

Trade & Cap

A Customer-Managed, Market-Based System for Trading Bandwidth Allowances at a Shared Link

Jorge Londoño[†]
jmlon@cs.bu.edu

Computer Science Department
Boston University, MA, USA

Azer Bestavros[‡]
best@cs.bu.edu

Nikolaos Laoutaris
nikos@tid.es
Telefonica Research
Barcelona, Spain

BUCS-TR-2009-25

Abstract—We propose Trade & Cap, an economics-inspired mechanism that incentivizes users to voluntarily coordinate their consumption of the bandwidth of a shared resource (e.g., a DSLAM link) so as to converge on what *they* perceive to be an equitable allocation, while ensuring efficient resource utilization. Under Trade & Cap, rather than acting as an arbiter, an Internet Service Provider (ISP) acts as an enforcer of what the community of rational users sharing the resource decides is a fair allocation of that resource. Our Trade & Cap mechanism proceeds in two phases. In the first, users engage in a strategic *trading* game in which each user agent selfishly chooses bandwidth slots to reserve in support of primary, *interactive* network usage activities. In the second phase, each user is allowed to acquire additional bandwidth slots in support of presumed open-ended need for *fluid* bandwidth, catering to secondary applications. The acquisition of this fluid bandwidth is subject to the remaining “buying power” of each user and by prevalent “market prices” – both of which are determined by the results of the trading phase and a desirable aggregate *cap* on link utilization. We present analytical results that establish the underpinnings of our Trade & Cap mechanism, including game-theoretic results pertaining to the trading phase, and pricing of fluid bandwidth allocation pertaining to the capping phase. Using real network traces, we present extensive experimental results that demonstrate the benefits of our scheme, which we also show to be practical by highlighting the salient features of an efficient implementation architecture. While our focus in this paper is on the rational coordination of the shared use of a DSLAM link, we also establish the generality of our Trade & Cap mechanism by presenting a number of other direct applications, ranging from coordination of energy-aware task schedules to coordination of ISP uplink bandwidth consumption.

I. INTRODUCTION

Motivation: The ever increasing appetite for Peer-to-Peer (P2P), media streaming, and Video on Demand (VoD) content is forcing service providers to constantly upgrade their infrastructures to keep-up with customers’ bandwidth demands. This state-of-affairs is significantly exacerbated by the prevalence of flat-pricing schemes and hence the lack of an incentive for users to moderate their hunger for network bandwidth,

especially around periods of peak network utilization, which are the primary determinants of an ISP costs (both in terms of infrastructure upgrade cycle and inter-AS traffic volume costs due to the 95/5 rule). Attempts by ISPs to deviate from flat pricing (including field-tested per-byte pricing [1]) have been widely rejected by customers [2]. This is also reinforced by the prevalence of flat pricing in the telephony market [3].

In addition to the significant capital investments that ISPs must shoulder to ensure that their networks are well provisioned during the few hours of peak demand, the new (Internet) world order of seemingly unbounded hunger for bandwidth further complicates fundamental issues that have confounded the networking community for decades, including the adoption of an acceptable notion of fairness as it relates to congestion management. Congestion increases delay and losses, reducing the perceived Quality of Service (QoS) of *interactive* applications such as web browsing, VoIP, and video streaming. Dealing with congestion requires that users (flows) “pay” for their share of the congestion they cause [4], resulting in a degradation in QoS (the congestion price). But, when interactive applications are forced to compete with non-interactive applications such as P2P filesharing, background backup services, or VoD downloads, the degradation in QoS becomes unacceptable.

Under flat pricing, during periods of peak demand, current congestion control practices could be seen as particularly “unfair” to users of low-volume, mostly-interactive applications who would be effectively subsidizing “bandwidth hogs.” This has prompted some ISPs to act as arbiters, proactively shaping user traffic by setting quotas,¹ or by preferentially treating different traffic payloads (e.g., web browsing vs.

[†] Supported in part by the Universidad Pontificia Bolivariana and COLCIENCIAS–Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología “Francisco José de Caldas”.

[‡] Supported in part by NSF awards CCF-0820138, CSR-0720604, EFR-0735974, CNS-0524477, and CNS-0520166.

¹ Incidentally, when demand is well below the provider’s nominal capacity, supporting bandwidth hogs is basically free, bringing to question the use of traffic volume “quotas” [5].

bittorrent downloads) during periods of peak demand.² These efforts have backfired, eliciting a public relation’s quagmire regarding violation of “Net Neutrality,” [17], [18], [19] which is perceived as the prime reason for the Internet being the cradle of innovation it is [20]. Proactive ISP intervention based on traffic payload also raises concerns regarding monopolistic practices, *e.g.*, blocking or taxing Video/VoIP services not provided by the same ISP [20].

Scope and Contributions: Rather than having ISPs act as arbiters who set the rules regarding what constitutes fair usage of a shared resource [21], in this paper, we propose a *voluntary*, market-based *Trade & Cap* system in which users converge on what *they* perceive as an equitable allocation of resources, irrespective of what these resources are used to support (HTTP vs P2P traffic) and irrespective of the absolute resource allocation (traffic volume) per user.³ In our setting, the role of an ISP becomes that of a “provider” and an “enforcer” of what the *community* of users decides is a fair allocation.

Effectively, our proposed Trade & Cap (T&C) mechanism sets up a marketplace. Given the fixed (flat-rate) payment to the provider, customers enter this marketplace with equal buying power, but their use of this fairly-allocated buying power depends on their flexibility. This allows customers to trade “volume” during low-utilization periods for “quality” during peak-utilization periods (or vice versa). The direction of the trade (not to mention the user’s willingness to even engage in trading) depends entirely on user preferences and flexibility (*e.g.*, tolerance for delaying a scheduled network backup job).⁴ In addition to empowering customers to trade bandwidth allocations, T&C has the desirable side effect of smoothing traffic utilization over time, thus reducing the ISP’s cost which is determined primarily by the 95/5 rule.

Outline and Summary of Results: We start this paper in Section II by overviewing the T&C mechanism as it applies to a Digital Subscriber Line Access Multiplexer (DSLAM) setting, and in Sections III and IV by presenting analytical results pertaining to convergence and efficiency of the marketplace underlying T&C. Formulating the problem as a game is not only useful for purposes of modeling and understanding the marketplace dynamics, but also it serves as the basis of a real mechanism that can be implemented and applied in practice. Thus, in Section V we discuss the salient features of an implementation architecture for T&C in a DSLAM setting. Our implementation allows the marketplace interactions to

² Along these lines, there is a growing body of academic [6], [7], [8], [9], [10], [11] and industry [12], [13], [14], [15], [16] work on delineating interactive from non-interactive traffic in order to police/balance consumption. Many of these systems depend on Deep Packet Inspection (DPI) techniques, raising concerns about consumer privacy. Moreover, the scalability and resilience of these techniques is also questionable as applications adapt quickly to avoid detection, *e.g.* by using encryption and randomization of port numbers.

³ We note that recent polls [22] indicate that consumers would accept traffic allocation mechanisms that ensure fairness as long as these mechanisms do not trample on net neutrality, privacy, *etc.*

⁴ We use the term “user” liberally since in practice, customer-side software agents would make most decisions on behalf of the user.

be carried out by software agents that run on behalf of the users and the ISP, and thus (with the exception of minimal configuration and parametrization) is quite transparent to the user. Next, in Section VI, we demonstrate the significant advantages of T&C by presenting results from extensive trace-driven simulations. For instance, we show that introducing a relatively small level of flexibility in the scheduling of user activities results in significant gains for both the users and the ISP. For example, allowing user agents to reposition bandwidth allocations within relatively small windows of time enables them to increase their share of fluid bandwidth (supporting non-interactive applications) by 20% to 40% depending on their flexibility. This benefits the ISP as well, resulting in as much as 16% to 31% reduction in the 95th percentile of the ISP’s 5-minute traffic volume, and (even more impressively) resulting in smoothing traffic volume, reducing the 95th-percentile/50th-percentile ratio from 1.58 to an almost perfect ratio of 1.004. While our focus in this paper is on the rational coordination of the shared use of a DSLAM link, we also establish the generality of our Trade & Cap mechanism by presenting in Section VII a number of other direct applications, ranging from coordination of energy-aware task schedules to coordination of ISP uplink bandwidth consumption. We conclude the paper in Section VIII with a review of the related literature.

II. TRADE & CAP IN A DSLAM SETTING

While our T&C mechanism is applicable to any setting in which it is desirable to coordinate the fractional acquisition by a set of rational parties of the *shared* capacity of a single resource, in this paper, and without loss of generality, we restrict ourselves to a specific setting – that of coordinating the utilization of a shared DSLAM link.

Figure 1 illustrates the basic architecture of Digital Subscriber Line (DSL) access technology. In this setting, DSL modems on the customer side connect hundreds to thousands of users to a single DSLAM server on the provider network. DSLAMs connect to a Broadband Remote Access Server (BRAS) which relays traffic to/from the Internet. In this setting, the DSLAM-BRAS link poses the most significant traffic management problems for ISPs and is thus the shared resource managed using our T&C mechanism.⁵

As we alluded before, we envision a marketplace where DSL customers are empowered to trade capacity over time, so as to facilitate the exchange of traffic volume for QoS. This exchange is desirable given the different utility that various applications attribute to traffic volume vs. QoS (*e.g.*, Fluid-Traffic (FT) applications value traffic volume whereas Interactive-Traffic (IT) applications value QoS).⁶ In the envisioned marketplace, the DSLAM server’s role is to enforce the

⁵ T&C is equally valuable and practical if the resource to be managed is not “physical” but rather “virtual” – *e.g.*, the aggregate inter-ISP (transit) traffic of a subnetwork. Our distributed implementation architecture discussed in Section V is particularly suited for managing such resources.

