Joint Admission Control and Antenna Assignment for Multiclass QoS in Spatial Multiplexing MIMO Wireless Networks

Dusit Niyato, *Member, IEEE*, Ekram Hossain, *Senior Member, IEEE*, and Dong In Kim, *Senior Member, IEEE*

***Abstract*—We consider the problem of quality-of-service (QoS) provisioning for multiple traffic classes in a MIMO wireless network. This QoS provisioning is posed as a radio resource management (RRM) problem at a wireless node (e.g., a wireless mesh router) with multiple antennas. We decompose this RRM problem into two tractable subproblems, namely, the antenna assignment and the admission control problems. The objective of antenna assignment is to minimize the weighted packet dropping probability for the different traffic classes under constrained packet delay. The objective of admission control is to maximize the revenue of the wireless node gained from the ongoing connections for different traffic classes under constrained connection blocking probability and average per-connection throughput. The decision of antenna assignment is made in a short-term basis (e.g., for every packet transmission interval) while that of admission control is made in a long-term basis (i.e., when a connection arrives). Constrained Markov decision process (CMDP) models are formulated to obtain the optimal decisions on antenna assignment and admission control. To provide efficient channel utilization, the RRM framework considers adaptive modulation at the physical layer which exploits channel state information. Performance evaluation results show that this joint antenna assignment and admission control framework can provide class-based service differentiation while satisfying both the connection-level and packet-level QoS requirements.**

***Index Terms*—Multiple-input multiple-output (MIMO) antenna system, antenna assignment, admission control, connection-level and packet-level QoS.**

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) antenna technology has been identified as one of the enabling physical layer technologies to provide high capacity wireless links for next generation wireless networks [1] such as IEEE 802.11n [2] and IEEE 802.16e/Mobile WiMAX [3] networks. MIMO systems are capable of enhancing the transmission rate and improving the transmission reliability. Specifically, when combined with adaptive modulation, a MIMO system can achieve much higher transmission rate than that of single-input single-output (SISO) system [4]. While numerous works in the literature studied the performance benefits of MIMO links from a physical layer point of view, only some recent studies (e.g., in [5]-[16]) have addressed the higher layer protocol design issues in single-hop as well as multi-hop MIMO wireless networks. To exploit the benefits of MIMO links in a wireless network and quality-of-service (QoS) provisioning for different classes of traffic, efficient radio resource management methods (e.g., for power allocation, antenna assignment, adaptive modulation, admission control) need to be designed.

In this paper, we address the problem of radio resource management (RRM) to support QoS for multiple traffic classes in a wireless node[1] with multiple antennas (for example, a mesh router/access point/base station). The packets from the ongoing connections in the same traffic class are buffered in the same queue. This is in line with the differentiated service (DiffServ) network architecture for QoS provisioning in wired networks [17] which has the characteristics of self-organization, auto-configuration, and better scalability. The spatial multiplexing mode for MIMO transmission is considered where multiple antennas can be used for transmission of packets from the same queue or from different queues. To improve transmission efficiency, adaptive modulation is used. To ensure both the packet-level and the connection-level QoS for different traffic classes in the wireless router and optimize the system performance, we develop a RRM framework for antenna assignment and admission control to control transmission of packets and to control arrival of time-correlated packet traffic, respectively, at a MIMO-enabled wireless router. In particular, the goal of antenna assignment is to achieve the packet-level QoS objectives (e.g., meet the packet delay and the packet dropping probability requirements). The goal of admission control is to guarantee connection-level QoS performances (i.e., connection blocking probability, average per-connection throughput) and maximize the revenue of the router from the ongoing connections of different traffic classes. These schemes are jointly optimized to maximize the utility of the wireless router while guaranteeing both the packet-level and connection-level QoS performances. The RRM problem is decomposed into two subproblems – one for antenna assignment and the other for admission control. For these subproblems, the entire system state space (i.e., the number of packets in the different queues and the number

Manuscript received December 16, 2008; revised March 26, 2009; accepted June 17, 2009. The associate editor coordinating the review of this paper and approving it for publication was M.-L. Merani.

D. Niyato is with the School of Computer Engineering, Nanyang Technological University (NTU), Singapore (e-mail: dniyato@ntu.edu.sg).

E. Hossain is with the Department of Electrical and Computer Engineering, University of Manitoba, Canada (e-mail: ekram@ee.umanitoba.ca).

D. I. Kim is with the School of Information and Communication Engineering, Sungkyunkwan University (SKKU), Korea (e-mail: dikim@ece.skku.ac.kr).

Digital Object Identifier 10.1109/TWC.2009.081648

[1]From now on, this wireless node will be referred to as a wireless *router* node.

1536-1276/09$25.00 © 2009 IEEE

of ongoing connections for the different traffic classes) and the action space (i.e., number of transmit antenna allocated to queues and admission/rejection of incoming connections) are decomposed accordingly. This improves the tractability of the corresponding analytical and optimization models. Specifically, optimization models based on constrained Markov decision process (CMDP) are formulated for each of these subproblems.

