Extending the Power of Backpressure to Correlated Traffic

Jeongho Jeon and Anthony Ephremides

Department of Electrical and Computer Engineering and Institute for Systems Research
University of Maryland, College Park, MD 20742
Email: {jeongho, etony}@umd.edu

***Abstract*—We study the impact of time-correlated arrivals on the performance of backpressure policy for stochastic network control. The arrival process considered in this work is fairly general in the sense that it may exhibit short/long-range dependence depending on the asymptotic shape of the autocorrelation function. In this paper, we show that that the backpressure policy stabilizes the network whenever the arrival rate vector is inside the stability region even though the arrivals have finite-length memory or infinite-length memory with monotonically decreasing autocorrelation functions. Apart from the stability, the effect of correlations appears in the upper bound on average network delay.**

I. Introduction

The stability region of constrained queueing systems was introduced in [1] which is defined as the set of arrival rate vectors for which the queues in the network reach their steady states. A maximum throughput policy which supports the entire stability region except at most at points on the boundary was proposed in the paper as well. The maximum throughput policy is comprised of two parts: maximum differential backlog routing and max-weight scheduling. Since the policy selects paths dynamically according to the differential backlog between neighboring nodes, it is often referred to as *backpressure* policy.

Despite the significance of the original work [1] and its extensions [2]–[8] on stochastic network control, one weakness might be the fact that they were derived under the assumption that each arrival process had independent and identically distributed (i.i.d.) number of arrivals in each time slot. Several measurement-based studies have pointed out, however, that scale-invariant burstiness, i.e., self-similarity, exists in local/wide-area network and Internet traffic [9]–[11]. Possible explanation includes heavy-tailed file size distribution, human interactions, and protocol-level dynamics. It was also shown that the variable-bit-rate video traffic is long-range dependent [12]. Self-similarity and long-range dependence are two distinct concepts and, hence, one does not necessarily imply the other. However, when both are viewed at the asymptotically large scale, they are identical [13].

In this paper, we show that the network is stabilized by the backpressure policy whenever the arrival rate vector is strictly inside the stability region even when the arrivals are possibly short/long-range dependent. To the best of our knowledge, there is no existing work in this context. In [14], the stability region of the finite-user slotted ALOHA system was studied under the assumption of a specific correlated arrival distribution. In [15], delay performance of max-weight scheduling was studied for both single-hop and multihop networks with two-state Markov modulated arrival processes that are certainly short-range dependent. The arrival process considered in [16] is as general as ours but was considered in the context of maximal scheduling for single-hop networks.

This work is supported by MURI grant W911NF-08-1-0238, NSF grant CCF-0728966, and ONR grant N000141110127.

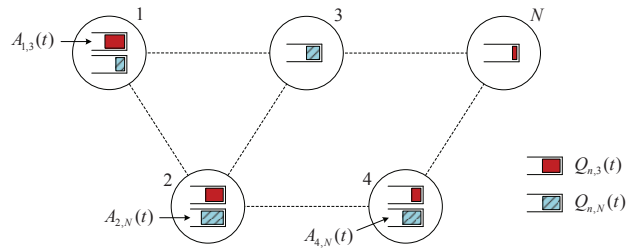

Fig. 1. Wireless multihop network with multiple input streams and per-destination queueing

The rest of the paper is organized as follows. In Section II, we provide descriptions on network model and time-correlated arrivals. In Section III, we present the main result on the backpressure policy with time-correlated arrivals. Finally, we draw conclusions in Section IV.