⁶ Our T&C mechanism ensures that resource allocations are based on true valuations by the users themselves (rather than assumed by the ISP).

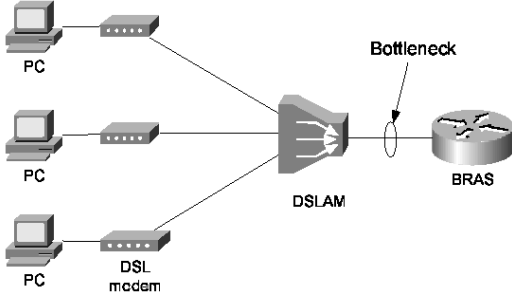


Fig. 1. Illustration of the DSL “last-mile” architecture.

capacity allocations agreed upon by the DSL customers. By doing so the ISP will benefit as well, as the T&C marketplace dynamics result in a more balanced load over time, improving user satisfaction and alleviating the need for infrastructure upgrades to accommodate peak demand.

For our purposes, we assume that the marketplace will operate over fixed, non-overlapping periods of time, which we call *epochs* (e.g., days), and that the trading and allocation of capacity will occur within T subdivisions of an epoch, which we call *time-slots* (e.g., 288 5-minute slots per day).

At the beginning of each epoch, the operator assigns each user (agent) $i = 1, 2, \dots, n$ an allowance or *budget* B_i in accordance with the user’s Service Level Agreement (SLA) (e.g., “Business” versus “Residential” plans). Under flat pricing, which we assume in this paper, all customers receive an equal budget. Our T&C mechanism proceeds in two phases:

(1) *The Bandwidth Trading Phase*: This phase proceeds as a pure-strategies, non-cooperative game among agents, who are allowed to rationally and selfishly decide *when* to schedule bandwidth allocations in support of their IT *sessions*. An IT session is a consecutive set of time-slots during which a particular IT application is active. For example, a user may have a browsing and an e-mail IT session from 7-8pm and another video-streaming IT session from 10-11pm. Although not necessary for analytical purposes, in our experiments as well as in our implementation architecture, we enforce the practical requirement that sessions be atomic – *i.e.*, a video-streaming session cannot be broken up or interrupted. The scheduling of IT sessions is subject to preset user preferences and constraints. The outcome of this game is a Nash-Equilibrium (NE) of IT bandwidth allocations to all participating agents, along with the corresponding cost incurred by each agent.

(2) *The Bandwidth Capping Phase*: This phase proceeds as a market-clearing phase, in which the operator distributes any remaining capacity among agents. The amount of “remaining” capacity distributed in this phase is set based on a desirable nominal utilization of the link (e.g., determined by the 95/5 rule threshold). The allocation of bandwidth in the capping phase rewards users who were able to preserve more of their budgets in the trading phase (due to a low IT volume or due to flexibility in scheduling such traffic), ensuring a market equilibrium of the resulting allocations.

III. THE BANDWIDTH TRADING PHASE

Each agent i represents its IT demand as a vector of requested bandwidth allocations: $T_i = (t_{i1}, \dots, t_{iT})$. An assignment of an agent’s demand is a mapping that pins each one of the components of the vector to a different time slot. A set of such assignments (one per agent) comprises a potential configuration, or *schedule* of IT utilization at the DSLAM.

Let $k = m_i(j)$ be the time slot assigned to the j^{th} component of player i ’s request vector. We denote by x_{ik} the actual allocation for player i in time-slot k , where $x_{ik} = t_{i,m_i(j)}$. The x_{ik} notation implicitly represents the mapping $m_i(\cdot)$, noting that for time-slots that are not used in the mapping, we assume that $x_{ik} = 0$. Thus, x_{ik} is defined for all time-slots.

Definition 1. (*Cost Function*) The cost for a player i is

$$c_i = \frac{1}{C} \sum_{p=1}^T x_{ip} U_p \quad (1)$$

where $U_p = \sum_{i=1}^n x_{ip}$ is the utilization of slot p and C is a constant.

The above cost function (which is proportional to the product of the current utilization and the demand of the player over all time slots) can be interpreted as a *cost-sharing* scheme where each user pays its fair share of the price of each time slot, which depends on the *square* of the time-slot’s utilization

$$c_i = \frac{1}{C} \sum_{p=1}^T U_p^2 \left(\frac{x_{ip}}{U_p} \right)$$

Our choice of the above cost function captures a notion of fairness and incentivizes users to shift (if possible) their traffic to lower-utilization time slots.⁷ Non-linear cost functions (of which ours is an instance) have been used before [23] to control congestion and achieve “proportional fairness.”

The strategy space S_i^* for agent i is the set of permutations of its request vector. As such, the strategy space is finite with cardinality $|S_i^*| = P_i^T$. The game’s strategy space is the Cartesian product of the strategy spaces of all players: $S = \times S_i$. Initially, we will assume that all the points in the strategy space are feasible, and we will incorporate capacity/budget constraints later.

Theorem 1. *The pure strategies game in which users adopt better/best responses to allocate atomic units of traffic in per-user, mutually-exclusive time-slots converges to a NE.*

Proof: We define the following potential function:

$$\Phi = \sum_{i=1}^n c_i = \frac{1}{C} \sum_{p=1}^T U_p^2$$

⁷ While instrumental in establishing the convergence property given in Theorem 1, the specification of a quadratic (square) form in our cost function is not essential as other cost functions may well yield the same desirable incentives.

When a player makes a cost-reducing move, $\Delta c_i < 0$,

$$\frac{1}{C} \sum_p (x'_{ip} U'_p - x_{ip} U_p) < 0 \quad (2)$$

Notice that for any other player $k \neq i$, its utilization of interval p does not change, but the change in the total utilization affects its cost as follows

$$\Delta c_k = \frac{1}{C} \sum_p x_{kp} (U'_p - U_p)$$

Adding the changes of the players other than i we get

$$\begin{aligned} \sum_{k \neq i} \Delta c_k &= \sum_{k \neq i} \left(\frac{1}{C} \sum_p x_{kp} (U'_p - U_p) \right) \\ &= \frac{1}{C} \sum_p \left((U'_p - U_p) \sum_{k \neq i} x_{kp} \right) \\ &= \frac{1}{C} \sum_p \left((x'_{ip} - x_{ip}) \sum_{k \neq i} x_{kp} \right) \end{aligned} \quad (3)$$

where in the last step we used the fact that $U'_p - U_p = x'_{ip} - x_{ip}$ because players other than p did not change their allocations. Since the components of x'_{ip} are the same as those of x_{ip} (but in different positions), we observe that $\sum_p x'_{ip}{}^2 = \sum_p x_{ip}{}^2$. With this, we can reorganize expression (2) as follows

$$\frac{1}{C} \sum_p (x'_{ip} U'_p - x_{ip} U_p) = \frac{1}{C} \sum_p \left((x'_{ip} - x_{ip}) \sum_{k \neq i} x_{kp} \right) < 0$$

which is exactly the same as (3), *i.e.* $\sum_{k \neq i} \Delta c_k = \Delta c_i < 0$. As the sum of negative quantities is negative, we get

$$\sum_i \Delta c_i = \Delta \Phi < 0$$

i.e. the potential is monotonically decreasing and is lower-bounded by some constant greater than zero. This lets us conclude that the game converges to a Nash Equilibrium. ■

As we alluded before, it may be the case that an agent may have additional constraints that limit its strategy space – *e.g.*, a 2-hour-long IT fixed bandwidth allocation must be assigned in consecutive time-slots, and be scheduled to start between 6pm and 8pm. Such constraints are easily captured by defining the player's strategy space as a subset of $S_i \subseteq S_i^*$.

In practice two type of constraints may be important to consider in the game as described so far: (1) *Capacity constraints* to ensure that the shared link capacity is never exceeded by the aggregate allocation – $\forall p : \sum_{i=1}^n x_{ip} \leq C$, and (2) *Budget constraints* to ensure that no agent is able to reserve resources beyond his “fair” share, which is upper-bounded by the agent's allowance – $\forall i : \frac{1}{C} \sum_{p=1}^T x_{ip} U_p \leq B_i$.

Notice that both sets of constraints correspond to the elimination of infeasible points in the strategy space \mathcal{S} . This removal can be easily accomplished by setting to ∞ the cost per player for those points.

Theorem 2. (Convergence to NE under constraints) *Given a pure strategies game, such that each player's action space is finite, and that converges under better/best response dynamics to a NE, then after adding constraints to the action space of one or more players, the game still converges, given that there exists feasible configurations after the addition of the constraints.*

Proof: Consider the following directed graph $G = \langle V, E \rangle$: There is a vertex $v_j \in V$ for every possible point in the strategy space $v_j = (a_{1j_1}, \dots, a_{nj_n})$, where a_{ij} denotes the j^{th} action of player i . There is an edge $e_{pq} \in E$ for any valid transition⁸ on the strategy space, *i.e.* the cost associated with player i at vertex p is larger than the cost at vertex q : $c_p(i) > c_q(i)$ and $a_{-i,p} = a_{-i,q}$, meaning that the actions of all players other than i are the same in p and q . Let us call G the transition graph of the game. Then, if the game always converges to a NE in the unconstrained case, it follows that G must be a Directed Acyclic Graph (DAG). Any path (sequence of actions) the players traverse when following their rational, selfish goal will always reach a vertex with no outgoing edges corresponding to a NE (of possibly many) of the game. The addition of constraints to the players actions corresponds to *removing* unfeasible vertices from V as well as the edges coming into or out of these vertices. Let G' be the residual transition graph after removing unfeasible vertices and edges. Suppose the new game with constraints does not always converge to a NE. Then, there must exist at least one cycle in this residual transition graph G' . The fact that G' is a subgraph of G implies that the same cycle must exist in the original graph G , contradicting the fact that G is a DAG. ■

Figure 2 illustrates the construction used in the proof of the above Theorem. Figure 2 (a) shows the DAG corresponding to the transitions of some hypothetical game, where states v_6 and v_8 are the NE. In Figure 2 (b), vertices v_4 and v_6 have been removed with their respective edges because they are unfeasible. The NE in the residual graph are v_3 and v_8 . Notice that the set of NE vertices after the addition of constraints need not to be the same as those of the unconstrained game. In particular, feasible vertices that were not a NE will become a NE if all their outgoing edges are removed.