The rest of this paper is organized as follows. Section II presents the related work. Section III describes the system model and assumptions. For the antenna assignment problem, an optimization formulation based on CMDP is presented in Section IV. Section V presents the optimization formulation for the admission control problem. The performance evaluation results for the proposed RRM framework are presented in Section VI. Section VII states the conclusions.

II. Related Work

The problem of resource (i.e., antenna and power) allocation and QoS provisioning for MIMO systems has been addressed in recent literature. In [5], the problem of antenna assignment to multiple users was formulated as a weighted bipartite matching problem and Hungarian algorithm was used to obtain the optimal antenna assignment when the number of antennas is equal to the number of active users. In [6], the problem of antenna assignment among users for downlink transmission of a MIMO cellular system was formulated as a combinatorial optimization problem. The objective is to maximize the capacity while maintaining user satisfaction and fairness. In a similar spirit, a scheduling scheme was proposed in [7] to obtain an optimal selection of users and the assignment of their corresponding data to the transmit antennas in order to maximize throughput under fairness constraint. Multiuser scheduling schemes based on convex optimization were also proposed in [8] considering the availability of full and partial channel state information (CSI). The objectives, however, were to guarantee the rate constraints for the users only. In [9], the problem of transmit power allocation to channels in a MIMO link was addressed considering simultaneous transmission of QoS-sensitive and best-effort traffic. In [10], the rate and power control problem for spatial multiplexing MIMO systems was posed as a stochastic optimization problem to minimize the total transmission power over all transmitter antennas under the constraint of average delay for unknown channel/traffic statistics. A joint optimization model for power control, beamforming, and link scheduling was proposed in [11] to achieve proportional fairness and to support QoS in terms of data rates of all users. The problem of power allocation under QoS constraints was also addressed in [12] and [13] in the context of a MIMO-OFDM system. Different from the above works, the problem of resource allocation and QoS provisioning in asynchronous/distributed and multihop MIMO wireless networks was addressed in [5]-[16] (also see some of the references there in).

The above works, however, did not consider the problem of QoS provisioning (both connection-level and packet-level) for multiple (more than two) classes of traffic. In this paper, we solve this problem by using efficient antenna assignment and admission control methods taking the radio link level queueing dynamics in a wireless router into consideration.

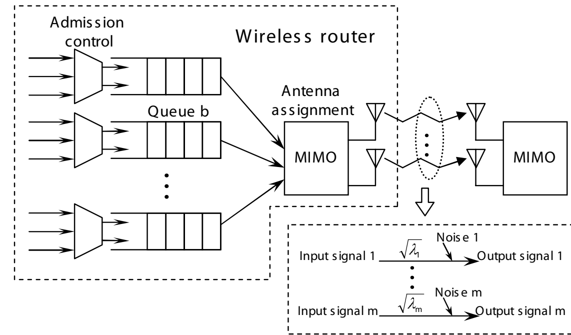


Fig. 1. System model of a MIMO-enabled wireless router with multiple traffic classes.

III. System Model and Assumptions

A. MIMO-Enabled Wireless Router

We consider a wireless router node with multiple transmit antennas which serve multiple traffic classes. With a total of B traffic classes, packets from all connections of traffic class b are buffered in queue b ($b = \{1, \dots, B\}$) as shown in Fig. 1. A single-hop spatial multiplexing MIMO transmission to another router node is considered where packet transmissions over the MIMO link are time-slotted. In each time slot, the available transmit antennas are allocated to the different queues. For each traffic class, admission control is used to decide whether an incoming connection can be accepted or not. The wireless router operates in a noise-limited environment in which the transmission channel is allocated to the router by a centralized controller. However, antenna assignment and admission control for the different traffic classes are performed locally at the router.

The packet arrivals for a connection in traffic class b follow a batch Markovian process with $\tilde{h}_b$ phases. The corresponding probability transition matrix is given by $\mathbf{A}_b'^{(a)}$ whose element is $A_b^{(a)}(h, h')$ for $a \in \{0, 1, \dots, \tilde{a}_b'\}$ arriving packets. $\tilde{a}_b'$ is the maximum batch size of a connection in class b. $A_b^{(a)}(h, h')$ denotes the probability that there are a arriving packets and the phase changes from h to h'. The matrix $\mathbf{A}_b'$ is defined as $\mathbf{A}_b' = \mathbf{A}_b'^{(0)} + \mathbf{A}_b'^{(1)} + \cdots + \mathbf{A}_b'^{(\tilde{a}_b')}$. Let $\boldsymbol{\nu} = \left[\begin{array}{ccccc} \nu(1) & \cdots & \nu(h) & \cdots & \nu(\tilde{h}_b') \end{array} \right]$ denote the steady state probability vector of packet arrival phase. The element $\nu(h)$ of this matrix is the steady state probability that the phase of packet arrival is h. This vector can be obtained by solving $\boldsymbol{\nu}\mathbf{A}_b' = \boldsymbol{\nu}$ and $\boldsymbol{\nu}\mathbf{1} = 1$. The average packet arrival rate of a connection in class b can be obtained from $\overline{a}_b' = \sum_{a=1}^{\tilde{a}_b'} a\boldsymbol{\nu}\left(\mathbf{A}_b'^{(a)}\mathbf{1}\right)$, where $\mathbf{1}$ is a vector of ones.

