Throughput Regions of Multiple-Access Fading Channels with Markov Arrivals and QoS Constraints

Mustafa Ozmen and M. Cenk Gursoy

***Abstract*—In this paper, throughput regions of multiple-access fading channels are characterized when multiple users, experiencing random data arrivals, transmit to a common receiver under statistical quality of service (QoS) constraints. Random arrivals are modeled as discrete Markov or Markov fluid processes. Different multiple-access transmission strategies, namely time-division multiple access and superposition coding with fixed or variable decoding orders, are considered. For these arrival and transmission models, throughput in terms of maximum average arrival rates is formulated in the presence of buffer constraints, employing the notions of effective bandwidth of time-varying sources and effective capacity of time-varying transmissions. Throughput regions are computed for the two-user case and the impact of source burstiness, buffer constraints, signal-to-noise ratio, and different communication strategies is investigated.**

***Index Terms*—Markov arrivals, multiple-access channel, quality of service, superposition coding, TDMA, throughput region.**

I. Introduction

MULTIPLE-ACCESS channel (MAC) model, in which multiple users share a communication medium to send their messages to a common receiver, is one of the main building blocks of multiple-user communication scenarios, modeling, for instance, uplink in cellular and satellite communications and wireless LANs. It is well-known that Gaussian MAC capacity region is achieved by having simultaneous transmissions from the users (i.e., superposition coding) with successive cancelation decoding at the receiver [1]. Similar transmission and reception strategies are optimal in multiple-access fading channels as well [2].

In this paper, we investigate the throughput regions of multiple-access fading channels when the users experience random arrivals and operate in the presence of quality-of-service (QoS) constraints. Consideration of QoS guarantees is motivated by the recent exponential growth of wireless transmissions of multimedia content. Indeed, with advances in the wireless technology and widespread use of social networking tools and video-sharing sites, multimedia data traffic (consisting of e.g., voice over IP (VoIP), multimedia streaming, interactive video) is poised to become the dominant traffic in mobile wireless communications. Such multimedia traffic requires certain QoS limitations (e.g., buffer/delay constraints) to be satisfied in order to provide required levels of performance and quality to the end-users. Effective bandwidth and effective capacity are employed in [3] and [4], respectively, as performance metrics to determine the maximum throughput in the presence of statistical QoS constraints in the form of limitations on buffer violation probability[1]. In [5] and [6], effective bandwidths of different source models are derived. Recently, in [7], we adopted the effective capacity to study the throughput regions of multiple-access fading channels under the assumption that arrival rates to all users are constant. In this paper, we consider a more general scenario in which users experience random Markov arrivals. In particular, we combine the theory of effective bandwidth of time-varying random arrivals and the theory of effective capacity of time-varying wireless transmissions in order to characterize the throughput regions in multiple-access fading channels.

II. Channel Model

We consider a multiple-access fading channel in which M users transmit to a common receiver. We assume that each user experiences Markov data arrivals. Randomly arriving data is initially buffered at each user before transmission over the multiple-access channel. For each random source traffic, certain statistical QoS constraints are imposed at each user in order to limit the buffer violation probability.

In the considered multiple-access channel, each link experiences flat-fading and the channel input-output relation can be expressed as

$$y = \sum_{i=1}^{M} h_i x_i + n \qquad (1)$$

where x_i is the channel input of the i^{th} user and y is the output at the receiver. Average transmitted signal energy of the i^{th} user is $\mathbb{E}\{|x_i|^2\} = \mathcal{E}_i$. Moreover, n denotes the zero-mean, circularly-symmetric, complex Gaussian background noise at the receiver with variance $\mathbb{E}\{|n|^2\} = N_0$. Hence, the signal-to-noise ratio (SNR) of the i^{th} user is $\text{SNR}_i = \frac{\mathbb{E}\{|x_i|^2\}}{\mathbb{E}\{|n|^2\}} = \frac{\mathcal{E}_i}{N_0}$ for $i = 1, ..., M$. Finally, in (1), h_i denotes the fading coefficient in the channel between the user i and the receiver. While fading coefficients can have arbitrary distributions with finite energies, we assume that block-fading is experienced. Hence, the realizations of the fading coefficients stay fixed for a block of symbols and change independently for the next block.