II. Network Model

We consider a time-slotted multihop wireless network with N nodes and L directed links. An example network topology is shown in Fig. 1. We denote by $\mathcal{N}$ and $\mathcal{L}$ the set of nodes and links, respectively. Let $A_{n,c}(t)$ represent the exogenously arriving amount of data at node n during time slot t which is destined to node c in units of bits/slot. At each node, all the exogenous and endogenous arrivals due to multihop relaying are classified and queued according to their destinations. We let $Q_{n,c}(t)$ be the backlog of destination c data that is awaiting transmission in node n at time slot t in units of bits and define $\overrightarrow{Q}(t) \triangleq (Q_{n,c}(t))$. It is assumed that the system starts with empty queues, i.e., $Q_{n,c}(0) = 0$ for all

(n, c) pairs. Denote by $\mu_l(t)$ and $\mu_{l,c}(t)$ the transmission rate over link l during time slot t and the amount of rate that is offered to destination c traffic among $\mu_l(t)$ in units of bits/slot, respectively. Consequently, $\sum_c \mu_{l,c}(t) \leq \mu_l(t)$ for all l. Denote by Γ_t the feasible region of link transmission rate vector $\overrightarrow{\mu}(t) \triangleq (\mu_l(t))$ at time slot t. In each time slot, a control policy chooses $\overrightarrow{\mu}(t)$ from the constrained set Γ_t and allocates the rate on each link to the traffics of each destination.

In this work, the arrival process $A_{n,c}(t)$ does not necessarily to be i.i.d. but only needs to be wide-sense stationary (WSS); that is a process with constant mean $\lambda_{n,c}$, finite variance $\sigma^2_{n,c}$, and an autocorrelation function $\rho_{n,c}(k) \triangleq E[(A_{n,c}(t) - \lambda_{n,c})(A_{n,c}(t-k) - \lambda_{n,c})]/\sigma^2_{n,c}$ that depends only on the time-lag between two samples. Denote by $H(t)$ the past history of all arrivals up to but not including time slot t. If there exists finite number $T \geq 0$ such that $E[A_{n,c}(t)|H(t-k)] = \lambda_{n,c}$ for $k \geq T$, then $\rho_{n,c}(k) = 0$ for $k > T$. Stochastic processes that fall in this category are said to have finite-length memory; an i.i.d. process is a special case with $T = 0$. If there is no such finite T, we further divide the processes according to the asymptotic shape of autocorrelation function $\rho_{n,c}(k)$. For most of the stochastic models including autoregressive moving average processes and Markov modulated processes, the autocorrelation functions are characterized by an exponential decay, i.e., $\rho_{n,c}(k) \sim \alpha^k$ as $k \to \infty$, where $0 < \alpha < 1$. The exponential tail of the function implies $\sum_k \rho_{n,c}(k) < \infty$. Stochastic processes belonging to this category are said to be short-range dependent. On the other hand, long-range dependent processes are characterized by a power-law decay of autocorrelation function, $\rho_{n,c}(k) \sim k^{-\beta}$ as $k \to \infty$, where $0 < \beta < 1$. As a result, the autocorrelation function is nonsummable, i.e., $\sum_k \rho_{n,c}(k) = \infty$, which implies that while high-lag correlations are individually small, their cumulative effect gives rise to features which are drastically different from those of short-range dependent processes [12]. Examples of long-range dependent processes are fractional Brownian motion and its discrete-time analog, fractional Gaussian noise.

The degree of self-similarity of a series is expressed using a single parameter H, called Hurst parameter. Self-similar processes and long-range dependent processes are related through the relation $H = 1 - \beta/2$; for a self-similar process with long-range dependence, we have $1/2 < H < 1$. Nevertheless, they are two different concepts; self-similarity involves all scales whereas long-range dependence only involves asymptotically large scale. Yet, if we are only interested in the steady-state performance measures such as average network delay, the role of self-similar input traffic is not different from that of long-range dependent traffic.