The packet arrival process of c' connections of class b can be obtained from

$$\mathbf{A}_b^{(a)}(c') = \sum_{\{g,g' | g+g'=a\}} \mathbf{A}_b^{(g)}(c'-1) \otimes \mathbf{A}_b'^{(g')} \qquad (1)$$

where $\mathbf{A}_b^{(a)}(1) = \mathbf{A}_b'^{(a)}$, and $\otimes$ is the Kronecker product. In this case, the aggregated packet arrival process of traffic class b with total c_b ongoing connections is defined as $\mathbf{A}_b^{(a)} = \mathbf{A}_b^{(a)}(c_b)$ for $a = \{0, 1 \ldots, \tilde{a}_b\}$, where $\tilde{a}_b = c_b \tilde{a}'_b$ is the maximum batch size. The average aggregated packet arrival rate of class b is obtained from $\overline{a}_b = c_b \overline{a}'_b$. The total number of phases of aggregated packet arrival is $\tilde{h}_b = (\tilde{h}'_b)^{c_b}$.

The state of the router can be defined by $(\mathcal{C}_1, \mathcal{X}_1, \mathcal{H}_1, \ldots, \mathcal{C}_b, \mathcal{X}_b, \mathcal{H}_b, \ldots, \mathcal{C}_B, \mathcal{X}_B, \mathcal{H}_B)$, where $\mathcal{C}_b$ is the number of ongoing connections for traffic class b, $\mathcal{X}_b$ is the number of packets in the corresponding queue, $\mathcal{H}_b$ is the phase of packet arrival of traffic class b, and B is the total number of traffic classes. Therefore, the size of the state space is $\prod_{b=1}^{B} c_b^* \prod_{b=1}^{B} (X_b + 1) \prod_{b=1}^{B} \tilde{h}_b$, where c_b^* is the maximum number of ongoing connections, X_b is the maximum queue size, and $\tilde{h}_b$ is the total number of aggregated packet arrival phases.

B. RRM in the Wireless Router

For the system model described above, the huge state space would make the analytical and optimization models for RRM in the wireless router intractable. The concept of hierarchical Markov decision process [18] is applied to address this tractability issue. In this paper, the RRM problem is decomposed into two subproblems, namely, the antenna assignment and admission control problems (Fig. 2(a)). With the decomposition of RRM, the decisions on antenna assignment and admission control can be made separately in the different time scales (i.e., on short-term and long-term basis, respectively) as shown in Fig. 2(b).

C. MIMO Transmission and Channel Model

A block fading frequency-flat MIMO channel with M transmit and N receive antennas is considered where the channel matrix $\mathbf{H}$ is an $N \times M$ matrix composed of independent complex Gaussian random variables with zero mean and unit variance [4]. For time-slotted transmissions, perfect channel information is assumed to be available at both the transmitter and receiver side. The channel matrix is assumed to remain constant within a time slot (i.e., quasi-static), but it varies independently from time slot to time slot. Using the Singular Value Decomposition (SVD) technique, the channel transfer matrix $\mathbf{H}$ can be decomposed into m parallel subchannels (Fig. 1, [4]), where where $m = \min(M, N)$, and the subchannel power gains are given by $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_m]$, where $\{\lambda_i\}_{i=1}^m$ are eigenvalues of $\mathbf{H}\mathbf{H}^H$ where $\mathbf{H}^H$ is the complex conjugate transpose of matrix $\mathbf{H}$. For total transmit power P, the transmit power at each sub-channel is $P_i = \frac{P}{m}$. The received SNR can then be calculated as $\gamma_i = \frac{P\lambda_i}{m\sigma^2} = \gamma_0 \lambda_i$, where σ^2 is the noise power. Note that, this power allocation is suboptimal given that perfect CSI is available. However, the proposed optimization formulation is also applicable when the adaptive weterfilling technique is used.

