On Asymptotic Optimality of Dual Scheduling Algorithm In A Generalized Switch

Lijun Chen, Steven H. Low and John C. Doyle
Engineering & Applied Science Division, California Institute of Technology
Pasadena, CA 91125, USA

Abstract— **Generalized switch is a model of a queueing system where parallel servers are interdependent and have time-varying service capabilities. This paper considers the dual scheduling algorithm that uses rate control and queue-length based scheduling to allocate resources for a generalized switch. We consider a saturated system in which each user has in£nite amount of data to be served. We prove the asymptotic optimality of the dual scheduling algorithm for such a system, which says that the vector of average service rates of the scheduling algorithm maximizes some aggregate concave utility functions. As the fairness objectives can be achieved by appropriately choosing utility functions, the asymptotic optimality establishes the fairness properties of the dual scheduling algorithm.**

The dual scheduling algorithm motivates a new architecture for scheduling, in which an additional queue is introduced to interface the user data queue and the time-varying server and to modulate the scheduling process, so as to achieve different performance objectives. Further research would include scheduling with Quality of Service guarantees with the dual scheduler, and its application and implementation in various versions of the generalized switch model.

I. Introduction

We consider a general model where a set S of queues (users) are served in discrete time by a generalized switch, as de£ned in [22]. The generalized switch can be viewed as a discrete-time, interdependent parallel server system. The servers are interdependent in that they cannot provide service simultaneously, and the dependency among them is re¤ected on the constraints that specify which subsets of servers can be active at the same time. Switch state h follows a discrete-time, £nite-state Markov chain. At each time slot t, the switch can choose a scheduling decision m from a £nite set M, which captures the constraints imposed by the interdependency among the servers. Each scheduling decision has the associated vector of service rates $\tilde{r}^m(h(t))$ at which queues are served, where $h(t)$ is the switch state at time t.

The generalized switch model has many applications in communication networks. For example, in cellular network in the downlink, the servers correspond to the wireless links from the base stations to the users, and the constraint is that each base station can transmit to at most one of the users and each user can be served by at most one of the base stations in each time slot, see e.g. [5], [10]. Other examples include multi-hop wireless network where each wireless link can be viewed as a server and the constraints disallow simultaneous transmission of neighboring links due to interference, see e.g. [26], [27], [28]. It also includes as a single-state special case input-queued cross-bar switch where a server corresponds to each input-output port pair, and the constraint is that each input port transmits to exactly one of the output ports and each output port receives from exactly one of the input ports at any time, see e.g. [17]. The same model can extend to handle the packet switch in wireless network, where the switch state (i.e., wireless line rates) is supposedly time-varying.

For such a generalized switch system with time-varying state, the service rate that can be offered to the users (queues) is both user-dependent and time-dependent. This, on one hand, opens up the possibility to use state-aware scheduling strategies, i.e., to exploit service variations to increase the throughput. On the other hand, the parallel servers are interdependent, and to serve (schedule) always the user with the highest potential rate maximizes overall throughput but usually results in the starvation of some users. So, we need to trade off throughput for fairness. However, the time-varying nature of the generalized switch, coupled with the user-dependent service rate and unknown data arrival, makes it very challenging to design scheduling policies to ful£l fairness and throughput requirements, as well as other performance objectives.

In this paper, we study the dual scheduling algorithms for the generalized switch. These algorithms are motivated by the dual subgradient algorithm of convex optimization problems [21], [6]. With an additional queue (termed M-queue) being introduced for each user, the dual scheduling algorithm is a combination of rate control (of the M-queue) and M-queue-length based scheduling. The rate control algorithm is motivated by utility framework for TCP congestion control (see e.g. [12], [16]), which shows that various TCP congestion control protocols can be interpreted as distributed primal-dual algorithms to solve aggregate network utility maximization. The queue-length based scheduling takes the form of a simple throughput-optimal scheduling [26], [27], [28]. As such, while the queue-length-based scheduling part keeps maximizing the throughput, the rate-control part modulates the scheduling process by choosing appropriate utility functions, so as to achieve various performance objectives.

Section III presents the details of system model and the dual scheduling algorithm for the generalized switch. In Section IV, we consider a saturated system in which each user has in£nite amount of data to be served. For such a system, fairness among the users is presumably the most important concern. We present a dual scheduling algorithm, which can be seen

0-7803-9550-6/06/$20.00

as an extension of the algorithm studied in [9], and prove its asymptotic optimality, which says that the vector of average service rates of the scheduling algorithm maximizes some aggregate concave utilities of the users. As is well-known, the fairness objectives can be achieved by appropriately choosing the utility functions. So, the asymptotic optimality establishes the fairness properties of the dual scheduling algorithm. Also, the proofs presented in this section are rather general. They only use general properties of convex/concave function, subgradient and convex set, and represent an integration of convex optimization and stochastic control, and can be readily extended to other systems.

