{1,\ldots,T\}}, \\ &\qquad (y^r, r\in\mathcal{O})\in\mathbb{R}_+^{\mathcal{O}} \end{aligned} \tag{26}$$

is solved by $y^r = y^r(p^r), x_t^r = x_t^r(p), r \in \mathcal{O}, t = 1,\ldots,T$, where $(x_t^r(p), r \in \mathcal{O}\cup\mathcal{I}, t = 1,\ldots,T)$ is any solution of V-SYSTEM$(p,\cdot)$ that makes both sides of (25) *equal.*

Proof: Define the function $H : \mathbb{R}_+^{\mathcal{O}} \to \mathbb{R}_+$ by $H(p) := G(p) + \sum_{r\in\mathcal{O}} p^r\epsilon^r$ where $G(p)$ is as in (11). It is readily seen that H is a convex function. We will show that H is a Lyapunov function for algorithm II.

Let $(p_\tau, \tau \geq 0)$ be an absolutely continuous trajectory of algorithm II for some arbitrary initial conditions. The time derivative of $H(p_\tau)$ is given by

$$\frac{d}{d\tau}H(p_\tau) = \sup_{g\in\partial H(p_\tau)} \langle g, p_\tau\rangle \,.$$

Also $\partial H(p) = \{g + (\epsilon^r, r \in \mathcal{O}) | g \in \partial G(p)\}$, where as seen in the proof of Proposition 4.1, $\partial G(p)$ consists of vectors $(-y^r(p^r)+\sum_t \hat x_t^r, r\in\mathcal{O})$ where $(\hat x_t^r, r \in \mathcal{O}\cup\mathcal{I}, t = 1,\ldots,T)$ runs over all solutions of V-SYSTEM(p). Let $(\hat x_t^r, r \in \mathcal{O}\cup\mathcal{I}, t = 1,\ldots,T)$ be any such solution and notice,

$$\begin{aligned} &\langle -y + \sum_t \hat x_t + \epsilon, p_\tau\rangle = \\ &\sum_{r\in\mathcal{O}}\left(-y^r(p^r) + \sum_t \hat x_t^r(p) + \epsilon^r\right)\cdot\left(\frac{y^r(p^r) - \sum_t x_t^r(p)}{\epsilon^r} - 1\right) \\ &= -\sum_{r\in\mathcal{O}} \frac{(y^r(p^r) - \sum_t x_t^r(p) - \epsilon^r)^2}{\epsilon^r} + \sum_{r\in\mathcal{O}} \frac{(y^r(p^r) - \sum_t x_t^r(p))}{\epsilon^r} \\ &\cdot\left(\sum_t \hat x_t^r(p) - \sum_t x_t^r(p)\right) - \sum_{r\in\mathcal{O}}\left(\sum_t \hat x_t^r(p) - \sum_t x_t^r(p)\right) \\ &\leq -\sum_{r\in\mathcal{O}} \frac{(y^r(p^r) - \sum_t x_t^r(p) - \epsilon^r)^2}{\epsilon^r} \end{aligned}$$

by Lemmas 4.3 and 4.4. Hence, $\dot H(p_\tau)$ is nonnegative except when p_τ satisfies,

$$y^r(p^r) = \sum_t x_t^r(p) + \epsilon^r\,, \quad \forall r\in\mathcal{O}\,. \tag{27}$$

We claim that such p exists and is unique. To see this consider problem (26) and its dual function

$$\mathbb{R}_+^{\mathcal{O}} \ni a \mapsto \sum_{r\in\mathcal{O}} [U_r(y^r(a^r)) - a^r y^r(a^r)] + \max_{(z_t^r):\sum_r z_t^r\leq C_1}\left[\sum_{r\in\mathcal{O}} a^r \sum_{t=1}^T z_t^r + \sum_{r\in\mathcal{I}} u^r(z^r)\right] + \sum_{r\in\mathcal{O}} a^r\epsilon^r\,.$$