We adopt quadrature-amplitude modulation (e.g., 4-QAM, 16-QAM) with total number of modulation modes K. The SNR at the receiver $\gamma_i \in [0, \infty)$ is partitioned into $K + 1$ finite intervals with the thresholds $J_0(= 0) < J_1 < J_2 < \cdots < J_{K+1}(= \infty)$. Note that this threshold can be obtained given the target bit error rate (BER) [19]. If $J_k \leq \gamma_i < J_{k+1}$ ($k = 0, 1, 2, \cdots, K$), which will be called channel state k, modulation mode k is employed. The probability that modulation mode k is used in subchannel i can be obtained as follows:

$$R_i(k) = \Psi_1(1, J_k/\gamma_0) - \Psi_1(1, J_{k+1}/\gamma_0) \qquad (2)$$

where $\Psi_1(c, x)$ is given by (6) in [4]. Let $\mathbf{r}_i = \left[\ R_i(0) \quad R_i(1) \quad \cdots \quad R_i(K)\ \right]$ and $\mathbb{S}_b$ denote the set of antennas allocated to traffic class b. Based on the results in [21], the average bit-error-rate (BER) can be approximated as in [22]. For packet size of L bits (header + payload) and block coding with up to e bits of error correction which can be corrected in a packet, the average packet error rate can be obtained as follows:

$$\overline{PER}_k = 1 - \sum_{j=0}^{e} \left(\begin{array}{c} L \\ j \end{array} \right) (\overline{BER}_k)^j (1 - \overline{BER}_k)^{L-j}. \qquad (3)$$

With an infinite-persistent automatic repeat request (ARQ) protocol, the probability of d packets successfully transmitted on subchannel i can be obtained from

$$\tilde{\beta}_i(d) = \sum_{k=1}^{K} R_i(k) \left(\begin{array}{c} \hat{d}_k \\ d \end{array} \right) (1 - \overline{PER}_k)^d \overline{PER}_k^{d'-d} \qquad (4)$$

where $\hat{d}_k$ is the total number of packets transmitted using transmission mode k. Let us define $\tilde{\boldsymbol{\beta}}_i$ as $\tilde{\boldsymbol{\beta}}_i = \left[\ \tilde{\beta}_i(0) \quad \cdots \quad \tilde{\beta}_i(d) \quad \cdots \quad \tilde{\beta}_i(\hat{d}_k)\ \right]$. The matrix indicating the probability of successfully packet transmission $\beta_b(d)$ can be obtained from

$$\boldsymbol{\beta} = \bigodot_{i \in \mathbb{S}_b} \tilde{\boldsymbol{\beta}}_i \qquad (5)$$

where $\beta_b(d)$ is an element of this matrix $\boldsymbol{\beta}$ for $d \in \{0, 1, \ldots, \tilde{d}_b\}$ where $\tilde{d}_b = \hat{d}_K |\mathbb{S}_b|$, $|\mathbb{S}_b|$ is the cardinality of set $\mathbb{S}_b$, and $\bigodot$ denotes the discrete convolution [20].

IV. Optimization Formulation for the Antenna Assignment Problem

A. State Space, Action Space, and Decision Epoch for Antenna Assignment

The composite state of the CMDP formulation for antenna assignment is defined as follows:

$$\begin{aligned} \Phi \;=\; & \{(\mathcal{X}_1, \mathcal{H}_1, \ldots, \mathcal{X}_b, \mathcal{H}_b, \ldots, \mathcal{X}_B, \mathcal{H}_B); \\ & \mathcal{X}_b \in \{0, 1, \ldots, X_b\}, \mathcal{H}_b \in \{1, \ldots, \tilde{h}_b\}\} \end{aligned} \qquad (6)$$

where $\mathcal{X}_b$ is the number of packets in queue, X_b is the maximum queue size, and $\mathcal{H}_b$ is the phase of packet arrival for traffic class b. The action space is defined as $\mathbb{U}$ which is the set of all possible antenna assignments. Given action $u \in \mathbb{U}$, the set of antennas allocated to traffic class b is defined as $\mathbb{S}_b(u)$. The number of packets in queue and the phase of packet arrival are observed at the end of a time slot. Given the observed state, at the beginning of the next time slot, the decision on antenna assignment is made (Fig. 2(b)). During a time slot, packets are retrieved from queue b and transmitted according to the achievable transmission rate which depends on the set of allocated antennas $\mathbb{S}_b(u)$ and the modulation

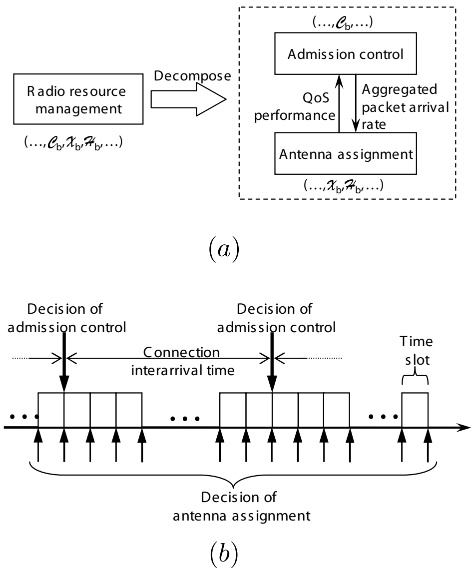


Fig. 2. (a) Decomposition of RRM problem for the wireless router and (b) decisions of antenna assignment and admission control.

modes used in the subchannel(s). The set $\mathbb{S}_b(u)$ affects the packet transmission probability (i.e., $\beta_b(d)$) as derived in (5).