The dual scheduling algorithm motivates a new architecture for scheduling in the generalized switch, in which an additional queue is introduced to interface the user data queue and the time-varying server and to modulate the scheduling process, so as to achieve different performance objectives such as fairness, and maximum throughput, etc. In Section V, we will brie¤y discuss some implementation issues and advantages of the dual scheduler, and Quality of Service scheduling in generalized switches.

II. Related Work

There exists lots of work on scheduling with different performance objectives for different versions of the generalized switch model. For fair scheduling, in the context of cellular network in the downlink, one of the principal policies is the Proportional Fair Scheduler of Qualcomm High Data Rate system [5], [10], which schedules the user with the largest ratio of the current achievable rate to the exponentially smoothed throughput. This scheduling algorithm has been shown to maximize the sum of the logarithm utilities of the long-run average data rates provided to the users [29], [13], [25], and thus achieve proportional fairness [11]. The generalization of the proportional fair scheduling algorithm to any concave utility function for a generalized switch has been studied[1], see e.g. [23]. Other work on fair scheduling includes, e.g., [14], [7], [15]. For throughput-optimal scheduling that attains maximum stability region of the system, one of the principal policies is the MaxWeight scheduling in the context of wireless network, see e.g. [26], [27], [28], [1], [20], [18], [8], and in the context of input-queued switch, see e.g. [17]. The stability region of a scheduling policy is the set of mean ¤ow rate vectors such that the queue-length process is stable under this policy. The throughput-optimal scheduling has its origin in [26], [27], [28], where it is shown that allocating resources to maximize a queue-length-weighted sum of rates is a stabilizing policy under any sustainable ¤ows. However, there is no fairness guarantee with throughput-optimal scheduling.

The dual scheduling algorithm for saturated systems can be seen as an extension of the algorithm studied in [9] to generalized switches with general user utilities. They also use different proof techniques for asymptotic optimality. The connection between fair resource allocation and duality can also be found in [19] and subsequent works. Similar result on asymptotic optimality is also obtained in [24] through a much different technique.

[1]We call this type of scheduling policies the primal scheduling algorithms, since they can be seen as the gradient algorithm to solve a concave utility maximization problem directly.

III. System Model

We consider a queueing system where a £nite set S of parallel queues (users), indexed by s, are served by a generalized switch. The generalized switch can be abstracted as an interdependent parallel server system. The servers are interdependent in that they cannot provide service simultaneously, and the dependency among them is re¤ected on the constraints that specify which subsets of servers can be active at the same time. For convenience, we use a "dependency" graph G to capture this interdependency. Each vertex in G represents a server, and an edge between two vertices means the corresponding servers cannot be active simultaneously. Thus, only those servers in an independent set[2] of the dependency graph can be active at the same time. We denote the set of independent sets by M, with each element indexed by m.

The system operates in discrete time $t = 0, 1, 2, \cdots$. By convention, we choose the duration of a time slot as the *unit* of time, and identify time t with the unit time interval $[t, t+1)$. The switch has a £nite set H of states. The switch state is £xed in one of the states $h \in H$ within a time slot but varies across slots according to an irreducible £nite-state Markov chain. Corresponding to the switch state h, the service rate to user s is $r_s(h)$ packets per time slot when the switch servers only s, and the service rate vectors $\tilde{r}^m(h),\ m \in M$ that can be offered to the users are

$$\tilde{r}_s^m(h) = \begin{cases} r_s(h) & \text{if } \ s \in m \\ 0 & \text{otherwise.} \end{cases}$$

By standard time-sharing argument, the feasible rate region $\Pi(h)$ in switch state h is de£ned to be the convex hull of these rate vectors [4]

$$\Pi(h) := \left\{ \tilde{r} : \tilde{r} = \sum_{i=1}^{M} t_i \tilde{r}^i(h),\ t_i \geq 0,\ \sum_{i=1}^{M} t_i = 1 \right\}, \qquad (1)$$

where we slightly abuse the notation and let M also denote the size of the set M. Let the switch state distribution be $d(h)$, we further de£ne the mean feasible rate (capacity) region as

$$\overline{\Pi} = \left\{ \overline{r} : \overline{r} = \sum_{h \in H} d(h) \tilde{r}(h), \tilde{r}(h) \in \Pi(h) \right\}. \qquad (2)$$

This mean rate region is a closed convex set, and is the best feasible rate region the system can support on average.

A. Queue Length Dynamics

Fig.1 shows the architecture of the dual scheduler from the perspective of one user. The system keeps separate *data queues* for the users to buffer the data intended to them. In addition, another queue, called M-queue, is introduced for each user.