An important consideration when considering equilibria of non-cooperative games is the Price of Anarchy (PoA) – the ratio of the social cost at the worst-case equilibrium compared to the best possible. In the case of the Bandwidth Trading game, the social cost (understood in our case to be the system metric we want to optimize) is the maximum slot utilization.

Theorem 3. (Price of Anarchy for Bandwidth Trading) *When user sessions are described as finite sequences of fixed size allocations, the PoA on the per-slot load is n .*

Proof: A loose bound on the PoA for the trading game is trivial: Given a maximum allocation per player, X_{max} , it

⁸ Observe that the set of edges is not limited to best-responses, but includes any feasible move.

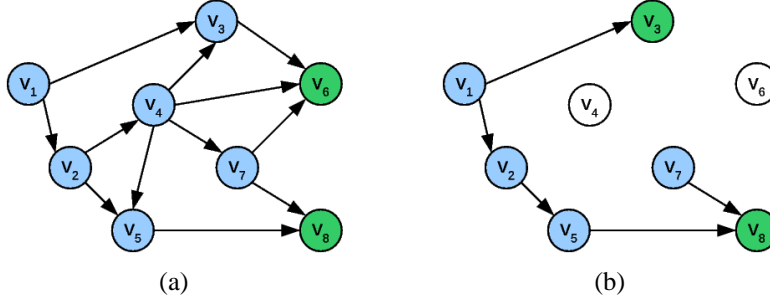


Fig. 2. Illustration of transition graphs for a pure strategies game. a) Without constraints, b) With constraints

may be the case that all the n players have an equally-large demand, and there exists a NE where these demands coincide in the same time slot. On the other hand, there is always going to be a slot with utilization of at least X_{max} , therefore this is a lower-bound on the slot utilization. Therefore we have the bounds

$$X_{max} \leq \max\{U_p\} \leq nX_{max}$$

These loose bounds immediately imply that

$$PoA \leq \frac{\text{worst-case } \max\{U_p\}}{\text{optimal } \max\{U_p\}} = n$$

To show that this bound is tight, we present in Fig. 3 an instance that realizes it. In this example there are n players, each one having a session of length $n + kn$, and the total number of time-slots is $n + kn + n - 1 = n(k + 2) - 1$. Fig. 3a shows the optimal allocation which yields an $\max\{U_p\} = X_{max}$, and part (b) shows a NE whose $\max\{U_p\} = nX_{max}$. Part (b) is a NE because any unilateral deviation by any player, gives a higher cost. In fact the player cost at NE is

$$c_i = \frac{1}{C} \sum_p x_{ip} U_p = nX_{max}^2 + k$$

And the cost for a player if he moves any integral number of positions (within the allowed time-slots) is

$$c'_i = \frac{1}{C} \sum_p x_{ip} U_p = X_{max}^2 + 2k$$

and $c'_i > c_i$ whenever $k \geq (n - 1)X_{max}^2$.

It is important to notice that realizing the above PoA bound requires a carefully crafted problem instance. In practice it is very unlikely to find instances with these characteristics. In fact, to evaluate the practical behavior of the PoA we conducted a series of simulations as described in the following procedure:

- 1) Create a problem instance whose optimal allocation is known. The load-balancing problem itself is NP-Complete⁹. On the other hand, constructing an instance

⁹ It is easy to see this by reduction to the 2-PARTITION[24] problem. If we had a polynomial algorithm that solves the load-balancing problem, we could run this algorithm on an instance of the partition problem with two slots. If the sum of the elements in the two slots is equal, the answer to the PARTITION problem is “yes”, otherwise is “no”

with a known optimal solution is simple: Take the slots, assume they are all equally filled say with 1 unit. Split the content of each slot in several fractions and then take sequences of elements from different slots to be the tasks of the players. Finally, shuffle around the tasks of the players to get a problem instance.

- 2) For different numbers of players (this defines the game size) and of time-slots we create multiple problem instances. In the results presented in this paper, we created 100 instances for each game size.
- 3) Run the game by letting the players take turns and play their best responses until the game reaches a NE. Take the maximum among all the instances of the same size, and then compute the ratio with respect to the known optimum. This gives the empirical ratio of the worst-case to the optimal.

The results of these simulations are illustrated in Figure 4, with 5 slots (a) and 10 slots (b). In practice, the PoA for the trading phase (game) is almost always below 2, and tends to be insignificant as the number of players (size of the game) increases, which bodes well for our setting.

IV. THE BANDWIDTH CAPPING PHASE

Once the trading phase is concluded (resulting in an assignment of IT to time slots), an agent i is allowed to use whatever is left over from its originally allocated budget B_i to acquire FT capacity. We denote by ℓ_i the left-over budget for user i .

$$\ell_i = B_i - c_i$$

As its namesake suggests, we will assume that FT traffic can be handled as a fluid, so the allocation of user i to slot p is $w_{ip} \in \mathbb{R}^+$.

During the capping phase, each agent computes the allocation of FT traffic that maximizes its total volume, subject to the constraint on the left-over budget. For this computation, agents use the same cost function given in equation (1), but now accounting for the overall utilization from both the trading and capping phases.

For notational clarity, we will use the bold symbol \mathbf{U}_p to denote the total utilization of slot p . The normal symbol U_p will continue to denote the utilization due only to IT allocation,

is closed and bounded. Let $Df(x)$ of a function $f : \mathbb{R}^p \rightarrow \mathbb{R}^n$ be the matrix¹⁰ whose $(i, j)^{th}$ element is

$$\frac{\partial f_i}{\partial x_j}(x), \quad i = 1, \dots, n, \quad j = 1, \dots, p$$

The function $g()$ is strictly convex as its Hessian matrix $D^2g(w_i) = 2I$ is positive definite. This is the case because for any $z \in \mathcal{D}_i, z \neq \mathbf{0}$,

$$z^T(2I)z = 2 \sum_p z_p^2 > 0$$

With this, we can show that \mathcal{D}_i is also a convex set. Take $x, y \in \mathcal{D}_i$, and define

$$z = \alpha x + (1 - \alpha)y$$

then as $g()$ is strictly convex, $g(z) < \alpha g(x) + (1 - \alpha)g(y) < 0$. Also, because $x, y \in \mathcal{D}_i$ it is the case that $\forall p : h_p(x) \geq 0$ and $\forall p : h_p(y) \geq 0$. Therefore $h_p(z) = \alpha h_p(x) + (1 - \alpha)h_p(y) \geq 0$. In conclusion $z \in \mathcal{D}_i$, hence \mathcal{D}_i is convex.

Before proceeding, let us define some notation (following [25]). Let

$$\varphi_i = \begin{cases} g(w_i), & \text{if } i = 1 \\ h_{i-1}, & \text{if } i \in \{2, \dots, T+1\} \end{cases}$$

as the set of functions encapsulating all the constraints. Also define $\rho(A)$ to be the rank of matrix A .

A constraint is called an *effective constraint* if it is satisfied with equality, so the set $E \subset \{1, \dots, T+1\}$ identifies the effective constraints, i.e. at the maximum w_i^* , for $k \in E$, $\varphi_k(w_i^*) = 0$. Let $\varphi_E = (\varphi_k)_{k \in E}$ be the matrix of effective constraints. It is easy to show that $\rho(D\varphi_E(w_i^*)) = |E|$. To see this, observe that $Dh_p(w_i) = 1$ in column p and zero otherwise. Therefore all the rows corresponding to the functions h_p are linearly independent. On the other hand, the row

$$Dg_i(w_i) = (\dots, \mathbf{U}_p^{-i} + 2w_{ip}, \dots)$$

is non-zero on those columns where $h_p(w_i) = 0$ (if $\mathbf{U}_p^{-i} = 0$, the cost of the slot is minimal and $w_{ip} \neq 0$). Therefore the row $Dg_i(w_i)$ is not linearly dependent on the rows $Dh_p(w_i)$, so we conclude that the rank of $D\varphi_E(w_i^*)$ is $|E|$.

As a result of this analysis, the problem meets all the conditions of Theorem 6.10 in [25] and therefore there exists a vector of Lagrange multipliers $(\lambda_i, \mu_1, \dots, \mu_p)$ such that

- 1) $\mu_p \geq 0$ and $\mu_p h_p(w_i^*) = 0$ for $p = 1 \dots T$.
 - 2) $DL(w_i^*, \lambda_i, \mu) = 0$
- where

$$L(w_i, \lambda_i, \mu) = f(w_i) + \lambda_i g(w_i) + \sum_{p=1}^T \mu_p h_p(w_i)$$

is called the Lagrangean.