As an example, let us consider the case of a router with two queues (i.e., $B = 2$), where $b = 1$ and $b = 2$ correspond to QoS-sensitive and best-effort traffic, respectively. Then the state space for antenna assignment is as follows: $\Phi = \{(\mathcal{X}_1, \mathcal{H}_1, \mathcal{X}_2, \mathcal{H}_2); \mathcal{X}_1 \in \{0, \ldots, X_1\}, \mathcal{X}_2 \in \{0, \ldots, X_2\}\}$, where $\mathcal{X}_1$ and $\mathcal{X}_2$ denote the number of packets in the queues for QoS-sensitive and best-effort traffic, respectively. The action space is $\mathbb{U} = \{1, \ldots, 4\}$. The sets of antennas allocated to QoS-sensitive and best-effort traffic are as follows: $\mathbb{S}_1(u = 1) = \emptyset$, $\mathbb{S}_2(u = 1) = \{1, 2\}$, $\mathbb{S}_1(u = 2) = \{1\}$, $\mathbb{S}_2(u = 2) = \{2\}$, $\mathbb{S}_1(u = 3) = \{2\}$, $\mathbb{S}_2(u = 3) = \{1\}$, $\mathbb{S}_1(u = 4) = \{1, 2\}$, and $\mathbb{S}_2(u = 4) = \emptyset$.

B. Transition Probability Matrix for Antenna Assignment

With general number of traffic classes $B \geq 2$, to obtain the transition probability matrix $\mathbf{P}(u)$ for the state space defined in (6), we first derive the transition probability matrix $\mathbf{Q}_b(u)$ for the queue of traffic class b. $\mathbf{Q}_b(u)$ is defined as in (7) where element $\hat{\mathbf{Q}}_b(x, x')$ is the transition probability matrix when the number of packets in queue of traffic class b changes from x in the current time slot to x' in the next time slot, $\tilde{d}_b$ is the maximum number of transmitted packets given a set of allocated antennas $\mathbb{S}_b(u)$, and $\tilde{a}_b$ is the maximum batch size of aggregated packet arrival of traffic class b. The element $\hat{\mathbf{Q}}_b(x, x')$ can be obtained as $\hat{\mathbf{Q}}_b(x, x+z) = \sum_{\{d,a|a-d=z\}} \mathbf{A}_b^{(a)} \beta_b(d)$, for $z = -d', \ldots, 0, \ldots, \tilde{a}_b$ where $d \in \{0, 1, \ldots, d'\}$ and $a \in \{0, 1, \ldots, \tilde{a}_b\}$. d' indicates the maximum number of transmitted packets which can be obtained from $d' = \min(\tilde{d}_b, x)$. The element $\hat{\mathbf{Q}}'_b(x, X_b)$ for $x + \tilde{a}_b > X_b$ can be obtained from $\hat{\mathbf{Q}}'_b(x, X_b) = \sum_{a=X_b-x}^{\tilde{a}_b} \hat{\mathbf{Q}}_b(x, x+a)$ and for $x = X_b$, we have $\hat{\mathbf{Q}}'_b(x, x) = \hat{\mathbf{Q}}_b(x, x) + \sum_{a=1}^{\tilde{a}_b} \hat{\mathbf{Q}}_b(x, x+a)$, where $\hat{\mathbf{Q}}_b(x, x')$ denotes the probability transition matrices when there is always enough space in queue to store the arriving packets.

Once the probability transition matrix $\mathbf{Q}_b(u)$ of traffic class b is obtained, then the probability transition matrix $\mathbf{P}(u)$ of the queues for all traffic classes can be obtained. According to the action u for antenna assignment, the packet arrivals and transmissions for each queue in a time slot are independent from those for other queues. Therefore, the probability transition matrix for the state space Φ can be obtained from

$$\mathbf{P}(u) = \bigotimes_{b=1}^{B} \mathbf{Q}_b(u). \tag{8}$$

where $\bigotimes$ is the Kronecker product [23].

C. CMDP Formulation: Objective and Constraints

The objective of antenna assignment is to minimize packet dropping probability (or equivalently to maximize throughput) while maintaining the packet delay below the threshold D_b^{max} for all traffic classes. For traffic class b, the long-term packet delay and packet dropping probability are respectively defined as follows:

$$\mathscr{J}_{D,b} = \lim_{t\to\infty} \sup \frac{1}{t} \sum_{t'=1}^{t} E\left(\mathscr{D}_b(\mathcal{S}(t'), \mathcal{U}(t'))\right) \tag{9}$$

$$\mathscr{J}_{L,b} = \lim_{t\to\infty} \sup \frac{1}{t} \sum_{t'=1}^{t} E\left(\mathscr{L}_b(\mathcal{S}(t'), \mathcal{U}(t'))\right) \tag{10}$$