Notice that the maximization center term is just V-SYSTEM(a). By strong duality there exists $a =: p$ such that any solution $(\hat x_t^r, r \in \mathcal{O}\cup\mathcal{I}, t = 1,\ldots,T)$ of V-SYSTEM(p) along with demand vector $(y^r(p^r), r \in \mathcal{O})$ solve (26). By strict monotonicity of $U_r, r \in \mathcal{O})$ this solution must satisfy the volume constraints in (26) with equality. But $x(p)$ is a solution of V-SYSTEM(p) since the volume constraints in V-SYSTEM(p,t) are inactive. Hence, (27) is satisfied for this p. Moreover this is unique, as problem (26) gives unique $(y^r, r \in \mathcal{O})$ by strict concavity of $U_r, r \in \mathcal{O}$.

It remains to show that p_τ indeed converges to p as $\tau \uparrow +\infty$. First note that the sublevel sets of H are compact since $H(a) \geq G(a)$ for all $a \in \mathbb{R}_+^{\mathcal{O}}$, and G has compact sublevel sets. In particular this is true for $D := \{a \in \mathbb{R}_+^{\mathcal{O}} | H(a) \leq H(p_0)\}$, a fact which implies that the trajectory is confined in D. For any $\delta > 0$ let $B(p, \delta)$ be the open ball with radius δ around p, and observe that

$$\sup_{\substack{a\in D\setminus B(p,\delta)\\ \bar x \text{ solves V-SYSTEM}(a,\cdot)}} -\sum_{r\in\mathcal{O}}\left(y^r(a^r) - \sum_t \bar x_t^r - \epsilon^r\right)^2 < 0 \tag{28}$$

holds. We prove this claim by contradiction: assume that the supremum is zero. Then there exists sequence $p_n \to \bar p$ as

$n \to \infty$ for which,

$$\lim_n \sup_{\bar{x} \text{ solves V-SYSTEM}(p_n,\cdot)} -\sum_{r\in\mathcal{O}} \left(y^r(p_n^r) - \sum_t \bar{x}_t^r - \epsilon^r \right)^2 = 0. \tag{29}$$

Now, using the fact that H is convex and attains its minimum at p we get

$$0 > H(p) - H(p_n) \geq \sum_{r\in\mathcal{O}} \left(-y^r(p_n^r) + \sum_t \hat{x}_t^r(p_n) + \epsilon^r \right) \cdot (p^r - p_n^r) \ . \underset{n}{\to} 0 \ , \tag{30}$$

for any solution $(\hat{x}_t^r(p_n), r \in \mathcal{O}, t = 1,\ldots,T)$ of V-SYSTEM(p_n). But (29) implies the existence of a sequence $(\bar{x}(p_n))$ of solutions to V-SYSTEM$(p_n,\cdot)$ such that for all n large enough, $y^r(p_n) > \sum_t \bar{x}_t^r(p_n)$ for all $r \in \mathcal{O}$. Thus $\bar{x}(p_n)$ solves V-SYSTEM(p_n) since the volume constraints in V-SYSTEM(p_n,t) are inactive for each $t = 1,\ldots,T$, This implies that the expression on the RHS of (30) converges to zero for $\hat{x}(p_n) = \bar{x}(p_n)$. But this is a contradiction since $H(p_n) \to H(\bar{p}) > H(p)$ by the uniqueness of optimum p.

Since (28) is true for arbitrary $\delta > 0$, p is a limit point of $(p_\tau, \tau \geq 0)$. Moreover, $\lim_\tau H(p_\tau) = H(p)$ using optimality of p and $\dot{H}(p_\tau) \leq 0$ for τ-a.e. If $\bar{p} \neq p$ is also a limit point then $H(\bar{p}) = H(p)$ should hold. But clearly, $H(\bar{p}) > H(p)$ so $p_\tau \to p$ holds. ∎

Lemma 4.3. *Let* $(x_t^r(p), r \in \mathcal{O}\cup\mathcal{I}, t = 1,\ldots,T)$, $(\hat{x}_t^r(p), r \in \mathcal{O}\cup\mathcal{I}, t = 1,\ldots,T)$ *solve V-SYSTEM*$(p,\cdot)$ *and V-SYSTEM*(p) *respectively. For every* $t = 1,\ldots,T$*,*

$$\sum_{r\in\mathcal{O}} \hat{x}_t^r(p) \geq \sum_{r\in\mathcal{O}} x_t^r(p) \ , \tag{31}$$

If $y^r(p^r) > \sum_t x_t^r(p)$ *for some* $r \in \arg\max\{p^s | s \in \mathcal{O}\}$, (31) *holds with equality. In particular, equality still holds if the summations in* (31) *are over* $\arg\max\{p^s | s \in \mathcal{O}\}$.