III. Preliminaries

A. *Effective Bandwidth of Markov Arrivals*

As noted above, we assume that the data to be sent is generated from Markov sources and is initially stored in a

Manuscript received April 18, 2013. The associate editor coordinating the review of this letter and approving it for publication was L. Badia.

The authors are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 (e-mail: {mozmen, mcgursoy}@syr.edu).

Digital Object Identifier 10.1109/WCL.2013.061913.130279

[1]We note that statistical QoS constraints considered here are generally more stringent than stability constraints for which the throughput region or more specifically the stability region is equal to the ergodic capacity region [8].

buffer before transmission. Statistical constraints are imposed on the buffer length. In particular, we assume that the buffer violation probability at user i satisfies

$$\lim_{\eta \to \infty} \frac{\log P(Q_i \geq \eta)}{\eta} = -\theta_i \quad (2)$$

where Q_i denotes the stationary queue length at the i^{th} user, and θ_i is the decay rate of the tail distribution of the queue length. The above limiting formula implies that for large $\eta_{\max}$, we have $P(Q_i \geq \eta_{\max}) \approx e^{-\theta_i \eta_{max}}$. Hence, for a sufficiently large threshold, the buffer violation probability should decay exponentially with rate controlled by the QoS exponent θ_i. Note that as θ_i increases, more strict queueing or QoS constraints are imposed.

In this paper, we consider two types of Markov sources, namely discrete Markov source and Markov fluid source, and concentrate on a simple two-state (ON-OFF) model. For both types of sources, we briefly describe below the effective bandwidth, which characterizes the minimum constant transmission (or service) rate required to support the given time-varying data arrivals while the buffer violation probability satisfies (2).

1) Discrete Markov Source: For this type of source, data arrivals are modeled as a discrete Markov process with a transition probability matrix $\mathbf{J}$. In the special case of a two-state (ON-OFF) Markov chain, we assume that r bits arrive (i.e., the arrival rate is r bits/block) in the ON state while there are no arrivals in the OFF state. With the state transition probability matrix

$$\mathbf{J} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}, \quad (3)$$

the effective bandwidth is given by [6]

$$a(\theta, r) = \frac{1}{\theta} \log_e \left(\frac{p_{11} + p_{22} e^{r\theta} + \sqrt{(p_{11} + p_{22} e^{r\theta})^2 - 4(p_{11} + p_{22} - 1) e^{r\theta}}}{2} \right) \quad (4)$$

where θ is the QoS exponent defined in (2), p_{11} denotes the probability of staying in the OFF state and p_{22} denotes the probability of staying in the ON state. The probabilities of transitioning from one state to a different one are therefore denoted by $p_{21} = 1 - p_{22}$ and $p_{12} = 1 - p_{11}$.

2) Markov Fluid Source: In the fluid model, we assume that the data arrivals are modeled as a continuous-time Markov process with a generating matrix $\mathbf{G}$. The generating matrix for the two-state case is in the form of

$$\mathbf{G} = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \quad (5)$$

where α and β are the transition rates from one state to another. When the arrival rates for two-state model are r and 0 and hence we basically have ON and OFF states, the effective bandwidth expression becomes [6]

$$a(\theta, r) = \frac{1}{2\theta} \left[\theta r - (\alpha + \beta) + \sqrt{(\theta r - (\alpha + \beta))^2 + 4\alpha\theta r} \right]. \quad (6)$$

B. Effective Capacity in Fading Channels

Effective capacity, as a dual concept to effective bandwidth, identifies the maximum constant arrival rate that can be supported by a given time-varying service process while satisfying (2). Under the block-fading assumption, the effective capacity can be expressed as [7]

$$C_E(\text{SNR}, \theta) = -\frac{1}{\theta} \log_e \mathbb{E}\{e^{-\theta R}\} \quad (7)$$

where R is the instantaneous transmission (or equivalently service) rate. For instance, the maximum service rate in a single-user fading Gaussian channel is given by the instantaneous channel capacity expressed as $R = \log_2(1 + \text{SNR} z)$ where $z = |h|^2$ is the magnitude-square of fading.