where $\mathcal{S}(t') \in \Phi$ and $\mathcal{U}(t') \in \mathbb{U}$ denote the state and action of the antenna assignment algorithm, respectively, at time t', and $E(\cdot)$ denotes the expectation. $\mathscr{D}_b(s, u)$ and $\mathscr{L}_b(s, u)$ for $s \in \Phi$ and $u \in \mathbb{U}$ denote, respectively, the immediate packet delay and packet dropping probability of traffic class b. These are functions of composite state s and action u. Note that the composite state s is defined as $s = (x_1, h_1, \ldots, x_b, h_b, \ldots, x_B, h_B)$. The element of this composite state s is the realization of the variables defined in the state space Φ, i.e., the number of packets in queue x_b and the phase h_b of packet arrival.

The average packet delay can be expressed as a function of average number of packets in the queue and the average packet arrival rate as follows:

$$\overline{D}_b = \lim_{t\to\infty} \frac{1}{t} \sum_{t'=1}^{t} \frac{E(\mathcal{X}_b(t'))}{\overline{a}_b}. \tag{11}$$

From long-term average delay defined in (11), the immediate packet delay can be derived directly as $\mathscr{D}_b(s, u) = \frac{x_b}{\overline{a}_b}$, where $s = (x_1, h_1, \ldots, x_b, h_b, \ldots, x_B, h_B)$. It can be observed that the immediate packet delay is independent of packet arrival phase and action. However, in the long term, this state and action will affect the probability of having x_b packets in queue, and subsequently the average packet delay.

The immediate packet dropping probability (i.e., loss probability) can be obtained from $\mathscr{L}_b(s, u) =$

$$\frac{\sum_{z=X_b-x_b+1}^{\tilde{a}_b} z \left(\sum_{a+d=z} \beta_b(d) \sum_{h'=1}^{\tilde{h}_b} A_b^{(a)}(h_b, h') \right)}{\sum_{a=1}^{\tilde{a}_b} a \sum_{h'=1}^{\tilde{h}_b} A_b^{(a)}(h_b, h')} \tag{12}$$

where the numerator gives the average number of dropped packets and the denominator gives the total number of arriving packets given the state x_b and h_b of traffic class b.

$$\mathbf{Q}_b(u) = \begin{bmatrix} \hat{\mathbf{Q}}_b(0,0) & \cdots & & \hat{\mathbf{Q}}_b(0,\tilde{a}_b) & & & \\ \hat{\mathbf{Q}}_b(1,0) & & \hat{\mathbf{Q}}_b(1,1) & \cdots & & \hat{\mathbf{Q}}_b(1,\tilde{a}_b+1) & \\ \ddots & & \ddots & \ddots & & \ddots & \\ & \hat{\mathbf{Q}}_b(X_b-1, X_b-\tilde{d}_b-1) & & \cdots & \hat{\mathbf{Q}}_b(X_b-1,X_b-1) & & \hat{\mathbf{Q}}'_b(X_b-1,X_b) \\ & & & \hat{\mathbf{Q}}_b(X_b, X_b-\tilde{d}_b) & & \cdots & \hat{\mathbf{Q}}'_b(X_b,X_b) \end{bmatrix} \quad (7)$$

To achieve the objectives under the specified constraints, an optimal decision (i.e., optimal policy) of the antenna assignment can be obtained from the CMDP formulation. A policy π is defined as the mapping of state s to action u, i.e., $u = \pi(s)$ for $u \in \mathbb{U}$ and $s \in \Phi$. We consider a randomized policy in which action u to be taken at state s is chosen randomly according to a certain probability distribution denoted by $\mu(\pi(s))$ such that $\sum_{\pi(s)\in\mathbb{U}} \mu(\pi(s)) = 1$ (i.e., $\mu(u)$ is the probability of taking action u). The CMDP formulation can now be expressed as follows:

$$\text{Min:} \quad \sum_{b=1}^{B} w_b \mathscr{J}_{L,b}(\pi) \quad (13)$$

$$\text{Subj to:} \quad \mathscr{J}_{D,b}(\pi) \le D_b^{max}, \quad \forall b \in \{1,\ldots,B\} \quad (14)$$

where w_b is the weight of packet dropping probability for traffic class b. Note that the long-term packet delay and packet dropping probability here are defined as functions of policy π.