Proof: Fix any $t = 1,\ldots,T$ and consider the problem V-SYSTEM(p,t). By strong duality, there exist Lagrange multipliers $(\mu_t^r, r \in \mathcal{O})$ such that $0 \leq \mu_t^r \leq p^r$ for which the problem

$$\max \sum_{r\in\mathcal{O}} (p^r - \mu_t^r) z_t^r + \sum_{r\in\mathcal{I}} U_r(t, z_t^r) \text{ such that } \sum_{r\in\mathcal{O}\cup\mathcal{I}} z_t^r \leq C, \tag{32}$$

over $(z_t^r, r \in \mathcal{O}\cup\mathcal{I}) \in \mathbb{R}_+^{\mathcal{O}\cup\mathcal{I}}$ is solved by $(x_t^r(p), r \in \mathcal{O}\cup\mathcal{I})$. Also notice that $(\hat{x}_t^r(p), r \in \mathcal{O}\cup\mathcal{I})$ solves

$$\max \sum_{r\in\mathcal{O}} p^r z_t^r + \sum_{r\in\mathcal{I}} U_r(t, z_t^r) \quad \text{such that} \sum_{r\in\mathcal{O}\cup\mathcal{I}} z_t^r \leq C, \ (z_t^r, r \in \mathcal{O}\cup\mathcal{I}) \in \mathbb{R}_+^{\mathcal{O}\cup\mathcal{I}}. \tag{33}$$

(Below we drop dependence of $x_t^r, \hat{x}_t^r$ on p since no confusion arises.) We will compare solutions of (33) with those of (32) by looking at their duals. Both minimize the same objective

$$\max_{(z_t^r, r\in\mathcal{I})\in\mathbb{R}_+^{\mathcal{I}}} \sum_{r\in\mathcal{I}} [U_r(t, z_t^r) - \lambda_t z_t^r] + \lambda_t C \ , \tag{34}$$

but over

$$\{\hat{\lambda}_t \in \mathbb{R}_+ | \hat{\lambda}_t \geq \max\{p^r | r \in \mathcal{O}\}\} \tag{35}$$

and

$$\{\lambda_t \in \mathbb{R}_+ | \lambda_t \geq \max\{p^r - \mu_t^r | r \in \mathcal{O}\}\} \tag{36}$$

for (33) and (32), respectively. If $\hat{\lambda}_t, \lambda_t$ are the respective unique dual solutions then it must be that $\hat{\lambda}_t \geq \lambda_t$. Also, for any $r \in \mathcal{I}$ with $\hat{x}_t^r > 0$ we have $U_r'(t, \hat{x}_t^r) = \hat{\lambda}_t$, while $\hat{x}_t^r = 0$ if $U_r'(t, \hat{x}_t^r) < \hat{\lambda}_t$. Thus, $\sum_{r\in\mathcal{I}} \hat{x}_t^r \leq \sum_{r\in\mathcal{I}} x_t^r$. Now, $\hat{\lambda}_t \geq \max\{p^r | r \in \mathcal{O}\} > 0$ so

$$\sum_{r\in\mathcal{O}} \hat{x}_t^r = C - \sum_{r\in\mathcal{I}} \hat{x}_t^r \geq C - \sum_{r\in\mathcal{I}} x_t^r \geq \sum_{r\in\mathcal{O}} x_t^r \ ,$$

as claimed.