III. Network with Stabilizable Arrival Rates

Define $\mathcal{I}_n$ and $\mathcal{O}_n$ as the set of incoming and outgoing links of node n, i.e., $\{\, l : \mathrm{rx}(l) = n \,\}$ and $\{\, l : \mathrm{tx}(l) = n \,\}$, where $\mathrm{tx}(l)$ and $\mathrm{rx}(l)$ are the transmitting and receiving nodes of link l, respectively. Then, the queue length process at node n for destination c data evolves as

$$Q_{n,c}(t+1) \leq \max\left[Q_{n,c}(t) - \sum_{l \in \mathcal{O}_n} \mu_{l,c}(t), 0\right] + \sum_{l \in \mathcal{I}_n} \mu_{l,c}(t) + A_{n,c}(t) \quad (1)$$

which is an inequality rather than an equality because the actual endogenous arrivals may be less than the allocated link rates if the corresponding transmitters do not have enough data. The stability region Λ of a system is defined as the set of arrival rate vectors $\overrightarrow{\lambda} \triangleq (\lambda_{n,c})$ for which all queues in the network are stable[1] by considering all the policies [1]. It is well-known that the following backpressure policy stabilizes a network if the arrival rate vector is inside the stability region and each arrival process has i.i.d. number of arrivals in each time slot [1], [3].

Algorithm 1: Backpressure Policy

- Differential Backlog Routing/Scheduling: The differential backlog of destination c data over link l is defined as $D_{l,c}(t) \triangleq Q_{\mathrm{tx}(l),c}(t) - Q_{\mathrm{rx}(l),c}(t)$. If the link is directly connected to the destination, i.e, $\mathrm{rx}(l) = c$, $D_{l,c}(t) \triangleq Q_{\mathrm{tx}(l),c}(t)$. The maximum differential backlog over link l is obtained as $D_l(t) = \max_c D_{l,c}(t)$ and the maximizing destination c^* data is chosen for potential transmission over link l.
- Max-Weight Rate Allocation: The link transmission rate vector $\vec{\mu}(t)$ is selected from feasible region Γ_t to maximize the weighted sum rate as

$$\arg\max_{\vec{\mu}(t) \in \Gamma_t} \sum_{l \in \mathcal{L}} D_l(t)\mu_l(t)$$

  where the weight of each link corresponds to the maximum differential backlog of the link.

Note that the policy lets the maximizing destination c^* data use all the rate allocated to the link. Therefore, in each time slot, $\mu_{l,c^*}(t) = \mu_l(t)$ and $\mu_{l,c}(t) = 0$ for all $c \neq c^*$. Before we proceed to present the performance of the backpressure policy with time-correlated arrivals, let us first describe the following constants. Denote by $A_{\max}$ the maximum amount of arrivals to any node in any slot, i.e., $\sum_c A_{n,c}(t) \leq A_{\max}$. Further define $\mu^{\mathrm{out}}_{\max}$ and $\mu^{\mathrm{in}}_{\max}$ as the maximum transmission rate out of, and into, any node for all time slots as $\sum_{l \in \mathcal{O}_n} \mu_l(t) \leq \mu^{\mathrm{out}}_{\max}$ and $\sum_{l \in \mathcal{I}_n} \mu_l(t) \leq \mu^{\mathrm{in}}_{\max}$, respectively.

Theorem 3.1: If arrival rate vector $\overrightarrow{\lambda}$ is strictly interior to stability region Λ and there exists finite integer $T > 0$ such that $E[A_{n,c}(t)|H(t-k)] = \lambda_{n,c}$ for $k \geq T$ and for all (n, c) pairs, the backpressure policy stabilizes the network and guarantees bounded average queue backlog as

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{n,c} E[Q_{n,c}(\tau)] \leq \frac{NB_1 + 2\sum_{n,c} \phi_{n,c}(T)}{2\nu_{\max}}$$

[1]A queue is said to be stable if it reaches a steady state and do not drift to infinity.

where

$$\phi_{n,c}(T) \triangleq \sigma_{n,c}^2 \sum_{k=1}^{T} \rho_{n,c}(k) + T\lambda_{n,c}\left(\mu_{\max}^{\text{in}} + \lambda_{n,c}\right)$$

and $B_1 \triangleq (\mu_{\max}^{\text{in}} + A_{\max})^2 + (\mu_{\max}^{\text{out}})^2$ and $\nu_{\max}$ is defined as the maximum of ν such that $\overrightarrow{\lambda} + \nu\overrightarrow{1} \in \Lambda$ where $\overrightarrow{1}$ is a vector whose cardinality is the same with that of $\overrightarrow{\lambda}$ and elements are all one.