Condition 1) states that either μ_p , or w_{ip}^* is zero. For those slots where $w_{ip}^* \neq 0$, condition 2) can be expanded as:

$$\frac{\partial L}{\partial w_{ip}} = 0 \quad (7)$$

$$\frac{\partial L}{\partial \lambda_i} = g(w_i) = 0 \quad (8)$$

Then, solving for w_{ip} from (7), we get:

$$\frac{\partial L}{\partial w_{ip}} = 1 + \lambda_i \mathbf{U}_p^{-i} + 2\lambda_i w_{ip} + \mu_p = 0$$

$$w_{ip} = -\frac{1 + \mu_p + \lambda_i \mathbf{U}_p^{-i}}{2\lambda_i} \quad (9)$$

and substituting from equation (9) into equation (8) and removing the slots where $\mu_p = 0$ gives us the following equation:

$$\frac{1}{4\lambda_i^2} \sum_{p|w_{ip}>0} \left(1 - (\lambda_i \mathbf{U}_p^{-i})^2\right) = \ell_i C$$

Solving for λ_i , we get:

$$\lambda_i = \sqrt{\frac{\sum_{p|w_{ip}>0} 1}{4\ell_i C + \sum_{p|w_{ip}>0} (\mathbf{U}_p^{-i})^2}} \quad (10)$$

The summations in equation (10) require the identification of the set $\{p|w_{ip} > 0\}$ denoting the slots with not null allocation. Observing that the slots where $w_{ip} = 0$ are infeasible slots (they are very expensive), a two-step recursive process that allows us to identify the slots to use is to: (1) assume that all slots are feasible, solve for λ_i and w_{ip} , and the resulting violations define the forbidden slots $\mathcal{F} = \{p|w_{ip} \leq 0\}$, and (2) solve again with the remaining slots $\bar{\mathcal{F}}$. This corresponds to redistributing the budget that could not be saved in the \mathcal{F} slots (by setting $w_{ip} < 0$) between feasible slots.

Substituting for λ_i from equation (10) into equation (9) yields the desired allocations of FT bandwidth in each one of the slots $\{p|w_{ip} > 0\}$.

Lemma 1. (Global optimality) w_i^* found by the previous procedure is a global optimum

Proof: Immediate from the fact that \mathcal{D}_i is convex and $f()$ is concave. (See Theorem 7.13 in [25]). ■

Lemma 2. (Uniqueness of the global optimum) w_i^* is unique

Proof: Suppose it is not. Let w_i^1 and w_i^2 be two distinct global optima. Let us also define $z = \alpha w_i^1 + (1 - \alpha)w_i^2$ for $\alpha \in (0, 1)$. By the convexity of \mathcal{D}_i , it is always the case that $z \in \mathcal{D}_i$. By the linearity of f we have

$$f(z) = \alpha f(w_i^1) + (1 - \alpha)f(w_i^2) = f(w_i^2)$$

i.e. all points z in the hyperline segment defined by w_i^1, w_i^2 are also global optima. The constraint $g(w_i)$ is strictly convex, therefore

$$g(z) < \alpha g(w_i^1) + (1 - \alpha)g(w_i^2)$$

¹⁰ This is the Jacobian of f .

$g(z)$ is the excess over the budget. At the optimal solutions $g(w_i^{\{1,2\}}) = 0$, which means that at point z there is still some unused budget. This means that the user can increase its allocations, therefore contradicting the optimality of z . ■

Theorem 4. (Existence and Uniqueness of Optimal FT Bandwidth Allocation) *There is a single maximum for the aggregated FT traffic and this maximum coincides with the per-user maximum.*

Proof: (Sketch) Consider the related problem of maximizing the total allocation of FT traffic. Namely, maximize

$$f(w) = \sum_{i=1}^n \sum_{p=1}^T w_{ip}$$

subject to

$$\begin{aligned} \frac{1}{C} \sum_p w_{ip} U_p &= \ell_i \quad \forall i \\ w_{ip} &\geq 0 \quad \forall i, p \end{aligned}$$

The corresponding Lagrangean is

$$L(w, \lambda, \mu) = f(w) + \sum_i \lambda_i g_i(w) + \sum_i \sum_p \mu_{ip} h_{ip}(w)$$

with

$$\begin{aligned} g_i(w) &= g(w_i) = \sum_p w_{ip} U_p - \ell_i C \\ h_{ip}(w) &= h_p(w_i) = w_{ip} \end{aligned}$$

The feasible space in this case is

$$\mathcal{D} = \{w \in \mathbb{R}^{nT} \mid \forall i : g_i(w) \leq 0 \wedge \forall i, p : h_{ip}(w) \geq 0\}$$

Observe also that

$$\mathcal{D} = \cap_{i=1}^n \mathcal{D}_i$$

and therefore it is convex (See Theorem 1.37 in [25]).

We note that the above is a generalization of the single user (distributed) optimization, including the Lagrange multipliers of all the users. In particular, the Karush-Kuhn-Tucker (KKT) conditions are a superset of the corresponding ones for the single user case and their solutions are the same. Therefore, the unique solution of the aggregated problem corresponds to the set of solutions of the (distributed) per-user problems. ■

An important consequence of this theorem is that the maximum is a NE as no user has an alternative solution with larger allocation of FT traffic having the same budget.

V. IMPLEMENTATION OF A T&C DSLAM MARKETPLACE

Architecture: We envision an implementation of a T&C-based system consisting two software agents: a provider-side agent and a one-per-customer agent. The provider-side agent is responsible for the marketplace functionality, accepting bids from the customer agents and reporting back the results. It can also provide a policing function, to ensure that traffic from each agent adheres to the allocations agreed upon in the marketplace, thus avoiding the possibility of customers tampering with their agents in an attempt to hack the system.

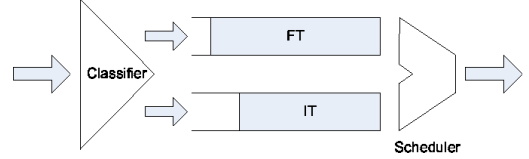


Fig. 5. Implementation using priority queues

The customer-side agent would be responsible for monitoring network traffic usage over time, and collecting feedback from the user regarding the user's experience. Using such data, the agent could easily infer, and accordingly set, trading preferences (e.g., flexibility of IT traffic).

The entire system operates on the local domain defined by the DSLAM and the finite (customer) population attached to it. For accounting and policing purposes, the system would need to uniquely identify each customer. Authentication – in many cases already in place at the physical or link layers, depending on the underlying technology (e.g., xDSL) – is needed to protect against “identity theft” whereby a customer would spoof the MAC address of another in the same DSLAM to avoid having its traffic counted against its own budget. Notice that to account for traffic during each epoch, the provider agent only needs the total allocation per customer. This information is enough to ensure that the customer is adhering to the outcomes of the T&C mechanism for each time slot. From the providers perspective it is irrelevant if the customer is using a bandwidth allotment for IT or FT bandwidth. In fact, this assures that the provider's policing mechanism is indeed *neutral* with regard to the customer's traffic.

Handling traffic on the customer side requires the implementation of a two-level priority queuing system, with the high priority assigned to IT demand and the low priority assigned to FT demand. This way, packets belonging to IT applications preempt any pending packets in the FT queue. In addition, the queues have a traffic shaper (e.g. token-bucket) to ensure that (for example) FT packets do not consume the bandwidth allocation during idle-periods of IT queue. The routing of packets to each one of these queues could be implemented in a number of ways, including using manual configuration on a per application basis, using an automatic traffic classifier ([6], [7], [9], [11]), or using special APIs that allow applications to bind to specific virtual interfaces.

Priority/weighted queueing systems have long been used in the QoS literature. An implicit assumption in that literature is that priorities/weights are assigned consistently by the end systems. However, when self-interested players compete for the same resource, their choice would be to assign themselves the highest priority, unless there is a cost associated with this choice. Our T&C mechanism incorporates such a cost, thus providing the needed incentive for players to act truthfully.

Algorithmic Complexity and Efficient Distributed Implementation: A scheme like ours would not be practical if associated processes are not efficient to compute.

The best-response computation in the trading phase is a

combinatorial problem, equivalent to a generalized knapsack problem. For that, we developed a dynamic programming solution which is pseudo-polynomial (complexity depends on the product of the number of sessions per user and the number of time slots) and which runs in a few seconds on current hardware for instances of practical sizes of hundreds of users and hundreds of time slots (108 and 288, respectively in our simulations). The dynamic programming solution, when finding the best response for user i proceeds as follows:

- 1) Let k be number of sessions of i , and T the number of time-slots
- 2) Initialize the matrix A of dimension $k \times T$. Each element a_{jp} of A will represent the cumulative cost of sessions $1 \dots j \leq k$, when the j^{th} session is allocated in slot p . All the matrix elements are initialized to infinity.
- 3) The first row is computed by assuming session 1 is placed in slot p and computing the resulting cost.
- 4) Subsequent rows ($j = 2 \dots k$) are computed according to eqn. (11). Here, $c(j, p)$ represents the cost of session j at slot p (from eq. 1). Observe also that a_{jp} is the minimum cost at which all the sessions up to j can be allocated in the time-slots up to p . Therefore, the minimum of the last row $\min\{a_{k,1..T}\}$ will give the optimal cost for the entire set of sessions of the user.

The feasibility condition in eqn. (11) refers to the constraints of the problem. Basically, different sessions do not overlap, all the components of the session fall within the allowed time-slots ($1 \dots T$), and the cumulative cost is less or equal to the budget. In particular, in the case of the user going over the budget, we adopted the policy of dropping arbitrary sessions until the budget constraint is satisfied.

In our experiments we did not implement the capacity constraint, although it could be easily incorporated into the procedure. In doing so, we allow for the utilization to grow as much as demanded, which gives even more conservative estimates of the worst-case performance metrics.

As for the fluid allocation computation in the capping phase, the solution using Lagrange multipliers presented in Section IV constitutes a straightforward distributed implementation, whereby at the Customer Premises Equipment (CPE) each agent computes its best response iteratively until it gets close enough to the global optimum.

Running both the trading and capping processes at the CPE is consistent with a network-neutral implementation. The only support needed from the DSLAM would be to offer a blackboard service where all the participants are able to register their (IT and FT) allocations and query the totals (U_p) per time-slot. Once the market reaches an equilibrium, the posted schedule is committed for the next epoch.