Now suppose $y^r > \sum_t x_t^r$ for some $r \in \arg\max\{p^s | s \in \mathcal{O}\}$. Then by complementary slackness $\mu_t^r = 0$ must hold for each $t = 1,\ldots,T$. Thus the dual deasible regions (35),(36) coincide so $\lambda_t = \hat{\lambda}_t$ for each t. Since $\lambda_t, \hat{\lambda}_t > 0$ we have $\sum_{s\in\mathcal{O}} x_t^s = \sum_{s\in\mathcal{O}} \hat{x}_t^s$ thereby showing the second claim. Finally, note that for any $s \in \mathcal{O}\setminus M$ we have $p^s - \mu_t^s - \lambda_t \leq p^s - \lambda_t < p^r - \lambda_t \leq 0$, so $x_t^s = \hat{x}_t^s = 0$ which proves the last assertion. ∎

Lemma 4.4. *Let* $(\hat{x}_t^r(p), r \in \mathcal{O}\cup\mathcal{I}, t = 1,\ldots,T)$ *be any solution of V-SYSTEM*(p)*. At each regular* $\tau \geq 0$*,* p_τ *satisfies*

$$\sum_{r\in\mathcal{O}} \frac{(y^r(p_\tau^r) - \sum_t x_t^r(p_\tau))}{\epsilon^r} \left(\sum_t \hat{x}_t^r(p_\tau) - \sum_t x_t^r(p_\tau) \right) \leq 0 \ ,$$

for some solution $(x_t^r(p), r \in \mathcal{O}\cup\mathcal{I}, t = 1,\ldots,T)$ *of V-SYSTEM*$(p,\cdot)$

Proof: Define $M := \arg\max\{p^s | s \in \mathcal{O}\}$ and notice that $\hat{x}_t^r(p_\tau) = 0$ for every t and $r \in \mathcal{O}\setminus M$. Hence,

$$\begin{aligned}
&\sum_{r\in\mathcal{O}} \frac{(y^r(p_\tau^r) - \sum_t x_t^r(p_\tau))}{\epsilon^r} \left(\sum_t \hat{x}_t^r(p_\tau) - \sum_t x_t^r(p_\tau) \right) \\
&= \sum_{r\in M} \frac{(y^r(p_\tau^r) - \sum_t x_t^r(p_\tau))}{\epsilon^r} \left(\sum_t \hat{x}_t^r(p_\tau) - \sum_t x_t^r(p_\tau) \right) \\
&\quad - \sum_{r\in\mathcal{O}\setminus M} \frac{(y^r(p_\tau^r) - \sum_t x_t^r(p_\tau))}{\epsilon^r} \sum_t x_t^r(p_\tau) \\
&\leq \sum_{r\in M} \frac{(y^r(p_\tau^r) - \sum_t x_t^r(p_\tau))}{\epsilon^r} \left(\sum_t \hat{x}_t^r(p_\tau) - \sum_t x_t^r(p_\tau) \right)
\end{aligned}$$

Now since $p_\tau^r = p_\tau^s$ for each $r, s \in M$ and τ is regular, we must have $\dot{p}_\tau^r = \dot{p}_\tau^s$ as well. Thus the expression in the last inequality can be rewritten as

$$\frac{y^s(p_\tau^s) - \sum_t x_t^s(p_\tau)}{\epsilon^s} \sum_{r\in M} \left(\sum_t \hat{x}_t^r(p_\tau) - \sum_t x_t^r(p_\tau) \right) \ , \tag{37}$$

for some $r \in M$. But if the term outside the sum is positive, by Lemma 4.3, $\sum_{r \in M} \sum_t \hat{x}_t^r(p_\tau) = \sum_{r \in M} \sum_t x_t^r(p_\tau)$. Hence, (37) equals zero, so the claimed inequality is true. ■

VI. Summary

References

[1] F. P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.
[2] P. Key, L. Massoulie, and M. Vojnovic, "Farsighted users harness network time-diversity," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 4, 2005, pp. 2383–2394.
[3] F. Kelly, A. Maulloo, and Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.
[4] F. Clarke, "Generalized gradients and applications," *Transcactions of the American Mathematical Society*, vol. 205, pp. 247–262, 1975.