Proof: Define $L(\overrightarrow{Q}(t)) \triangleq \sum_{n,c} Q_{n,c}(t)^2$ as a Lyapunov function for scalar measure of the network congestion whose conditional drift given the current queue backlog is defined by

$$\Delta(\overrightarrow{Q}(t)) \triangleq E[L(\overrightarrow{Q}(t+1)) - L(\overrightarrow{Q}(t))|\overrightarrow{Q}(t)]$$

From the queueing dynamics in (1), it can be verified that [3]

$$\Delta(\overrightarrow{Q}(t)) \leq NB_1 + 2\sum_{n,c} Q_{n,c}(t) E\left[A_{n,c}(t)|\,\overrightarrow{Q}(t)\right] - 2\sum_{n,c} Q_{n,c}(t) E\left[\sum_{l\in\mathcal{O}_n}\mu_{l,c}(t) - \sum_{l\in\mathcal{I}_n}\mu_{l,c}(t)\middle|\overrightarrow{Q}(t)\right] \quad (2)$$

Note that the control variables on the right-hand side (RHS) of the drift expression are link transmission rates $\mu_{l,c}(t)$. By simply converting the corresponding terms containing control variables from node-centric to link-centric expressions, we obtain the following identity:

$$\sum_{n,c} Q_{n,c}(t)\left(\sum_{l\in\mathcal{O}_n}\mu_{l,c}(t) - \sum_{l\in\mathcal{I}_n}\mu_{l,c}(t)\right) = \sum_{l}\sum_{c}\mu_{l,c}(t)\left(Q_{\text{tx}(l),c}(t) - Q_{\text{rx}(l),c}(t)\right) \quad (3)$$

which reveals the rationale behind the design of backpressure policy; it is aimed to minimize the RHS of (2) by maximizing (3). The original proof in [1] showed that for sufficiently large sum of queue backlogs, the RHS of (2) under the backpressure policy becomes negative which corresponds to the Forster's criteria for stability of irreducible Markov chains [17]. In [2]–[4], the Lyapunov drift technique is extended so that an explicit upper bound on the sum of average queue backlogs can be obtained which suffices for the stability proof as well. We follow similar procedures here. Note that if the arrival rate vector is strictly inside the stability region, there must exist some constant $\nu > 0$ such that $\overrightarrow{\lambda} + \nu\overrightarrow{1} \in \Lambda$. From Corollary 3.9 of [3], we know that there exists a stationary randomized policy that makes decision based only on the current channel states and independent of queue backlogs such that

$$E\left[\sum_{l\in\mathcal{O}_n}\mu_{l,c}(t) - \sum_{l\in\mathcal{I}_n}\mu_{l,c}(t)|\overrightarrow{Q}(t)\right] = \lambda_{n,c} + \nu$$

Since the stationary policy is simply a particular rate allocation policy, the performance under the backpressure policy is no worse than any other policies following the construction of the algorithm. Consequently, we have

$$\Delta(\overrightarrow{Q}(t)) \leq NB_1 + 2\sum_{n,c} Q_{n,c}(t) E\left[A_{n,c}(t)|\overrightarrow{Q}(t)\right] - 2\sum_{n,c}(\lambda_{n,c}+\nu)Q_{n,c}(t) \quad (4)$$