VI. EXPERIMENTAL EVALUATION

In this section we use trace-driven simulations to (1) highlight the benefits that a user in our system begets by exhibiting some flexibility in scheduling its IT sessions under T&C, (2)

Period	2009-03-31 00:00 – 2009-03-31 23:59
Total packets	1,551,089,845
TCP packets	1,194,409,653
UDP packets	4,321,852
Total TCP bytes (payload)	924,540,189,060

TABLE I
CHARACTERISTICS OF THE WAN TRACE USED IN OUR EVALUATION.

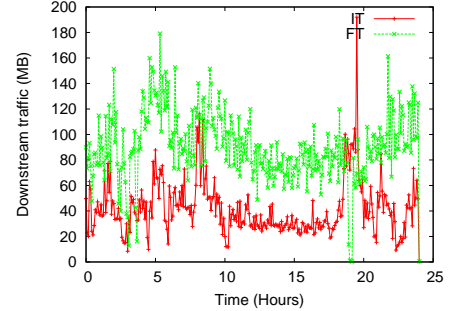


Fig. 6. Downstream trace for a subnet of broadband users

demonstrate the gains that an ISP stands to realize as a result of the overall smoother traffic profile of T&C, and (3) illustrate how various parameters affect the performance of T&C.

Traces and Trace Pre-Processing: As an alternative to direct DSLAM traces (which unfortunately are not available), we used publicly available WAN traces [26] to extract a slice of traffic associated with a customer access network. Table I shows the main characteristics of these WAN traces. Capturing a slice (portion) of the customer network’s traffic results in less pronounced diurnal peak-to-valley ratios, which limits the performance gains realized by T&C. Thus, the performance gains reported in this section should be viewed as “conservative”. Figure 6 shows the traffic aggregated over $5min$ time-slots for the subnetwork we selected for our evaluation.

To extract traffic associated with a customer access network, we applied the following pre-processing steps. First, we identified subnets most likely associated with broadband users, based on the upstream/downstream ratios, the activity per port number, and diurnal activity patterns. Next, assuming that each IP address is a single user/household, we classified the traffic per user as either IT or FT. This was done based on association of traffic activity with privileged port numbers. Finally, we identified the various IT sessions per user, with their corresponding demands per time-slot. Session identification was done by setting a threshold on the length of periods of high activity. We call this threshold S_{max} and it is given as a number of time-slots. For most of our experiments we considered the values $S_{max} = 6$ and $S_{max} = 12$ corresponding to half an hour and one hour respectively. If any sequence of time-slots has length greater than S_{max} , then we subtracted the minimum from this interval under the assumption that it was due FT. By repeating this process on any subinterval of length greater than S_{max} we obtained a set of disjoint IT sessions for the user.

T&C operates by letting user agents express their flexibility or willingness to move IT components (forward or backward

$$a_{jp} = \begin{cases} \infty & \text{if session } j \text{ is unfeasible at slot } p \\ \min\{a_{j-1,1\dots p-1}\} + c(j,p) & \text{otherwise} \end{cases} \quad (11)$$

in time) some number of time slots. We define a session’s *slack* to be the number of time slots that an agent is willing to shift its session (back or forth in time). A slack of 0 implies no flexibility. A slack of 1 implies a willingness to shift sessions by 5 minutes (our time slot) back or forth, if such a shift is advantageous. Notice that *moving a session* means a shift of the traffic attributed to that session for *all* time slots spanned by that session (*i.e.*, traffic in all time slots of a single session is shifted equally to preserve session atomicity). In our simulation we also enforced the condition that no shifting sessions could overlap. This is consistent with users not doing more activities on the same time-slot. Similarly, we also enforced the condition of preserving the session ordering. Although not required by our model, it implies less effort on the part of the user, and any results thus obtained are even more conservative.

How Does T&C Impact the ISP’s Bottom Line? Our first experiment aims to evaluate how the 95th percentile of the ISP’s 5-minute traffic volume (the 95% traffic envelop) changes as a result of letting users schedule their IT sessions according to the trading phase of T&C. For brevity, we assume that all users adopt the same *slack* value for all their sessions. Figure 7 shows two examples of the outcome after the market reaches an equilibrium. On the left is the traffic per time-slot, and on the right is the CDF of traffic per time-slot. Top row is for session length threshold of $S_{max} = 6$, and the bottom row is for $S_{max} = 12$ time-slots. Clearly, the session thresholding process has little effect on the trace, being the most noticeable effect the larger peak (from 130MB to 150MB). Table II shows the values of the 95% traffic envelop. These results underscore that selfishly scheduling IT sessions yields an equilibrium with *significant* reduction in the 95% traffic envelop – up to 31% reduction when slack is 1 hours. Even for a small slack of 15 minutes, the savings amount to 16%. It is important to note that these savings are likely to be *much more impressive* for real DSLAM traces which exhibit diurnal peak-to-valley ratios that are *much* larger than those evident in our WAN-based traces [27], [28].

Slack	$S_{max} = 6$		$S_{max} = 12$	
	95%(MB)	Savings%	95%(MB)	Savings%
0	36.3	0.0	47.7	0.0
3	30.6	15.6	42.1	11.7
6	27.4	24.4	33.6	29.6
12	24.9	31.4	30.9	35.2

TABLE II
95% UTILIZATION RESULTING FROM BANDWIDTH TRADING.

We now consider experiments in which both phases of T&C are carried out. In particular, after completing the trading phase – thus scheduling all IT sessions in the trace – users allocate as much fluid traffic as possible in accordance with their remaining budgets. Thus, an important consideration in

setting-up these experiments is the budget assignment. In particular, we used the following policy: Let V denote the nominal traffic per time-slot that results in a total volume equal to the total traffic originally in the trace. We introduce a control parameter R (for resistance) which allows the provider to adjust the resulting traffic on the shared link. By setting $C = V/R$ (this is the C of the cost function in equation 1), and the budget per customer to $B_i = CT/n$, the expected utilization (without IT) is precisely C . In our traces (as observed generally on the Internet) the FT component is much larger than the IT component, therefore the IT stage is rarely affected by the budget constraint.¹¹

Figure 8 shows the outcome of the two phases of T&C for a value of $R = 1.0$ and various slack values. The y-axis is normalized with respect to V (the nominal volume under perfectly balanced conditions, with no IT components). Due to the presence of IT components, this quantity is always (slightly) larger than 1.0. The session identification process also capture a much larger peak in the case of $S_{max} = 12$. Table III shows the 95% and 50% (median) of the time-slot utilizations, as well as the ratio between them. These results suggest that with T&C in place, the ratio is nearly 1.0, resulting in a perfect flattening of traffic over time slots, thus eliminating cost problems derived from spikes when using the 95/5 rule.

		95%	Median	Ratio
Original		197.15	124.56	1.583
T&C	$S_{max} = 6$	136.52	135.93	1.004
T&C	$S_{max} = 12$	138.05	137.33	1.005

TABLE III
TRAFFIC VOLUME STATISTICS (IN MB) WITH AND WITHOUT T&C

How Does T&C Enable an ISP to Cap its Aggregate Traffic Volume? The ISP is able to specify a target total traffic volume on the managed link through its choice of the resistance parameter R (which directly affects the constant C and hence the budget B_i allocated to each user). Figure 9a shows the total allocation per time-slot as a function of R , when *slack*=0 (which is the worst-case in the sense that under this scenario, the budgets are constrained the most). As expected, R effectively controls the the aggregate traffic volume resulting from T&C. This volume is almost flat due to the “fluid” nature of FT bandwidth allocation. The exception is due to spikes underscoring the presence of large IT sessions that could not be smoothed out under the chosen slack value. Naturally, these spikes dissipate when larger slack values are used (see Figure 8).

¹¹ For large values of R , the budget constraint may impact IT allocations. In the rare event when this happens, the policy we adopted was to randomly drop user sessions in case the user runs out of budget in the trading phase.

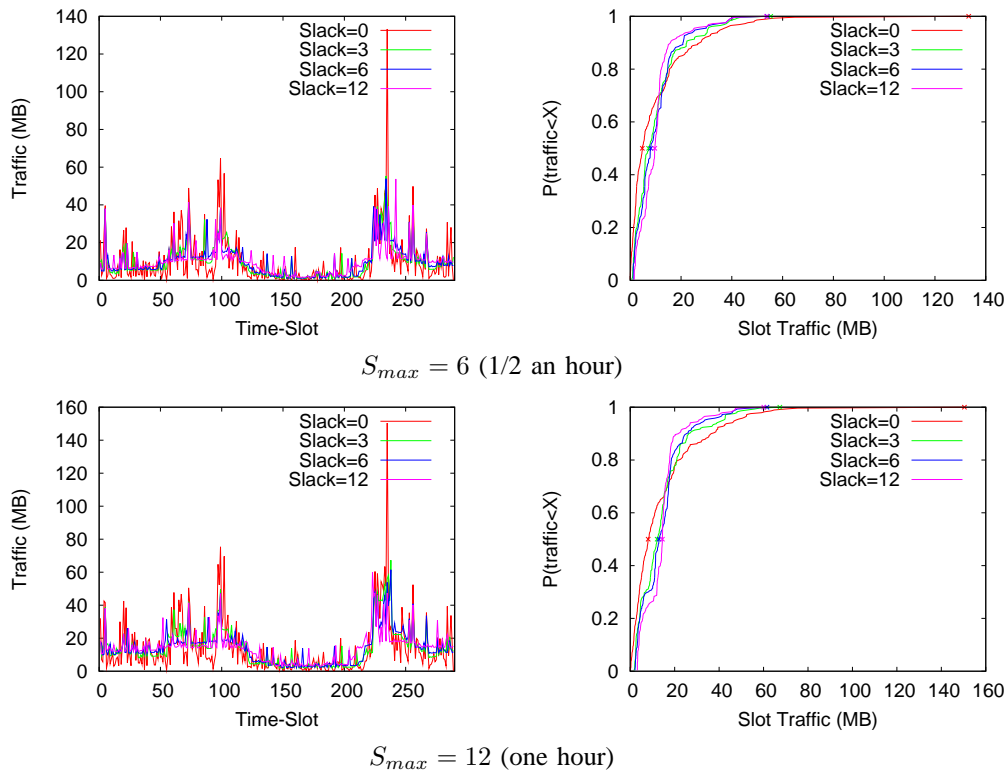


Fig. 7. Utilization over time for IT sessions with various slack values.