[2]An independent set of vertices is de£ned as a set of vertices that have no edges between each other. An empty set is an interdependent set.

The M-queue interfaces the data queue and the time-varying server, in that the data will depart from the data queue to enter the M-queue, and the server will directly serve the M-queue.

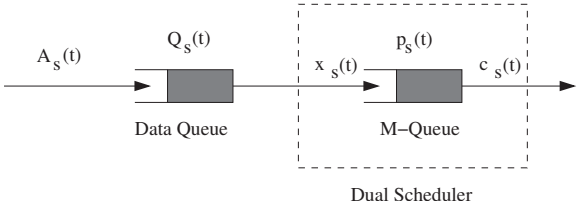

Fig. 1. The Architecture of the Dual Scheduler.

Denote the size of the data queue and M-queue for user s at the beginning of the time slot t by $Q_s(t)$ and $p_s(t)$ respectively, the number of arrivals to the data queue and M-queue of user s in time slot t by $A_s(t)$ and $x_s(t)$ respectively, and the amount of service offered to the M-queue of user s in time slot t by $c_s(t)$. The evolutions of the data queue and M-queue length for user s are given by

$$Q_s(t+1) = Q_s(t) + A_s(t) - x_s(t) \ , \quad (3)$$
$$p_s(t+1) = [p_s(t) + x_s(t) - c_s(t)]^+, \quad (4)$$

where '$+$' denotes the projection onto the set $\Re^+$ of non-negative real numbers.

We further introduce a small parameter $\gamma > 0$, and for convenience, de£ne a new quantity $q_s(t) = \gamma p_s(t)$ for each user s. In Section IV we will see that γ characterizes the asymptotic optimality and fairness of the dual scheduling algorithm. We call q the scaled queue-length, since it is the M-queue length scaled by γ. By equation (4), the evolution of the scaled queue-length is given by

$$q_s(t+1) = [q_s(t) + \gamma(x_s(t) - c_s(t))]^+. \quad (5)$$

With the dual scheduling algorithm, the system controls the arrival rate into the M-queues and determines service rates offered to the M-queues based on queue-length.

B. Dual Scheduling Algorithm

We assume that each user s attains a utility $U_s(x_s)$ when its arrival rate to the M-queue is x_s packets per time slot. $U_s(\cdot)$ may be dependent of data queue size Q_s, but is assumed to be continuously differentiable, increasing and strictly concave with respect to x_s. In time slot t, given the current M-queue length $p_s(t)$, the maximal arrival rate to the M-queue of user s is speci£ed as following

$$x_s(t) = \min\left\{U_s'^{-1}(\gamma p_s(t)), \alpha_s\right\}$$
$$= \min\left\{U_s'^{-1}(q_s(t)), \alpha_s\right\}, \quad (6)$$

where $\alpha_s > \max_h r_s(h)$ is the upper bound speci£ed on the arrival rate, and thus $x_s(t)$ maximizes $U_s(x_s) - q_s x_s$ over $0 \leq x \leq \alpha_s$. Note that we choose packet of equal length as the unit of data. x_s will be rounded to closest integer automatically.

We now consider service allocation. In time slot t, given the current M-queue length $p(t)$, the switch selects a (physical) service rate vector[3]

$$c(t) \in \arg\max_{c\in\Pi(h(t))} p(t)^T c = \arg\max_{c\in\Pi(h(t))} q(t)^T c \ , \quad (7)$$

where we will always pick an extreme point maximizer[4]. Equation (7) takes the form of simple throughput-optimal scheduling as proposed in [26], [27], which schedules the transmissions dynamically based only on current system backlog and switch state.

Equations (3)-(7) de£ne the dual scheduling algorithm. When the M-queue length process is stable, x_s *will be the service rate offered to user* s. This scheduling algorithm can be seen as motivated by the dual subgradient algorithm of concave maximization problem $\max_x \sum_s U_s(x_s)$, and is a combination of rate control [12], [16] and queue-length-based scheduling. As the queue-length-based scheduling part keeps maximizing the throughput, the rate-control part modulates the scheduling process by choosing appropriate utility functions, so as to achieve various performance objectives.

Given a scheduling algorithm, two of important issues that need to be addressed are to characterize its fairness property and its stability region. The fairness property governs the resource allocation among the competing users, and the stability region determines the ef£ciency of the scheduling algorithm as a whole. We will study the fairness property in this paper, and leave the stability region and other user-level performance for future work.

IV. Asymptotic Optimality and Fairness

We consider a saturated system in which each user has in£nite amount of data to be served, i.e., the user data queue is in£nitely backlogged. So, the data queue is irrelevant and the choice of utility function $U_s(\cdot)$ is independent of Q_s. The dual scheduling algorithm is thus de£ned by equations (4)-(7). We will show that the dual scheduling algorithm maximizes some aggregate concave utilities and establish its fairness properties through its asymptotic optimality.