In the previous works assuming i.i.d. arrivals, the relation $E[A_{n,c}(t)|\overrightarrow{Q}(t)] = \lambda_{n,c}$ holds which simplifies the rest of the analysis. However, if arrivals are correlated, $Q_{n,c}(t)$ is also correlated through the queueing dynamics in (1) and $A_{n,c}(t)$ and $Q_{n,c}(t)$ are no longer independent of each other. We note from the queueing dynamics that

$$Q_{n,c}(t) \leq Q_{n,c}(t-T) + \sum_{k=1}^{T}\left(\sum_{l\in\mathcal{I}_n}\mu_{l,c}(t-k) + A_{n,c}(t-k)\right)$$

and, thus, it follows that

$$E[Q_{n,c}(t)A_{n,c}(t)] \leq E\left[Q_{n,c}(t-T)A_{n,c}(t)\right] + E\left[\sum_{k=1}^{T}\left(\sum_{l\in\mathcal{I}_n}\mu_{l,c}(t-k) + A_{n,c}(t-k)\right)A_{n,c}(t)\right] \quad (5)$$

Because all the arrival processes are assumed to have finite memory of at most length T, we have

$$E\left[Q_{n,c}(t-T)A_{n,c}(t)\right] = \lambda_{n,c}E\left[Q_{n,c}(t-T)\right] \quad (6)$$

For the remaining of the RHS of (5), we obtain

$$E\left[\sum_{k=1}^{T}\left(\sum_{l\in\mathcal{I}_n}\mu_{l,c}(t-k) + A_{n,c}(t-k)\right)A_{n,c}(t)\right] \leq \sigma_{n,c}^2\sum_{k=1}^{T}\rho_{n,c}(k) + T\lambda_{n,c}\left(\mu_{\max}^{\text{in}} + \lambda_{n,c}\right) \quad (7)$$

Plugging (6) and (7) into the RHS of (5) yields an upper bound on $E[Q_{n,c}(t)A_{n,c}(t)]$. Taking the expectation of (4) with respect to the distribution of queue backlogs and applying the bound yields the unconditional Lyapunov drift satisfying

$$E[L(\overrightarrow{Q}(t+1)) - L(\overrightarrow{Q}(t))] \leq NB_1 - 2\sum_{n,c}(\lambda_{n,c}+\nu)E\left[Q_{n,c}(t)\right] + 2\sum_{n,c}\lambda_{n,c}E\left[Q_{n,c}(t-T)\right] + 2\sum_{n,c}\left(\sigma_{n,c}^2\sum_{k=1}^{T}\rho_{n,c}(k) + T\lambda_{n,c}\left(\mu_{\max}^{\text{in}} + \lambda_{n,c}\right)\right)$$

Summing the inequality over $t \in \{0, \ldots, M-1\}$ yields

$$\begin{aligned} E[L(\overrightarrow{Q}(M)) - L(\overrightarrow{Q}(0))] &\leq NMB_1 \\ &- 2\sum_{\tau=0}^{M-1}\sum_{n,c}(\lambda_{n,c}+\nu)E\left[Q_{n,c}(\tau)\right] \\ &+ 2\sum_{\tau=0}^{M-1}\sum_{n,c}\lambda_{n,c}E\left[Q_{n,c}(\tau-T)\right] \\ &+ 2M\sum_{n,c}\left(\sigma_{n,c}^2\sum_{k=1}^{T}\rho_{n,c}(k) + T\lambda_{n,c}\left(\mu_{\max}^{\text{in}}+\lambda_{n,c}\right)\right) \end{aligned}$$

where $E[Q_{n,c}(t)] = 0$ for $t \leq 0$. Dividing the above by $2M$, rearranging terms, and using the fact that the system starts with empty queues and the non-negativity of Lyapunov function, we obtain

$$\begin{aligned} \frac{1}{M}\sum_{\tau=0}^{M-1}\sum_{n,c}(\lambda_{n,c}+\nu)E\left[Q_{n,c}(\tau)\right] &\leq \frac{NB_1}{2} \\ &+ \frac{1}{M}\sum_{\tau=0}^{M-1}\sum_{n,c}\lambda_{n,c}E\left[Q_{n,c}(\tau-T)\right] \\ &+ \sum_{n,c}\left(\sigma_{n,c}^2\sum_{k=1}^{T}\rho_{n,c}(k) + T\lambda_{n,c}\left(\mu_{\max}^{\text{in}}+\lambda_{n,c}\right)\right) \end{aligned}$$

Taking a $\limsup$ as $M \to \infty$, and noting that

$$\limsup_{M\to\infty}\frac{1}{M}\sum_{\tau=0}^{M-1}E[Q_{n,c}(\tau-T)] = \limsup_{M\to\infty}\frac{1}{M}\sum_{\tau=0}^{M-1}E[Q_{n,c}(\tau)] \tag{8}$$

and optimizing over ν yields the result.