IV. MAC Throughput Region

In this section, we initially describe our throughput metric as the maximum average arrival rate that can be supported in a setting in which arrivals are modeled as ON-OFF Markov processes, service rates are given by the instantaneous channel capacities, and buffer overflow probabilities are limited as described in Section III-A. In particular, we formulate the maximum average arrival rates by using both effective bandwidth and effective capacity formulas. We subsequently consider three different strategies for communication in multiple-access fading channels, namely time-division with power control, superposition coding with fixed decoding order, and superposition coding with variable decoding order. Each scheme leads to different service rates at different users and results in different throughput regions.

A. Maximum Average Arrival Rate

We formulate the maximum average arrival rate for wireless transmissions in the presence of Markov arrivals and queueing constraints. Specifically, we consider two-state Markov arrival models in which the instantaneous arrival rates are r and 0 in the ON and OFF states, respectively. Probability of the ON state in steady-state is denoted as P_{ON}. Therefore, the average arrival rate is simply $r_{\text{avg}} = P_{\text{ON}} r$ which is equal to the average departure rate when the queue is in steady state (under the assumption that buffers are of infinite size) [9]. Now, we seek to determine our throughput metric, which is the maximum average arrival rate r_{avg}^* that can be supported in a fading channel while satisfying the statistical QoS limitations given in the form in (2). As shown in [9, Theorem 2.1], if the effective bandwidth of the arrival process is equal to the effective capacity of the service process, i.e.,

$$a(\theta, r) = C_E(\text{SNR}, \theta), \quad (8)$$

then, (2) is satisfied, i.e., buffer violation probability decays exponentially fast with rate controlled by the QoS exponent θ. Hence, we can determine from (8) the maximum arrival rate $r^*(\text{SNR}, \theta)$ that can be supported by the wireless channel for given SNR and QoS exponent θ. Then, the maximum average arrival rate is

$$r_{\text{avg}}^*(\text{SNR}, \theta) = r^*(\text{SNR}, \theta) P_{\text{ON}}. \quad (9)$$

1) Discrete Markov Source: In the case of ON-OFF discrete Markov source, using effective bandwidth expression in (4), we can express (8) in the following equivalent form:

$$\left(p_{11} + p_{22} e^{r\theta} - 2e^{\theta C_E}\right)^2 = (p_{11} + p_{22} e^{r\theta})^2 - 4(p_{11} + p_{22} - 1) e^{r\theta}.$$

Solving for r in the above equality, we can obtain the maximum arrival rate $r^*(\text{SNR}, \theta)$ and then express the maximum average arrival rate as a function of the effective capacity C_E as

$$r^*_{\text{avg}}(\text{SNR}, \theta) = P_{\text{ON}}\left[C_E(\text{SNR}, \theta) + \frac{1}{\theta}\log_e\left(\frac{e^{\theta C_E(\text{SNR},\theta)} - p_{11}}{1 - p_{11} - p_{22}\left(1 - e^{\theta C_E(\text{SNR},\theta)}\right)}\right)\right].$$

Note that probability of the ON state is given by $P_{\text{ON}} = \frac{1-p_{11}}{2-p_{11}-p_{22}}$.

2) Markov Fluid Source: Similarly as above, for the ON-OFF Markov fluid source, using (6), we can rewrite (8) as

$$(\theta r - (\alpha + \beta) - 2\theta C_E)^2 = (\theta r - (\alpha + \beta))^2 + 4\alpha\theta r \quad (10)$$

from which we determine the maximum average arrival rate as

$$r^*_{\text{avg}}(\text{SNR}, \theta) = P_{\text{ON}}\frac{\theta C_E(\text{SNR}, \theta) + \alpha + \beta}{\theta C_E(\text{SNR}, \theta) + \alpha} C_E(\text{SNR}, \theta). \quad (11)$$

Note that probability of ON state is given by $P_{\text{ON}} = \frac{\alpha}{\alpha+\beta}$.