A. An Ideal Reference System

Before preceding, let us £rst de£ne an ideal reference system problem,

$$\max_{x_s\geq 0, c_s\geq 0} \quad \sum_s U_s(x_s) \quad (8)$$
$$\text{subject to} \quad x \leq c \ \& \ c \in \overline{\Pi}. \quad (9)$$

The £rst constraint says that the arrival rate to the M-queues should not exceed the physical service rate. The second constraint says that the physical service rate should be in the mean rate region, which is the best feasible rate region the system can support. We will characterize the performance of

[3]We call the service rate allocated to the M-queue the physical service rate, in order to distinguish from the service rate received by the user data queue which will be x_s if M-queue is table.

[4]A point in a convex set is an extreme point if it cannot be written as a convex combination of other points in the convex set.

the dual scheduling algorithm with respect to this reference system.

Proposition 1: The solution x^* to problem (8)-(9) exists and is unique.

Proof: The proof is trivial, since the objective function is strictly concave and the constraint set is a closed, convex set [6]. ■

Consider the dual problem of the reference system problem (8)-(9)

$$\min_{u\geq 0} D(u) \tag{10}$$

with partial dual function

$$D(u) = \max_{x_s\geq 0, c_s\geq 0} \quad \sum_s U_s(x_s) - u^T(x-c) \tag{11}$$

$$\text{subject to} \quad c \in \overline{\Pi}, \tag{12}$$

where we relax only the constraint $x \leq c$ by introducing Lagrange multiplier u.

Proposition 2: The solution u^* to dual problem (10) exists. Moreover, there is no dual gap between the primal problem (8)-(9) and the dual problem (10).

Proof: The proof is trivial, since problem (8)-(9) is a convex optimization problem [6]. ■

Having established the properties of the ideal reference system problem and its dual, in the next subsection we will characterize the dual scheduling algorithm with respect to them.

Remark 1: Roughly speaking, the primal scheduling algorithm is a scheduling policy whose vector of average service rates solving the problem

$$\max_{x_s\geq 0} \quad \sum_s U_s(x_s)$$
$$\text{subject to} \quad x \in \overline{\Pi}.$$

This problem is equivalent to problem (8)-(9), since mathematically c can be seen as an auxiliary variable. The primal scheduling algorithm can be seen as being motivated by the gradient algorithm to solve this problem [23], while the dual scheduling algorithm can be seen as being motivated by the dual gradient algorithm to solve the same problem.

B. Stochastic Stability

Note that M-queue length $p(t)$ (and scaled queue-length $q(t)$) evolves according to a discrete-time, discrete-space Markov chain. We first show that this Markov chain is stable, i.e., the queue-length process reaches a steady state and does not go unbounded to infinity. It is easy to check that the Markov chain has a countable state space, but is not necessarily irreducible. In such a general case, the state space is partitioned in transient set T and different recurrent classes R_i. We define *the system to be stable if all recurrent states are positive recurrent and the Markov process hits the recurrent states with probability one* [26]. This will guarantee that the Markov chain will be absorbed/reduced into some recurrent class, and the positive recurrence ensures the ergodicity of the Markov chain over this class.

Theorem 3: The Markov chains described by equations (4) and (5) are stable.

Proof: Consider the the Lyapunov function $V(q) = \|q - u^*\|_2^2$. By equations (5)-(7) and define $g(q) = c(q) - x(q)$, we have

$$\begin{aligned}
& E[\Delta V_t(q)|q] \\
= \;& E[V(q(t+1)) - V(q(t)) \mid q(t) = q] \\
= \;& E[V([q(t) - \gamma g(q(t))]^+) - V(q(t)) \mid q(t) = q] \\
\leq \;& E[V(q(t) - \gamma g(q(t))) - V(q(t)) \mid q(t) = q] \\
= \;& E[-\gamma g(q(t))^T (2(q(t) - u^*) - \gamma g(q(t))) \mid q(t) = q] \\
= \;& 2\gamma \overline{g}(q)^T (u^* - q) + \gamma^2 E[\|g(q(t))\|_2^2 \mid q(t) = q] \\
\leq \;& 2\gamma \overline{g}(q)^T (u^* - q) + \gamma^2 G^2,
\end{aligned}$$

where G is the upper bound of the norm of $g(q(t))$, and

$$\overline{g}(q) = \overline{c}(q) - x(q) \quad \text{with} \quad \overline{c}(q) \in \arg\max_{c\in\overline{\Pi}} q^T c\ . \tag{13}$$