D. Optimal Policy for Antenna Assignment

The solution of the CMDP formulation for antenna assignment is referred to as the optimal policy π^*. To obtain π^*, the CMDP formulation in (13)-(14) can be transformed into an equivalent linear programming problem [24]. In particular, there is a one-to-one mapping between the optimal solution ϕ^* of the linear programming problem and the optimal policy π^*. Also, the solution of linear programming is feasible if and only if a solution of the CMDP formulation is feasible. Let $\phi(s,u)$ denote the steady state probability that action u is taken when the state is s. The linear programming problem corresponding to the CMDP formulation can be expressed as follows:

$$\text{Min:} \quad \sum_{s\in\Phi}\sum_{u\in\mathbb{U}} \phi(s,u) \sum_{b=1}^{B} w_b \mathscr{L}_b(s,u) \quad (15)$$

$$\text{Subj to:} \quad \sum_{s\in\Phi}\sum_{u\in\mathbb{U}} \mathscr{D}_b(s,u)\phi(s,u) \le D_b^{max}, \quad \forall b \quad (16)$$

$$\sum_{u\in\mathbb{U}} \phi(s',u) = \sum_{s\in\Psi}\sum_{u\in\mathbb{U}} P(s'|s,u)\phi(s,u) \quad (17)$$

$$\sum_{s\in\Psi}\sum_{u\in\mathbb{U}} \phi(s,u) = 1, \quad \phi(s,u) \ge 0 \quad (18)$$

for $s' \in \Phi$, where $P(s'|s,u)$ (which is an element of matrix $\mathbf{P}(u)$ defined in (8)) is the probability that the state changes from s to s' when action u is taken. The objective and the constraint defined in (15) and (16) correspond to those in (13) and (14), respectively. The constraint in (17) satisfies the Chapman-Kolmogorov equation and (18) satisfies the basic property of probability.

Let $\phi^*(s,u)$ denote the optimal solution of the linear programming problem defined in (15)-(18). The optimal policy π^* is a randomized policy which can be uniquely mapped from the optimal solution of the linear programming problem as follows:

$$\mu(u = \pi^*(s)) = \frac{\phi^*(s,u)}{\sum_{u'\in\mathbb{U}} \phi^*(s,u')}. \quad (19)$$

The optimal solution $\phi^*(s,u)$ can be obtained by using a standard method for solving linear programming. The optimal policy can be calculated off-line and stored in a look-up table to minimize computational overhead at a wireless router.

E. Packet-Level QoS Performance Measures

To obtain the packet-level QoS performance measures, the steady state probability (when optimal policy π^* is applied) would be required. The steady state probability for the system to be in state s is denoted by $p_{\pi^*}(s)$ for $s \in \Phi$, which can be obtained by solving the following set of equations: $\mathbf{p}_{\pi^*}\mathbf{P}(\pi^*) = \mathbf{p}_{\pi^*}$ and $\mathbf{p}_{\pi^*}\mathbf{1} = 1$, where $\mathbf{p}_{\pi^*} = \left[\cdots \quad p_{\pi^*}((x_1,h_1,\ldots,x_b,h_b,\ldots,x_B,h_B)) \quad \cdots \right]$. $\mathbf{P}(\pi^*)$ is the transition probability matrix when the optimal randomized policy $\pi^*(s)$ of antenna assignment is applied.

The average number of packets in queue for traffic class b can be obtained from

$$\overline{x}_b = \sum_{x_b=1}^{X_b} x_b \sum_{\forall x_{b'},\forall h_{b^\dagger}} p_{\pi^*}((\ldots,x_{b'},h_{b^\dagger},\ldots)) \quad (20)$$

for $b' \neq b$ and $b^\dagger \in 1,\ldots,B$.

The *average packet dropping rate* can be calculated from

$$\overline{x}_b^{dr} = \sum_{x_b=0}^{X_b} \sum_{z=X_b-x_b+1}^{\tilde{a}_b} z \left(\sum_{h_b=1}^{\tilde{h}_b} \sum_{a+d=z} \beta_b(d) \sum_{h'=1}^{\tilde{h}_b} A_b^{(a)}(h_b,h') \right) \times \sum_{\forall x_{b'},\forall h_{b'}} p_{\pi^*}((\ldots,x_{b'},h_{b'},\ldots)) \quad (21)$$

for $b' \neq b$. Then, the *packet dropping probability* for traffic class b is obtained as

$$L_b = \frac{\overline{x}_b^{dr}}{\overline{a}_b}. \quad (22)$$

The *queue throughput* is obtained from

$$Th_b = \overline{a}_b - \overline{x}_b^{dr} \quad (23)$$

and *average packet delay* can be obtained from Little's law as follows: $\overline{D}_b = \frac{\overline{x}_b}{Th_b}$.

V. Optimization Formulation for Connection Admission Control (CAC)

A. Admissible Region

In a wireless router, the admissible region is defined by the maximum number of ongoing connections for which the packet-level QoS requirements can be satisfied for all traffic classes. For example, the constraints on packet dropping probability can be considered when determining the admissible region. Given the maximum packet dropping probability threshold L_b^{max}, the maximum number of ongoing connections of traffic class b (i.e., boundary of admissible region) can be obtained from

$$c_b^* = \arg\max c_b \quad (24)$$
$$\text{Subj to: } L_b(c) \leq L_b^{max}, \quad \forall b \in \{1, \ldots, B\} \quad (25)$$

where $L_b(c)$ is the packet dropping probability of class b which is defined as a function of c. This packet dropping probability can be obtained as in (22). Here c is a composite state of number of ongoing connections of all traffic classes defined as follows: $c = (c_1, \ldots, c_b, \ldots, c_B)$. In this case, the packet dropping probability is obtained from (22) when the optimal policy for antenna assignment is used. The admissible region is defined as the following set:

$$\mathbb{A} = \{(0, \ldots, 0, \ldots, 0), \ldots, (c_1, \ldots, c_b^*, \ldots, c_B)\}.$$