Having formulated, for both Markov arrival models, the maximum average arrival rates in terms of the effective capacity, we next consider three different communication schemes over the multiple-access fading channel and identify the corresponding effective capacities.

B. TDMA

Time division is a simple strategy in which the users send their signals in non-overlapping intervals. Hence, interference is avoided in this case at the cost of reduced transmission rates. User i transmits τ_i fraction of the time with energy $\mathcal{E}_i/\tau_i$. Therefore, instantaneous service rates in bits/channel use are

$$R_i(\text{SNR}_i) = \tau_i \log_2\left(1 + \frac{\text{SNR}_i z_i}{\tau_i}\right) \text{ for } i = 1, ..., M \quad (12)$$

where again $z_i = |h_i|^2$. With these service rates, the effective capacity expressions of the users become

$$\mathsf{C}_{Ei}(\text{SNR}_i) = -\frac{1}{\theta_i}\log_e \mathbb{E}\left\{e^{-\frac{\theta_i\tau_i}{\log_e 2}\log_e\left(1+\frac{\text{SNR}_i z_i}{\tau_i}\right)}\right\}. \quad (13)$$

C. Superposition Coding with Fixed Decoding Order (SC-FDO)

In this strategy, transmitters simultaneously send the data and the receiver decodes the received sum-signal in a fixed-order denoted by π_k (for $k = 1, \ldots, M!$) during τ_k fraction of the time. Note that signals of M users can be decoded in $M!$ different orders. Note also that the time fractions $\{\tau_k\}$ satisfy $\tau_k \geq 0$ and $\sum_{k=1}^{M!}\tau_k = 1$. The throughput region is characterized by varying the values of $\{\tau_k\}$. In τ_k fraction of the time, instantaneous service rate of user i in bits/channel use is given by

$$R_{\pi_k(i)} = \log_2\left(1 + \frac{\text{SNR}_i z_i}{1 + \sum_{\pi_k(j)>\pi_k(i)}\text{SNR}_j z_j}\right). \quad (14)$$

Note from the above rate expression that user j with $\pi_k(j) > \pi_k(i)$ is decoded later than user i when decoding order π_k is employed and hence user i sees user j's signal as interference. Through successive interference cancelation, the signals of the previously-decoded users do not interfere. Accordingly, the effective capacity expression is given by

$$\mathsf{C}_{Ei}(\text{SNR}_i) = -\frac{1}{\theta_i}\log_e \mathbb{E}\left\{e^{-\theta_i\sum_{m=1}^{M!}\tau_m R_{\pi_m(i)}}\right\}. \quad (15)$$

D. Superposition Coding with Variable Decoding Order (SC-VDO)

In this method, users again transmit simultaneously. However, differently from the previous scheme in which the decoding is fixed in each fraction of time, we now consider varying the decoding order depending on the channel states or more specifically channel fading magnitude-squares $\mathbf{z} = [z_1, ..., z_M] \in \mathbb{R}^M_+$. Assume that the space of fading powers $\mathbb{R}^M_+$ is partitioned into $M!$ regions denoted by $\{\Gamma_k\}_{k=1}^{M!}$. If $\mathbf{z} \in \Gamma_k$, the decoding order π_k is used at the receiver. For a given partition, the effective capacity expression of the i^{th} user can now be expressed as

$$C_{Ei}(\text{SNR}_i) = -\frac{1}{\theta_i}\log_e \mathbb{E}\left\{e^{-\theta_i\sum_{k=1}^{M!}R_{\pi_k(i)}\mathbf{1}\{\mathbf{z}\in\Gamma_k\}}\right\} \quad (16)$$