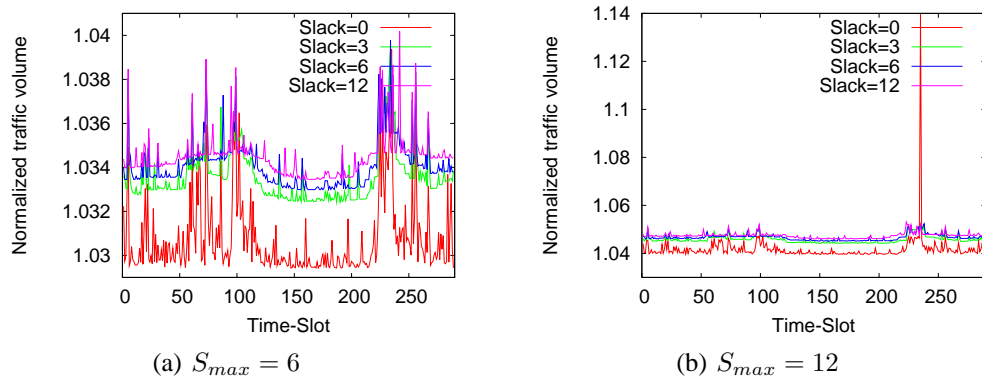
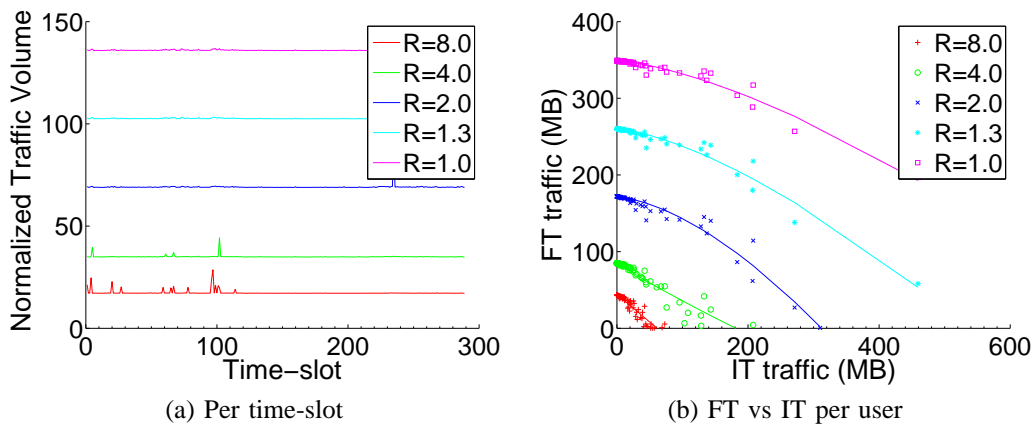


Fig. 8. Total Traffic (IT+FT) for various slack values.

How Does ISP Resistance Impact the Allocation of FT Traffic Relative to IT Traffic? Figure 9b compares the per-user bandwidth allocations for different values of the resistance, R . As before, the general trend is that the more IT bandwidth requested by a user during the trading phase, the less FT allocation the user is able to secure during the capping phase. Increasing the values of R results in a corresponding reduction in the aggregate allocation of FT bandwidth, with large IT bandwidth consumers impacted the most.

How Does T&C Impact the User’s Bottom Line? To evaluate T&C on a per-user basis, we compare how IT and FT allocations vary across users. Figure 10 (left) shows a clear negative correlation between the allotment of FT and IT bandwidth. The relationship is not monotonic or deterministic because it depends on the outcomes from the trading phase, which affect the left-over budget for each player. It is always

the case though that the larger the slack, the larger the FT allocation for any given user (points along the same vertical line in the plot). An agent with fixed IT demand increases its allocation of FT bandwidth when it adds more flexibility to its IT sessions. The results in Table IV expose this tradeoff for selected levels of IT demand and resistances. For example, when $R = 4$, a user with a nominal 100MB of IT bandwidth is able to capture 32% more FT traffic by accepting a minimal slack of 3 for its IT sessions. A rather surprising (and also desirable) finding – evident from Figure 10 and Table IV – is that the user begets *most* of the benefit by introducing a minimal amount of slack. Increasing the slack much beyond that results in only marginal increases in FT allocation. In the above example, by doubling its slack from 3 to 6, the user is able to capture only 3% more FT traffic. The message is clear: it “pays” to be flexible, even if minimally so.

Fig. 9. Traffic allocations for variable R .

	R=4.0	R=2.0	R=1.0
Slack	100MB	200MB	400MB
3	1.3190	1.2836	1.1931
6	1.3497	1.3338	1.2329
12	1.4079	1.3769	1.2520

TABLE IV

FT BANDWIDTH GAIN FOR VARIOUS VALUES OF R AND IT DEMAND.

Figure 10 (right) shows the same results on a semi-log scale to expose the outcome for users with negligible demand for IT bandwidth. In this case, the capping phase assigns to all such users almost equal share of the capacity (as expected). It is only the heavy IT bandwidth hogs who are unable to claim much FT bandwidth, which is precisely the premise of T&C.

VII. OTHER APPLICATION SCENARIOS FOR T&C

Non-cooperative load-balancing problems arise in a multitude of situations of which the DSLAM is a very practical example – an example that we chose to highlight in this paper. Our T&C formulation and associate model is quite general and can be applied to many other scenarios where customers’ tasks can be modeled as a combination of atomic and fluid processes and all the customers compete to complete their tasks with the lowest cost. We present additional examples below.

A. Energy-Aware Task Scheduling

Greenberg et al [29] provide a detailed analysis of the costs associated with modern cloud datacenters. Two important components in provisioning datacenter resources are energy and network capacity. Both resources are typically charged based on the 95/5 rule, and for the case of the datacenter this is a direct cost, making the incentive for the reduction of peak utilization more direct, but without changing the fundamental characteristics of the resource marketplace we have presented. In particular, energy requirements of different tasks can be described as vectors of power consumption per time-slot, $T_i = (t_{i1}, \dots, t_{in})$. Tasks may also be constrained to be executed within some time-interval and the charge associated with the execution of the tasks is determined by the total energy consumption according to eq. (1). Then, the customers can schedule the execution of their tasks by using

the trading mechanism as already described in §III. Similarly, there are fluid tasks that may need to use all the capacity available to them, and which run forever. Examples of such tasks are the crawling, indexing and ranking processes on a web search engine. We can think of these tasks as fluid-tasks and assign them a variable amount of resources per time-slot as to maximize the total amount of work they can perform at the lowest cost. In addition, the possibility of assigning budgets to different tasks permits adjusting the fraction of the resources they get. In fact Greenberg et al suggest using pricing and “urgency of execution” as parameters to reduce the peak-to-valley ratio on the utilization of these resources, precisely the notions captured by our mechanism.

B. Bulk Data Transfers

Another example application is proposed by Laoutaris et al [28]. They realize the possibility of transferring *Delay Tolerant Bulk (DTB)* data, such as large datasets, content replication batches, etc., at no cost by using a water-filling technique on the transit links. Under their model there is a single enterprise-customer who wants to transfer the DTB data between two points u and v , and the 95/5 rule determines its cost. So, as long as all the bulk data can be scheduled during the valleys of the demand curve, it is possible to preserve the 95% thus avoiding any extra costs.

Our fluid-allocation scheme allows to generalize this model to the case where many “fluid consumers” want to share those valleys in the transit links. In this case the proper settings of the budget and resistance parameters allows to control the shares per user (inducing a notion of fairness) and to cap the total utilization (so the 95% is preserved).

C. Large Enterprises Sharing Bottleneck Links of a Common Provider

Our T&C formulation of the DSLAM problem allows for a generalization whereby many large customers connect to an ISP, but the traffic cost depends mainly on the total traffic to upper tier providers. Such arrangement is illustrated in Figure 11, where the customers C_i share the links $ISP_1 - ISP_2$ and $ISP_1 - ISP_3$. If the main drivers of the cost for ISP_1 are

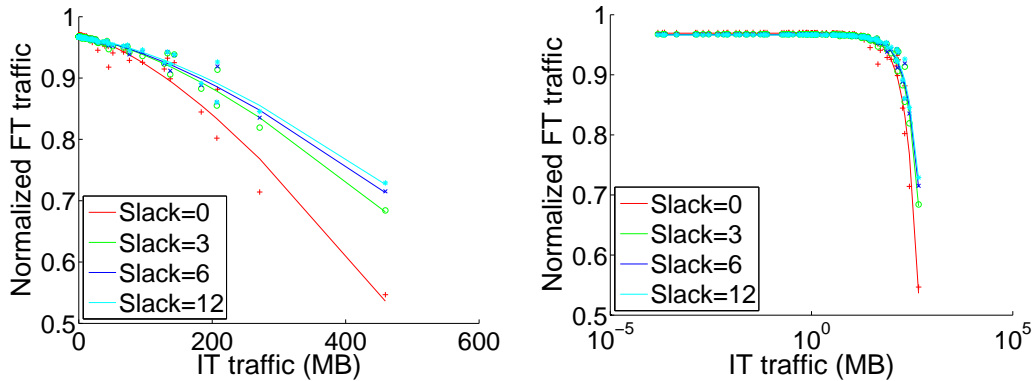


Fig. 10. IT and FT allocations per user for different slack values.