It is easy to check that $\overline{g}(q)$ is a subgradient[5] of the dual function $D(q)$ at point q, thus

$$\overline{g}(q)^T (u^* - q) \leq D(u^*) - D(q)\ .$$

So,

$$E[\Delta V_t(q)|q] \quad \leq \quad 2\gamma(D(u^*) - D(q)) + \gamma^2 G^2.$$

Note that $D(q)$ is a continuous function. Let

$$\delta = \max_{D(q)-D(u^*)\leq \gamma G^2} \|q - u^*\|_2$$

and define $\mathcal{A} = \{q:\ \|q - u^*\|_2 \leq \delta\}$. We can get

$$E[\Delta V_t(q)|q] \quad \leq \quad -\gamma^2 G^2 \mathcal{I}_{q\in\mathcal{A}^c} + \gamma^2 G^2 \mathcal{I}_{q\in\mathcal{A}}\ ,$$

where $\mathcal{I}$ is the index function. Thus, by *Theorem 3.1* in [26], which is a trivial extension of Foster's criterion for irreducible chain [3], the Markov chain $q(t)$ is stable. Since the M-queue length $p(t) = \gamma q(t)$, the Markov chain $p(t)$ is also stable. ■

The above proof shows that the distance to the optimal u^* has negative conditional mean drift for all scaled queue-length that have sufficiently large distance to u^*, and implies that the scaled queue-length will stay near u^* when γ is small enough.

Remark 2: We can make the Markov chain $p(t)$ irreducible over its state space, by making the arrival $x(t)$ a random variable with mean $\min\{U_s'^{-1}(\gamma p_s(t)), \alpha_s\}$, as that assumed in [9]. We can also make the system to reach a specific state infinitely often with finite mean recurrence times, which will ensure that the system reduces to one recurrent class whatever the initial state is.

[5]Given a convex function $f\ : \mathcal{R}^n \mapsto \mathcal{R}$, a vector $d \in \mathcal{R}^n$ is a subgradient of f at a point $u \in \mathcal{R}^n$ if $f(v) \geq f(u) + (v-u)^T d,\ v \in \mathcal{R}^n$ [21], [6].

C. Asymptotic Optimality and Fairness

In this subsection, we will prove the asymptotic optimality of the dual scheduling algorithm in terms of dual and primal functions of the reference system problem (8)-(9).

Theorem 4: The dual scheduling algorithm (4)-(7) converges statistically to within a small neighborhood of the optimal value $D(u^*)$, i.e.,

$$D(u^*) \leq D(E[q(\infty)]) \leq D(u^*) + \gamma G^2/2, \tag{14}$$

where $q(\infty)$ is a notation used to denote the state of the Markov chain $q(t)$ in the steady state.

Proof: The £rst inequality $D(u^*) \leq D(p)$ always holds, since $D(u^*)$ is the minimum of the dual function $D(u)$.

Now we prove the second inequality. From the proof of Theorem 3, we have

$$\begin{aligned} E[\Delta V_t(q)|q] &= E[V(q(t+1)) - V(q(t)) \mid q(t) = q] \\ &\leq 2\gamma(D(u^*) - D(q)) + \gamma^2 G^2. \end{aligned}$$

Taking expectation over q, we get

$$\begin{aligned} E[\Delta V_t(q)] &= E[V(q(t+1)) - V(q(t))] \\ &\leq 2\gamma(D(u^*) - E[D(q)]) + \gamma^2 G^2. \end{aligned}$$

Taking summation from $\tau = 0$ to $\tau = t-1$, we obtain

$$\begin{aligned} E[V(q(t))] \leq\ & E[V(q(0))] - 2\gamma \sum_{\tau=0}^{t-1} E[D(q(\tau))] \\ & + 2\gamma t D(u^*) + t\gamma^2 G^2. \end{aligned}$$

Since $E[V(q(t))] \geq 0$, we have

$$2\gamma \sum_{\tau=0}^{t-1} E[D(q(\tau))] - 2\gamma t D(u^*) \leq E[V(q(0))] + t\gamma^2 G^2.$$

From this inequality we obtain

$$\frac{1}{t}\sum_{\tau=0}^{t-1} E[D(q(\tau))] - D(u^*) \leq \frac{E[V(q(0))] + t\gamma^2 G^2}{2t\gamma}.$$

Note that $q(t)$ is stationary and ergodic in some steady state by Theorem 3, and so is $D(q(t))$. Thus,

$$\lim_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} E[D(q(\tau))] = E[D(q(\infty))].$$

So,

$$E[D(q(\infty))] - D(u^*) \leq \gamma G^2/2.$$

Since $D(q)$ is a convex function, by Jensen's inequality,

$$D(E[q(\infty)]) - D(u^*) \leq \gamma G^2/2,$$

i.e., the algorithm converges statistically to within $\gamma G^2/2$ of the optimal value $D(u^*)$. ∎

Since $D(q)$ is a continuous function, Theorem 4 implies that the scaled queue-length q approaches u^* statistically when γ is small enough.