■

Theorem 3.1 guarantees the stability of the network under backpressure policy when the arrival rate vector is inside the stability region and the arrival processes are time-correlated over fixed length of interval. This is done by explicitly showing that the sum of average queue backlogs in the network is upper bounded by some finite number. However, if there is at least one long memory arrival such that the condition $E[A_{n,c}(t)|H(t-k)] = \lambda_{n,c}$ for $k \geq T$ is invalidated for any finite integer T, the theorem fails simply because the upper bound becomes infinite. In the case that all arrivals are at most short-range dependent, it might be argued that the theorem still holds because the correlation decreases exponentially fast as the time-lag increases. Consequently, for sufficiently large but finite T, the correlation becomes effectively zero. However, if there exist long-range dependent arrivals, we need a stronger argument to guarantee the stability of the network as in the following theorem which only requires very mild conditions on the arrival process such as the monotonicity of its absolute autocorrelation function.

Theorem 3.2: If arrival rate vector $\overrightarrow{\lambda}$ is strictly interior to stability region Λ and the absolute autocorrelation functions $|\rho_{nc}(k)|$ of the arrival processes are monotonically decreasing for all (n,c) pairs, the backpressure policy stabilizes the network and guarantees bounded average queue backlog as

$$\limsup_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\sum_{n,c}E[Q_{n,c}(\tau)] \leq \frac{NB_1 + 2\sum_{n,c}\phi_{n,c}(T_\delta)}{2(\nu_{\max}-\delta)}$$

where δ is an arbitrary constant satisfying $0 < \delta < \nu_{\max}$ and T_δ is the minimum of T such that $|E[A_{n,c}(t)|H(t-k)] - \lambda_{n,c}| \leq \delta$ for $k \geq T$ and for all (n,c) pairs.

Proof: As in the proof of Theorem 3.1, the main difficulty of analyzing the Lyapunov drift is due to the correlation between $Q_{n,c}(t)$ and $A_{n,c}(t)$. Assume that δ and corresponding T_δ are chosen such that the conditions described in the theorem are met. Then, for $T \geq T_\delta$, we have

$$E\left[Q_{n,c}(t-T)A_{n,c}(t)\right] \leq (\lambda_{n,c}+\delta)E\left[Q_{n,c}(t-T)\right]$$

for all (n,c) pairs and inequality (7) holds for any $T > 0$. Therefore, the unconditional Lyapunov drift satisfies

$$\begin{aligned} E[L(\overrightarrow{Q}(t+1)) - L(\overrightarrow{Q}(t))] &\leq NB_1 \\ -2\sum_{n,c}(\lambda_{n,c}+\nu)E\left[Q_{n,c}(t)\right] &+ 2\sum_{n,c}(\lambda_{n,c}+\delta)E\left[Q_{n,c}(t-T)\right] \\ &+ 2\sum_{n,c}\left(\sigma_{n,c}^2\sum_{k=1}^{T}\rho_{n,c}(k) + T\lambda_{n,c}\left(\mu_{\max}^{\text{in}}+\lambda_{n,c}\right)\right) \end{aligned}$$

for $T \geq T_\delta$. The rest of the proof is identical with that of Theorem 3.1. ■