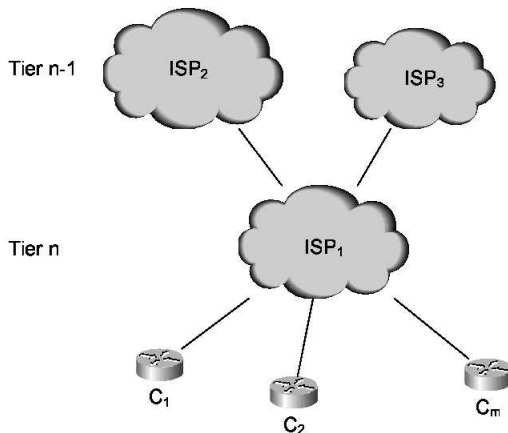


Fig. 11. Large customers sharing the links to upper tier providers

the cost of the links to upper-tier providers, then it is the goal of ISP_1 to incentive its customers to load-balance their traffic over time. This reduces the cost for ISP_1 , as well as the cost for the customers. In fact, a cost function of the form

$$c_i = \sum_{j=1}^k a_j f(U_j, w_{ij})$$

where each customer is charged according to the utilization U_j of link j and its contribution w_{ij} leads to a situation where the customers play to minimize the cost of each one of the k links to upper-tier providers independently. To observe this, notice that under standard routing protocols (*e.g.* BGP) the link used by each one of the sessions of C_i is predetermined by the routing protocol which is out of the control of the customer. This limits the options of the customers simply to play with the schedule of its sessions (bandwidth trading) and its allocation of fluid traffic (capping phase) independently over each link. Being the contributions to the cost per link independent of each other, the minimum is obtained as the sum of the minimums of all the links.

D. Iterated Trading for Multi-class QoS

The different IT sessions of a single user may capture many different QoS requirements. So for example a single user may have concurrent sessions of HTTP browsing, e-mail, VoIP,

online gaming, VoD, instant messaging, etc. Now, if the ISP offers a differentiated services [30] mechanism to handle the QoS mappings, then the cost function of eqn. (1) needs to be adjusted to reflect the cost of each one of the traffic classes. More precisely, let $k = 1 \dots K$ identify the service class. Then, the total cost of user i is

$$c_i = \sum_{k=1}^K c_{ik}$$

where the cost per class is

$$c_{ik} = \frac{1}{C_k} \sum_{p=1}^T x_{ikp} U_{kp} \quad (12)$$

Observing that the costs per class are independent of each other, we are then left with K independent cost minimization problems (under the assumption that the budget is enough to pay for all the IT traffic). In fact, the bandwidth trading phase for each class may be executed concurrently with the others, and the minimum user cost is the sum of the minimums per class, therefore it is still a NE. A subtle but important observation is that in making eq. (12) dependent only on the utilization per class, other classes have no influence in the performance of a given class. This is consistent with the Weighted Fair Queueing (WFQ) systems[31], but also requires the intervention of an admission control mechanism that avoids saturating any given class. The admission control mechanism corresponds in our model to having additional constraints on the set of actions available to the players, which as discussed in §III does not affect the properties about existence or convergence to a NE.

As for the FT, the left-over budget is computed as

$$\ell_i = B_i - \sum_{k=1}^K c_{ik}$$

and it proceeds the same as before. Because of its fluid nature, it is implicit that fluid traffic will be assigned to the lowest priority class.

VIII. RELATED WORK

While the application of game-theoretic and micro-economic approaches to networking problems is not novel [32], [23], [4],

[33], [34], our approach of strategically trading-off allocation slots based on desirable properties for different traffic classes is new and quite promising.

Laoutaris and Rodriguez [5] recognized that the problems associated with rampant delay-tolerant (what we call FT) traffic are due to the lack of incentives for end-users to properly schedule such delay-tolerant traffic and the lack of network mechanisms to identify and handle such traffic. As a solution to the first problem, they suggest giving users “higher-than-purchased” access rate during off-peak hours as a reward for time-shifting their FT traffic. As a solution to the second problem, they propose the introduction of a store-and-forward service to handle the network transfer of bulk delay-tolerant data during off-peak hours.

Fairness is a very controversial issue with no universally-accepted definition. The most commonly used definition is that of *max-min fairness*, whereby no user can increase its rate at the expense of other users with lower rates. Max-min fairness deals with instantaneous rates, and thus is useless over long time scales under time-varying demands. In many contexts, fairness is a property established across flows (*e.g.*, TCP’s max-min fairness). Clearly, this definition breaks when a single entity (user) is able to open multiple concurrent flows, as it is indeed the case in many applications. Briscoe [35] gives a very thorough discussion of the issues involved. He advocates a notion of *cost fairness* between economic entities, thus avoiding both the per-flow and the instantaneous connotations. This is consistent with T&C’s assignment of budgets to user agents as the primary mean for ensuring fairness.

Recently, Briscoe et al [36] proposed an architecture that operates at the network edges and realizes the *cost fairness* model without directly charging users (hence, compatible with flat pricing). This work introduces *re-feedback*, a mechanism that allows measurement of downstream path metrics, such as delay and congestion. This information can then be used to police the compliance of end-users with a predetermined policy (*e.g.* backoff the sending rate in case of congestion). The network itself can perform the policing function requiring only a shaper at the ingress point and a dropper at the egress point. When doing so, it is the dominant strategy for end-points to report the correct metrics. This is a congestion control mechanism, that provides the necessary feedback for flows to adjust their rates, and for the network to police response to congestion. It is strictly a best-effort scheme, and unlike T&C it does not provide the means for applications with specific QoS goals to make trade-offs that satisfy their requirements.

Approaches for *congestion-pricing* with explicit payments have been considered in a number of studies. Henderson *et al* [37] present a review of the benefits and limitations of these proposals. Examples include *Smart Markets* [33], [38] and *Split-Edge Pricing* [39]. Of particular interest is the scheme proposed by Ganesh *et al* [23], which assigns costs to packets depending on congestion. Under a family of non-linear cost functions that depend on the utilization of the congested link and the flow’s demand, they showed convergence to steady-

state equilibrium. While our mechanism and system model are entirely different, our cost function has similar characteristics.

Several works have also studied the [34], [40], [41] priority queueing systems (a la Diffserv) under game-theoretic frameworks. So for example, Marback [34] analyzes a priority queueing scheme where packets get charged based on their priority, and selfish users compete for bandwidth. Among other things, he shows that such a scheme leads to a Wardrop equilibrium and that allocation does not depend on the prices of each traffic class. A fundamental distinction in this case is that T&C enables different valuations for different classes of traffic, and uses these valuations to leverage the trading system. Park et al [40] consider a QoS class assignment game where users share a single WFQ queue and they can assign the class for the traffic. Users do so, to meet the QoS requirements of their application, which are dependent on the total allocation due to all the users. In this work, they consider both, the case where traffic may be arbitrarily splitted between the many service classes and the *unsplittable* case where all the traffic is assigned to the same class. In the splittable case, NE need not exist, but it is proven that in the unsplittable case NE always exists. Fundamental differences with respect to this model are the fact that users are not traffic maximizers (their traffic and QoS parameters are fixed) so that there is no need to concern about fairness; each user represents a single application, and the game’s action space are the instantaneous rate allocations, thus it does not consider finite sessions with schedule flexibility. In [41], the authors consider the assignment of service classes to each user’s traffic at each one of the routers in a path. In this analysis, each user provides a QoS vector and a utility function, and the user actions are the choices of service classes at each router, such that his traffic will meet the QoS goals with minimum cost. This model is limited to the *unsplittable case*, meaning that all the traffic from a user is assigned the same service class. The incentive for the users is implicit in the price-by-class scheme, where users requesting higher priority classes pay more. In addition, payment has to be made to all intermediary nodes on a route. Chen et al [42] also provide an efficient distributed implementation and evaluation of their multi-switch QoS assignment game, where agents running at the routers and end-points compute the game outcome on behalf of the users. The performance evaluation shows a significant improvement on the per-application QoS metrics with respect to a static reservation mechanism.

A fundamental distinction between T&C and the various congestion pricing schemes considered in the literature ([36], [43], [37], [23]) is that none of these schemes takes into account the dual nature (IT vs. FT) of applications. Therefore, all these schemes impose penalties (*e.g.* larger cost, increase drop rates) to all the traffic from a user during congestion periods. Because they operate over short-time-scales (targeting an instantaneous response to congestion), none of these approaches exploits the extra degree of freedom offered by the possibility of time-shifting the execution of IT tasks, or

adjusting the rate of FT tasks.

IX. CONCLUSION

Trade & Cap (T&C) is an effective bandwidth management mechanism that enables self-interested user agents to collectively converge on what *they* perceive to be an equitable allocation, based on their individual, private valuation of network utility (e.g., raw volume vs. QoS over time). Trade & Cap not only benefits users by allowing them to extract better utility from the network, but also benefits the ISP by yielding smoother aggregate traffic volumes, which lowers traffic transit costs and reduces the currently unsustainable pressure on ISPs to upgrade their networks in order to keep up with peak demand. Under Trade & Cap, rather than acting as an arbiter, an ISP acts as an enforcer of what the community of rational users sharing the resource decides is a fair allocation of that resource. This is a welcome departure from current practices that force ISPs to use artificial notions of fairness to police shared bandwidth use, with negative implications to privacy and network neutrality.