where $\mathbf{1}\{\}$ is the indicator function, and $R_{\pi_k(i)}$ is given in (14). Since determining the optimal partition of the fading state space is in general a difficult task, we consider suboptimal strategies in order to demonstrate the possible improvements of adopting variable decoding order. In particular, one strategy is

$$\frac{\lambda_{\pi(1)}}{\text{SNR}_{\pi(1)}z_{\pi(1)}} \leq \frac{\lambda_{\pi(2)}}{\text{SNR}_{\pi(2)}z_{\pi(2)}} \leq ... \leq \frac{\lambda_{\pi(M)}}{\text{SNR}_{\pi(M)}z_{\pi(M)}} \quad (17)$$

whose performance was shown in [7] to be close to that of the optimal one, which maximizes the weighted sum-throughput in the special case of two users and constant arrival rates.

V. Numerical Results

While the analysis above is general, we in this section provide numerical results considering the case of two users, i.e., $M = 2$, and assuming a symmetric setting in which $\text{SNR}_1 = \text{SNR}_2 = \text{SNR}$ and $\theta_1 = \theta_2 = \theta$. The channel fading magnitude-squares z_1 and z_2 are exponentially distributed with arbitrary correlation ρ. Also, for discrete Markov sources, we set $p_{11} = 1-q$ and $p_{22} = q$. In all figures, we plot the regions of average arrival rates (or equivalently the throughput regions) achieved by TDMA, SC-FDO and SC-VDO.

In Figure 1, we plot the throughput regions for $p_{22} = q = 0.5$ and $q = 0.3$ when the arrivals are modeled as a discrete Markov process and we have $\theta = 1$, $\text{SNR} = 10$, and $\rho = 0$. For these parameter values, we notice that the SC-VDO using the strategy in (17) provides the largest throughput region, demonstrating the benefits of variable-decoding order. Interestingly, sum-rate achieved by TDMA exceeds that of SC-FDO. In this figure, we also observe the impact of burstiness on rate regions. When $p_{22} = q$, which is the probability for ON state, has a lower value, data arrivals of given rate r occur less frequently and hence the source is more bursty. It is clearly seen that increased source burstiness reduces the throughput regions of all strategies in a similar fashion.

In Figs. 2 and 3, we plot the throughput regions again considering discrete Markov arrivals. In Fig. 2, we observe

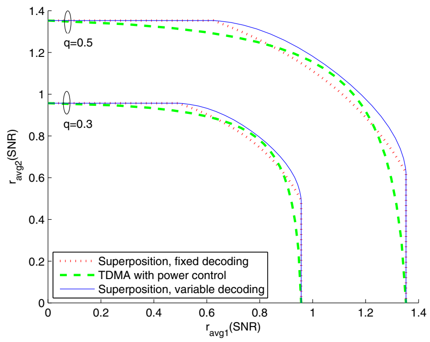

Fig. 1. Throughput regions for discrete Markov sources when $\theta = 1$, SNR $=$ 10 and $\rho = 0$.

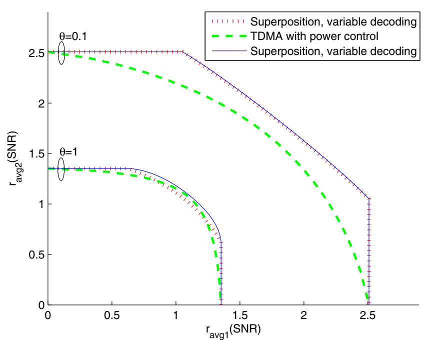

Fig. 2. Throughput regions for discrete Markov sources when $q = 0.5$, SNR $= 10$ and $\rho = 0$.

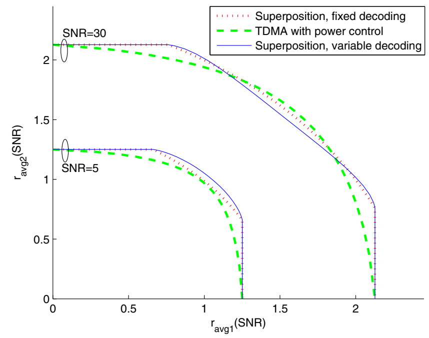

Fig. 3. Throughput regions for discrete Markov sources when $q = 0.5$, $\theta = 0.7$ and $\rho = 0$.