Corollary 5: $x(t)$ is a stable Markov chain. Moreover, the average arrival rates $E[x(\infty)] \in \overline{\Pi}$, where $x(\infty)$ denotes the state of the process $x(t)$ in the steady state.

Proof: $x(t)$ is a deterministic, £nite-value function of $q(t)$. $x(t)$ is a stable Markov chain, since $q(t)$ is. $E[x(\infty)] \in \overline{\Pi}$, otherwise the average scaled queue-length $E[q(\infty)]$ will go unbounded, which contradicts to Theorem 4. ∎

Theorem 6: Let $P(x)$ be the primal function of the reference system problem (8)-(9). The dual scheduling algorithm (4)-(7) converges statistically to within a small neighborhood of the optimal value $P(x^*)$, i.e.,

$$P(x^*) \geq P(E[x(\infty)]) \geq P(x^*) - \frac{\gamma G^2}{2}. \tag{15}$$

Proof: The £rst inequality $P(x^*) \geq P(E[x(\infty)])$ holds, since $E[x(\infty)] \in \overline{\Pi}$.

Now we prove the second inequality. By equation (5), we have

$$\begin{aligned} & E[||q(t+1)||_2^2 | q(t)] \\ =\ & E[||[q(t) - \gamma g(q(t))]^+||_2^2 | q(t)] \\ \leq\ & E[||q(t) - \gamma g(q(t))||_2^2 | q(t)] \\ =\ & ||q(t)||_2^2 - 2\gamma \overline{g}(q(t))^T q(t) + \gamma^2 E[||g(q(t))||_2^2 | q(t)] \\ =\ & ||q(t)||_2^2 + 2\gamma \sum_s U_s(x_s(t)) \\ & -2\gamma \sum_s (U_s(x_s(t)) - q_s(t)x_s(t)) \\ & -2\gamma \sum_s q_s(t)\overline{c}_s(t) + \gamma^2 E[||g(p(t))||_2^2 | q(t)] \\ \leq\ & ||q(t)||_2^2 + 2\gamma \sum_s U_s(x_s(t)) \\ & -2\gamma \sum_s (U_s(x_s^*) - q_s(t)x_s^*) \\ & -2\gamma \sum_s q_s(t)\overline{c}_s(t) + \gamma^2 E[||g(p(t))||_2^2 | q(t)] \\ =\ & ||q(t)||_2^2 + 2\gamma P(x(t)) - 2\gamma P(x^*) \\ & -2\gamma \sum_s q_s(t)(\overline{c}_s(t) - x_s^*) + \gamma^2 E[||g(q(t))||_2^2 | q(t)] \\ \leq\ & ||q(t)||_2^2 + 2\gamma P(x(t)) - 2\gamma P(x^*) \\ & +\gamma^2 E[||g(q(t))||_2^2 | q(t)] \\ \leq\ & ||q(t)||_2^2 + 2\gamma P(x(t)) - 2\gamma P(x^*) + \gamma^2 G^2, \end{aligned}$$

where $\overline{g}(q)$ is de£ned as in equation (13), the second inequality follows from the fact that $x_s(t)$ is the maximizer of $\max_{x_s}(U_s(x_s) - q_s x_s)$, and the third inequality follows from the fact that $\overline{c}(t)$ is the maximizer in equation (13) and $x^* \in \overline{\Pi}$.

Taking expectation over q, we get

$$\begin{aligned} E[||q(t+1)||_2^2] \leq\ & E[||q(t)||_2^2] + 2\gamma E[P(x(t))] \\ & -2\gamma P(x^*) + \gamma^2 G^2. \end{aligned}$$

Applying the inequalities recursively, we obtain

$$\begin{aligned} E[||p(t)||_2^2] \leq\ & E[||p(0)||_2^2] + 2\gamma \sum_{\tau=0}^{t-1} (E[P(x(\tau))] \\ & - P(x^*)) + t\gamma^2 G^2. \end{aligned}$$

Since $E[||p(t)||_2^2] \geq 0$, we have

$$2\gamma \sum_{\tau=0}^{t-1} (E[P(x(\tau))] - P(x^*)) \geq -E[||p(0)||_2^2] - t\gamma^2 G^2.$$

From this inequality we obtain

$$\frac{1}{t}\sum_{\tau=0}^{t-1} E[P(x(\tau))] - P(x^*) \geq \frac{-E[||p(0)||_2^2] - t\gamma^2 G^2}{2t\gamma}.$$

Note that $x(t)$ is stationary and ergodic in some steady state by Corollary 5. Thus,

$$\lim_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} E[P(x(\tau))] = E[P(x(\infty))].$$