REFERENCES

- [1] M. H. Bosworth, "Time warner: Metered broadband will prevent "internet brownouts"," April 10 2009. [Online]. Available: <http://tinyurl.com/cz4req>
- [2] W. Gruener, "Time warner shelves metered internet plans - for now," April 16 2009. [Online]. Available: <http://www.tgdaily.com/content/view/42055/103/>
- [3] A. Odlyzko, "Internet pricing and the history of communications," *Computer Networks*, vol. 36, pp. 493–517, 2001. [Online]. Available: <http://dx.doi.org/10.2139/ssrn.235283>
- [4] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997. [Online]. Available: <http://dx.doi.org/10.1002/ett.4460080106>
- [5] N. Laoutaris and P. Rodriguez, "Good Things Come to Those Who (Can) Wait or how to handle Delay Tolerant traffic and make peace on the Internet," in *HotNets'08*, 2008. [Online]. Available: <http://conferences.sigcomm.org/hotnets/2008/program.html>
- [6] L. Berraill, R. Teixeira, and K. Salamati, "Early application identification," in *CoNEXT '06: Proceedings of the 2006 ACM CoNEXT conference*. New York, NY, USA: ACM, 2006, pp. 1–12. [Online]. Available: <http://dx.doi.org/10.1145/1368436.1368445>
- [7] T. Karagiannis, D. Papagiannaki, and M. Faloutsos, "BLINC: multilevel traffic classification in the dark," in *SIGCOMM '05: Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM, 2005, pp. 229–240. [Online]. Available: <http://dx.doi.org/10.1145/1080091.1080119>
- [8] T. Karagiannis, A. Broido, M. Faloutsos, and K. claffy, "Transport layer identification of P2P traffic," in *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement (IMC)*. New York, NY, USA: ACM, 2004, pp. 121–134. [Online]. Available: <http://dx.doi.org/10.1145/1028788.1028804>
- [9] A. W. Moore and D. Zuev, "Internet traffic classification using bayesian analysis techniques," in *SIGMETRICS*. New York, NY, USA: ACM, 2005, pp. 50–60. [Online]. Available: <http://dx.doi.org/10.1145/1064212.1064220>
- [10] A. W. Moore and K. Papagiannaki, "Toward the accurate identification of network applications," in *Passive and Active Network Measurement (PAM'05)*, 2005, pp. 41–54. [Online]. Available: <http://dx.doi.org/10.1007/b135479>
- [11] S. Sen, O. Spatscheck, and D. Wang, "Accurate, scalable in-network identification of P2P traffic using application signatures," in *WWW '04: Proceedings of the 13th international conference on World Wide Web*. New York, NY, USA: ACM, 2004, pp. 512–521. [Online]. Available: <http://dx.doi.org/10.1145/988672.988742>
- [12] L. G. Roberts, "A radical new router," *IEEE Spectr.*, vol. 46, no. 7, pp. 34–39, July 2009. [Online]. Available: <http://dx.doi.org/10.1109/MSPEC.2009.5109450>
- [13] F5 Networks, Inc., "Bandwidth management for P2P applications." [Online]. Available: <http://www.f5.com/pdf/white-papers/rateshaping-wp.pdf>
- [14] iPoque, "Bandwidth management with deep packet inspection," 2009. [Online]. Available: <http://www.ipoque.com/>
- [15] E. Orion, "Comcast internet throttling is up and running," Jan 6 2009. [Online]. Available: <http://tinyurl.com/n3m8lw>
- [16] N. Anderson, "DPI vendor says 90% of ISP customers engage in traffic discrimination," Aug 3 2009. [Online]. Available: <http://tinyurl.com/kvtbc6>
- [17] J. Crowcroft, "Net neutrality: the technical side of the debate: a white paper," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 1, pp. 49–56, Jan 2007. [Online]. Available: <http://dx.doi.org/10.1145/1198255.1198263>
- [18] N. Anderson, "Network neutrality in congress, round 3: Fight!" Aug 3 2009. [Online]. Available: <http://tinyurl.com/mn7pgg>
- [19] K. Bogardus and K. Hart, "Obama's FCC to enforce 'net neutrality,'" Aug 25 2009. [Online]. Available: <http://tinyurl.com/l63z8j>
- [20] A. Odlyzko, "Pricing and architecture of the internet: Historical perspectives from telecommunications and transportation," in *Proceedings of TPRC*, 2004. [Online]. Available: <http://www.dtc.umn.edu/~odlyzko/doc/pricing.architecture.pdf>
- [21] C. Labovitz, "The internet after dark (part 1)," Aug 24 2009. [Online]. Available: <http://asert.arbornetworks.com/2009/08/the-internet-after-dark/>
- [22] The Canadian Press, "Most Canadians support reasonable internet traffic management, poll suggests," July 14 2009. [Online]. Available: <http://tinyurl.com/m6atbr>
- [23] A. Ganesh, K. Laevens, and R. Steinberg, "Congestion pricing and user adaptation," in *Proceedings IEEE INFOCOM*, 2001, pp. 959–965. [Online]. Available: <http://dx.doi.org/10.1109/INFCOM.2001.916288>
- [24] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA: W.H. Freeman and Co., 1979.
- [25] R. K. Sundaram, *A First Course in Optimization Theory*. Cambridge University Press, 1996. [Online]. Available: <http://www.cambridge.org/us/catalogue/catalogue.asp?isbn=9780521497701>
- [26] MAWI Working Group, "Traffic archive," 2009. [Online]. Available: <http://tracer.csl.sony.co.jp/mawi/>
- [27] R. D. Torres, M. Y. Hajjat, S. G. Rao, M. Mellia, and M. M. Munafo, "Inferring undesirable behavior from P2P traffic analysis," in *SIGMETRICS*. New York, NY, USA: ACM, 2009, pp. 25–36. [Online]. Available: <http://dx.doi.org/10.1145/1555349.1555353>
- [28] N. Laoutaris, G. Smaragdakis, P. Rodriguez, and R. Sundaram, "Delay tolerant bulk data transfers on the internet," in *SIGMETRICS*. New York, NY, USA: ACM, 2009, pp. 229–238. [Online]. Available: <http://dx.doi.org/10.1145/1555349.1555376>
- [29] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: Research problems in data center networks," *CCR Online*, Jan 2009. [Online]. Available: <http://ccr.sigcomm.org/online/?q=node/436>
- [30] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services." [Online]. Available: <http://www.ietf.org/rfc/rfc2475.txt>
- [31] A. Demers, S. Keshav, and S. Shenker, "Analysis and simulation of a fair queueing algorithm," in *SIGCOMM '89: Symposium proceedings on Communications architectures & protocols*. New York, NY, USA: ACM, 1989, pp. 1–12. [Online]. Available: <http://dx.doi.org/10.1145/75246.75248>
- [32] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker, "A BGP-based mechanism for lowest-cost routing," *Distrib. Comput.*, vol. 18, no. 1, pp. 61–72, 2005. [Online]. Available: <http://dx.doi.org/10.1007/s00446-005-0122-y>
- [33] J. MacKie-Mason and H. Varian, *Public Access to the Internet*. MIT Press, 1995, ch. Pricing the Internet. [Online]. Available: <http://mitpress.mit.edu/catalog/item/default.asp?type=2&tid=5603>
- [34] P. Marbach, "Analysis of a static pricing scheme for priority services," *IEEE/ACM Trans. Netw.*, vol. 12, no. 2, pp. 312–325, Apr 2004. [Online]. Available: <http://dx.doi.org/10.1109/TNET.2004.826275>
- [35] B. Briscoe, "Flow rate fairness: Dismantling a religion," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 2, pp. 63–74, April 2007. [Online]. Available: <http://dx.doi.org/10.1145/1232919.1232926>

- [36] B. Briscoe, A. Jacquet, C. D. Cairano-Gilfedder, A. Salvatori, A. Soppera, and M. Koyabe, "Policing congestion response in an internetwork using re-feedback," in *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 4. ACM Press, Aug 2005, pp. 277–288. [Online]. Available: <http://dx.doi.org/10.1145/1080091.1080124>
- [37] T. Henderson, J. Crowcroft, and S. Bhatti, "Congestion pricing. Paying your way in communication networks," *IEEE Internet Comput.*, vol. 5, no. 5, pp. 85–89, Sep/Oct 2001. [Online]. Available: <http://dx.doi.org/10.1007/978-3-540-72990-7>
- [38] J. MacKie-Mason and H. Varian, "Pricing congestible network resources," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1141–1149, Sep 1995. [Online]. Available: <http://dx.doi.org/10.1109/49.414634>
- [39] B. Briscoe, "The direction of value flow in connectionless networks," in *Networked Group Communication*, 1999, pp. 244–269. [Online]. Available: <http://dx.doi.org/10.1007/b72228>
- [40] K. Park, M. Sitharam, and S. Chen, "Quality of service provision in noncooperative networks: heterogenous preferences, multi-dimensional QoS vectors, and burstiness," in *ICE '98: Proceedings of the first international conference on Information and computation economies*. New York, NY, USA: ACM, 1998, pp. 111–127. [Online]. Available: <http://dx.doi.org/10.1145/288994.289022>
- [41] S. Chen and K. Park, "An architecture for noncooperative QoS provision in many-switch systems," in *Proceedings IEEE INFOCOM*, vol. 2, Mar 1999, pp. 864–872. [Online]. Available: <http://dx.doi.org/10.1109/INFCOM.1999.751475>
- [42] —, "A distributed protocol for multi-class QoS provision in noncooperative many-switch systems," in *Proceedings of the Sixth International Conference on Network Protocols*, Oct 1998, pp. 98–107. [Online]. Available: <http://dx.doi.org/10.1109/ICNP.1998.723730>
- [43] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998. [Online]. Available: <http://dx.doi.org/10.1057/palgrave.jors.2600523>