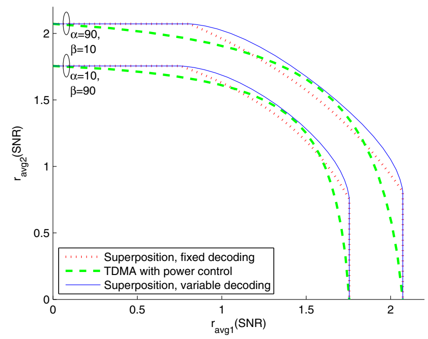

Fig. 4. Throughput regions for Markov fluid sources when $\theta = 1$, SNR $= 10$ and $\rho = 0.1$.

that increasing θ from 0.1 to 1 (or equivalently imposing more stringent QoS constraints) reduces the throughput regions. We also notice that while TDMA results in the smallest throughput region for less strict QoS constraints (i.e., when $\theta = 0.1$), TDMA sum rate exceeds that of SC-FDO when θ is increased to 1. Hence, buffer constraints have significant impact on the performance of different communication strategies. In Fig. 3, we see that increasing SNR expectedly improves the throughput regions. Surprisingly, TDMA sum-rate becomes the largest at SNR$=30$, which is in stark contrast to the results in the absence of buffer constraints in which TDMA is always suboptimal with respect to superposition transmissions.

In Fig. 4, we consider Markov fluid arrivals. Fading correlation is $\rho = 0.1$. We demonstrate the effect of different values of the transition rates α and β on the throughput region. Having α small and β large (with $\alpha + \beta$ fixed) results in a smaller probability for the ON state, representing a more bursty source. Again, as in Fig. 1, we note that increased burstiness hurts the throughput.

VI. Conclusion

We have studied the throughput regions of TDMA, SC-FDO and SC-VDO schemes in multiple-access fading channels in the presence of Markov data arrivals and QoS constraints. We have obtained expressions for the maximum average arrival rates in terms of the effective capacity of different multiple-access transmission schemes. We have demonstrated that source burstiness and stricter QoS constraints adversely affect the throughput. We have also seen that while SC-VDO can provide improvements on the throughput region, the simple strategy of TDMA becomes favorable in terms of sum rates if QoS constraints are strict or SNR is high. Under stricter QoS constraints, transmitters become more conservative and admit only smaller arrival rates which can be effectively supported by TDMA. When SNR is high, interference becomes more significant, and interference avoidance via TDMA tend to improve the performance.

References

[1] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.

[2] D. Tse and S. Hanly, "Multi-access fading channels—part I: polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[3] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queuing networks," *IEEE Trans. Auto. Control*, vol. 39, no. 5, pp. 913–931, May 1994.

[4] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643. July 2003.

[5] A. I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," *IEEE/ACM Trans. Networking.*, vol. 1, no. 3, pp. 329–343, Jun. 1993.

[6] G. Kesidis, J.Walrand, and C.-S. Chang, "Effective bandwidths for multiclass Markov fluids and other ATM sources," *IEEE/ACM Trans. Networking.*, vol. 1, no. 4, pp. 424–428, Aug. 1993.

[7] D. Qiao, M. C. Gursoy, and S. Velipasalar, "Transmission strategies in multiple-access fading channels with statistical QoS constraints," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1578–1593. Mar. 2012.

[8] E. M. Yeh, "Fundamental performance limits in cross-layer wireless optimization: throughput, delay, and energy," *Foundations and Trends in Commun. and Inf. Theory*, vol. 9, no. 1, pp. 1–112, 2012.

[9] C.-S. Chang and T. Zajic, "Effective bandwidths of departure processes from queues with time varying capacities," in *Proc. 1995 IEEE INFOCOM*.