So,

$$E[(P(x(\infty))] - P(x^*) \geq -\frac{\gamma G^2}{2}.$$

Since P is a concave function, by Jenson's inequality,

$$P(E[x(\infty)]) - P(x^*) \geq -\frac{\gamma G^2}{2},$$

i.e., the algorithm converges statistically to within $\gamma G^2/2$ of the optimal value $P(x^*)$. ■

Since $P(x)$ is a continuous function, Theorems 6 implies that the average arrival rates to the M-queues approaches the optimal of the ideal reference system (8)-(9) when γ is small enough. Note that, when the M-queue length is stable, x will be the service rates offered to the users. Theorem 4 and 6 shows that, surprisingly, the vector of average service rates offered by the dual scheduling algorithm (4)-(7) approximately solves the ideal reference system problem, which is to maximize the aggregate concave utilities over the best feasible rate region that the network can support.

As is well-known, the fairness objectives can be achieved by appropriately choosing the concave objective functions. So, the asymptotic optimality establishes the fairness properties of the dual scheduling algorithm. For example, if we choose logarithm utility function $U_s(x_s) = \log(x_s)$, the dual scheduler will achieve proportional fairness [11], [9].

Note that the above proofs for stability and performance bounds are rather general. They only use general properties of convex/concave function, subgradient and convex set, and represent an integration of convex optimization and stochastic control. These stability and optimality results can be readily extended to other systems.

V. A New Scheduling Architecture

The dual scheduling algorithm motivates a new architecture for scheduling in the generalized switch (please see Fig.1 for a pictorial depiction). In this new architecture, a queue, termed M-queue, is introduced to interface the user data queue and the time-varying server. Data will depart from the data queue to enter the M-queue, and the generalized switch serves directly the M-queue. Through controlling the arrival process to the M-queue (or the departure process from the data queue), we can modulate the scheduling process, in order to achieve different performance objectives such as fairness, and maximum throughput, etc.

The dual scheduler would not incur much additional complexity. M-queues are distributed at each user, and can be "virtual" or be implemented as physical queues. The control of the M-queue arrival process is also distributed at each user and depends on only the "local" queue length of each user. The dual scheduler provides some advantages over other scheduling algorithms. For example, in the cellular network in the downlink, even though the primary scheduling algorithm can achieve fair resource allocation, it requires to estimate the average throughput of the users, while with the dual algorithm we only need to simply measure the M-queue length. Also, the dynamics of the M-queue ($p(t)$ and $x(t)$) is feedback-controlled, and thus will be relatively smooth, comparing with the dynamics of the switch. So, the dual scheduler can provide a relatively reliable and smooth service to the users, and can behave as a good interface between higher layer protocols and the scheduling at the link layer and ensure a better performance of the higher-layer protocols such as that of TCP congestion control.

To provide Quality of Service in generalized switches is a dif£cult problem. In the context of wireless networks, the interdependence of wireless links in combination with the time-varying nature of wireless channel makes QoS scheduling fairly challenging and the available results are mostly on stability guarantees. In the context of input-queued cross-bar switch, input buffering makes scalable switch design possible but makes QoS guarantees very challenging and again most available results are on maximizing the throughput. The dual scheduler might be promising in providing QoS in generalized switches, through carefully designing the M-queue arrival process. Further study is needed on related issues.

VI. Conclusions

In this paper, we consider the dual scheduling algorithm for a generalized switch. For a saturated system, we prove the asymptotic optimality of the dual scheduling algorithm and thus establish its fairness properties. The dual scheduling algorithm motivates a new architecture for scheduling, in which an additional queue is introduced to interface the user data queue and the time-varying server and to modulate the scheduling process, so as to achieve different performance objectives. Further research stemming out of this article includes scheduling with Quality of Service guarantees with the dual scheduler, and its application and implementation in various versions of the generalized switch model.

In the context of cellular networks, the dual scheduling algorithms provide an alternative to the primal scheduling algorithms such as the Proportional Fair Scheduler of Qualcomm High Data Rate system. We will also study user-perceived performance of the dual scheduling algorithm in future work. As the number of data ¤ows (and even the number of users) in progress is highly dynamic, increasing at the instants of some arrival process and decreasing as the transfer of £nite size

data is completed, there exists a strong interaction between the stochastic process describing the number of data ¤ows in progress and the way in which different users are served. This motivates us to further study user-perceived performance, such as cell capacity, transfer delay and transfer time, and blocking rate, etc, in the context of cellular network in the downlink.

ACKNOWLEDGMENTS

The authors would like to thank Felisa Vazquez-Abad for helpful discussions, and the anonymous reviewers for helpful comments and pointing out related references.