Theorem 3.2 can be applied to the network containing any long memory arrivals whose absolute autocorrelation functions are monotonically decreasing. It is expected that long-range dependent arrivals would induce longer T_δ satisfying the condition when compared to short-range dependent arrivals due to the hyperbolic shape of the autocorrelation function. Once the autocorrelation functions are given, one can compute T_δ deterministically. From Theorem 3.1 and Theorem 3.2, we know that the backpressure policy stabilizes the network whenever the arrival rate vector is inside the stability region no matter whether arrival processes are i.i.d. or time-correlated. Specifically, in Theorem 3.1, the upper bound on the sum of average queue backlogs is always finite whenever $\nu_{\max} > 0$ which corresponds to the condition that the arrival rate vector is strictly inside the stability region. In Theorem 3.2, there must exist δ such that $0 < \delta < \nu_{\max}$ because $\nu_{\max} > 0$ and T_δ must be finite given δ because the autocorrelation functions are assumed to be monotonically decreasing. Therefore, the upper bound on the sum of average queue backlogs is always finite. Note that in case of i.i.d. arrivals, the upper bound is given by $NB_1/2\nu_{\max}$ [3]. Thus, apart from the stability issue, the upper bound on average network delay will increase when arrivals are time-correlated because the average network delay is proportional to average network backlog by Little's law [18].

IV. Concluding Remarks

The performance of backpressure policy with time-correlated arrivals was studied using the Lyapunov drift technique. It was shown that the backpressure policy stabilizes

the network whenever the arrival rate vector is inside the stability region no matter whether the arrivals are i.i.d. or time-correlated. The impact of input correlations are reflected in the upper bound on the average network delay. To handle the Lyapunov drift with correlated terms, we first expressed the evolution of the system dynamics over multiple slots and used the fact that the conditional expectation of the arrival process given the past history falls within arbitrarily small constant range around its unconditional expectation if the time-lag between the arrival process and the past history becomes sufficiently large. This is true for both finite memory and infinite memory arrivals with monotonically decreasing autocorrelation functions.

REFERENCES

[1] 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*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.

[2] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 89–103, Jan. 2005.

[3] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends in Networking*, vol. 1, no. 1, pp. 1–149, 2006.

[4] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," in *Proceedings of IEEE INFOCOM*, Mar. 2005.

[5] M. J. Neely, "Energy optimal control for time varying wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 2915–2934, Jul. 2006.

[6] ——, "Super-fast delay tradeoffs for utility optimal fair scheduling in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1489–1501, Aug. 2006.

[7] ——, "Order optimal delay for opportunistic scheduling in multi-user wireless uplinks and downlinks," in *Proceedings of the 44th Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2006.

[8] L. B. Le, E. Modiano, and N. B. Shroff, "Optimal control of wireless networks with finite buffers," in *Proceedings of IEEE INFOCOM*, 2010.

[9] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 1–15, Feb. 1994.

[10] V. Paxson and S. Floyd, "Wide-area traffic: The failure of poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3.

[11] M. E. Crovella and A. Bestavros, "Self-similar in world wide web traffic: Evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6.

[12] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in viriable-bit-rate video traffic," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4.

[13] G. Samorodnitsky, "Long range dependence," *Foundations and Trends in Stochastic Systems*, vol. 1, no. 3, pp. 163–257, 2007.

[14] V. Anantharam, "The stability region of the finite-user slotted aloha protocol," *IEEE Transactions on Information Theory*, vol. 37, no. 3, pp. 535–540, May 1991.

[15] L. B. Le, K. Jagannathan, and E. Modiano, "Delay analysis of maximum weight scheduling in wireless ad hoc networks," in *Proceedings of Conference on Information Sciences and Systems (CISS)*, 2009.

[16] M. J. Neely, "Delay analysis for maximal scheduling with flow control in wireless networks with bursty traffic," *IEEE/ACM Transactions on Networking*, vol. 17, no. 4, pp. 1146–1159, Aug. 2009.

[17] S. Asmussen, *Applied Probability and Queues*. New York: Springer-Velag, 2003.

[18] D. Bertsekas and R. Gallager, *Data Networks*. Saddle River, NJ: Prentice Hall, 1992.