REFERENCES

[1] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar and P. Whiting, Providing quality of services over shared wireless link, *IEEE Communication Magazine*, Feburary 2001.

[2] M. Andrews, Instability of the proportional fair scheduling algorithm for HDR, *IEEE Transactions on Wireless Communications*, **3**(5):104-116, 2004.

[3] S. Asmussen, *Applied Probability and Queues*, 2nd edition, Springer, 2003.

[4] A. Bar-Noy, A. Mayer, B. Schieber and M. Sudan, Guaranteeing fair service to persistent dependent tasks, *SIAM J. COMPUT.*, **27**(4):1168-1189, August 1998.

[5] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana and A. Viterbi, CDMA/HDR: a bandwidth-ef£cient high-speed wireless data service for nomadic users, *IEEE Communication Magazine*, **38**(7):70-77, 2000.

[6] D. Bertsekas, *Nonlinear Programming*, 2nd edition, Athena Scienti£c, 1999.

[7] S. Borst, User-level performance of channel-aware scheduling algorithm in wireless data networks, *Proc. IEEE Infocom*, 2003.

[8] A. Eryilmaz, R. Srikant and J. Perkins, Stable scheduling policies for fading wireless channels, *IEEE/ACM Transactions on Networking*, 2005.

[9] A. Eryilmaz and R. Srikant, Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control, *Proc. IEEE Infocom*, 2005. Submitted for journal publication.

[10] A. Jalali, R. Padovani and R. Pankaj, Data-throughput of CDMA-HDR a high ef£ciency data rate personal communication wireless system, *Proc. 50th IEEE VTC*, 2000.

[11] F. P. Kelly, Charging and rate control for elastic traf£c, *European Transactions on Telecommunications*, **8**:33-37, 1997.

[12] F. P. Kelly, A. K. Maulloo and D. K. H. Tan, Rate control for communication networks: Shadow prices, proportional fairness and stability, *Journal of Operations Research Society*, **49**(3):237-252, March 1998.

[13] H. J. Kushner and P. A. Whiting, Convergence of proportional-fair sharing algorithms under general conditions, *IEEE/ACM Transactions on Wireless Communications*, **3**(4):1250-1259, 2004.

[14] X. Liu, E. Chong and N. Shroff, Opportunistic transmission scheduling with resource-sharing constraints in wireless networks, *IEEE J. Sel. Area Comm.*, **19**(10):2053-2064, 2001.

[15] Y. Liu and E. Knightly, Opportunistic fair scheduling over multiple wireless channels, *Proc. IEEE Infocom*, April 2003.

[16] S. H. Low and D. E. Lapsley, Optimal ¤ow control I: Basic algorithm and convergence, *IEEE/ACM Transactions on Networking*, **7**(6):861-874, December 1999.

[17] N. McKeown, V. Anantharam and J. Walrand, Achieving 100% throughput in an inputed-queued switch, *Proc. IEEE Infocom*, 1996.

[18] M. Neely, E. Modiano and C. Rohrs, Dynamic power allocation and routing for time varying wireless networks, *Proc. IEEE Infocom*, April 2003.

[19] N. Neely, Dynamic power allocation and routing for satellite and wireless networks with time varying Channels, Ph.D. Thesis, November 2003.

[20] S. Shakkottai and A. Stolyar, Scheduling for multiple ¤ows sharing a time-varying channel: the exponential rules, *Transactions of the AMS*, Series 2, A volume in memory of F. Karpelevich, 2002.

[21] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, 1985.

[22] A. L. Stolyar, MaxWeight scheduling in a generalized switch: state space collapse and workload minimization in heavy traf£c, *Ann. Appl. probab.*, **14**(1):1-53, 2004.

[23] A. L. Stolyar, On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation, *Operation Research*, **53**(1):12-25, 2005.

[24] A. L. Stolyar, Maximizing queueing network utility subject to statbility: greedy primal-dual algorithm, *Queueing Systems*, 50(4):401-457, 2005.

[25] V. Subramanian and R. Agrawal, A stochastic approximation analysis of channel condition aware wireless scheduling algorithms, *Proc. INFORMS Telecommun. Conf.*, 2002.

[26] L. Tassiulas and A. Ephremides, Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks, *IEEE Transactions on Automatic Control*, **37**(12):1936-1948, December 1992.

[27] L. Tassiulas and A. Ephremides, Dynamic server allocation to parallel queues with randomly varying connectivity, *IEEE Transactions on Information Theory*, **39**:466-478, 1993.

[28] L. Tassiulas, Scheduling and performance limits of networks with constantly changing topology, *IEEE Transactions on Information Theory*, **43**(3):1067-1073, 1997.

[29] D. N. Tse, Multiuser diversity and proportionally fair scheduling, *In preparation*, 2004.