The *simple* admission control policy is to accept an incoming connection as long as the number of ongoing connections lies in the admissible region. However, this simple admission control policy cannot ensure the highest utility (or revenue) for the wireless router and also may not meet the target connection blocking probability requirement for each traffic class. Therefore, an optimization problem based on CMDP is formulated to obtain the optimal policy for admission control.

B. State Space, Action Space, and Decision Epoch for CAC

For the admission control problem, the system state is given by the number of ongoing connections for the different traffic classes as follows: $\Omega = \{(\mathcal{C}_1, \ldots, \mathcal{C}_b, \ldots, \mathcal{C}_B) \in \mathbb{A}\}$, where $\mathcal{C}_b$ is the number of ongoing connections for traffic class b. The action space is defined as $\mathbb{V}$, where action $v \in \mathbb{V}$ corresponds to the decision of the admission controller to reject or accept an incoming connection of that traffic class. This action of admission controller is defined as $v = (v_1, \ldots, v_b, \ldots, v_B)$, where v_b denotes the admission control decision for an incoming connection of traffic class b. Specifically, $v_b = 1$ if the incoming connection is accepted, otherwise $v_b = 0$. The system state is observed when there is either an arrival or a departure of connection at the router. The decision on admission control is made when a connection arrives.

As an example, let us again consider the case of a router serving QoS-sensitive and best-effort traffic. The state space for admission control is as follows: $\Omega = \{(\mathcal{C}_1, \mathcal{C}_2)\}$, where $\mathcal{C}_1$ and $\mathcal{C}_2$ denote the number of ongoing QoS-sensitive and best-effort connections, respectively. The action space is $\mathbb{V} = \{(v_1, v_2)\} = \{(0,0), (1,0), (0,1), (1,1)\}$ whose elements correspond to the following actions: reject all incoming connections (i.e., $v_1 = 0$, $v_2 = 0$), accept only QoS-sensitive connection (i.e., $v_1 = 1$, $v_2 = 0$), accept only best-effort connection (i.e., $v_1 = 0$, $v_2 = 1$), and accept all incoming connections (i.e., $v_1 = 1$, $v_2 = 1$).

C. Transition Probability Matrix for CAC

We assume that connection arrivals follow a Poission process, while connection holding time is exponentially distributed. For traffic class b, let α_b and $1/\theta_b$ denote, respectively, the average connection arrival rate and average connection holding time. The transition rate matrix $\mathbf{N}(v)$ can be obtained for two cases, i.e., when a connection is accepted and when it is rejected.

The transition matrix in this case can be expressed as in (26) where $\mathbf{N}_B(c_B, c_B')$ is the transition matrix when the number of ongoing connections of class B changes from c_B to c_B'. These matrices can be obtained from

$$\mathbf{N}_B(c_B, c_B+1) = \begin{cases} \begin{bmatrix} \alpha_B & & \\ & \ddots & \\ & & \alpha_B \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}, & v_B = 1 \\ \mathbf{0}, & v_B = 0, \end{cases}$$
$$\mathbf{N}_B(c_B, c_B-1) = \begin{bmatrix} c_B\theta_B & & \mathbf{0} \\ & \ddots & \vdots \\ & c_B\theta_B & \mathbf{0} \end{bmatrix} \quad (27)$$

where $\mathbf{0}$ is a matrix of zeros. The matrix $\mathbf{N}_B(c_B, c_B)$ is defined as in (28) where $\mathbf{N}_{B-1}(c_{B-1}, c_{B-1}')$ is the transition matrix when the number of ongoing connections of class $B-1$ changes from c_{B-1} to c_{B-1}'. The elements $\mathbf{N}_{B-1}(c_{B-1}, c_{B-1}+1)$ and $\mathbf{N}_{B-1}(c_{B-1}, c_{B-1}-1)$ can be obtained in a manner similar to that in (27) in which α_B and θ_B become α_{B-1} and θ_{B-1}, respectively. The matrix $\mathbf{N}_{B-1}(c_{B-1}, c_{B-1})$ can be obtained in a manner similar to that in (28). The transition matrices $\mathbf{N}_b(c_b, c_b-1)$, $\mathbf{N}_b(c_b, c_b+1)$, and $\mathbf{N}_b(c_b, c_b)$ for $b = \{B-2, B-3, \ldots, 2, 1\}$ can be obtained similarly to those in (27) and (28).

Note that the diagonal element of the transition matrices $\mathbf{N}(v)$ for $v \in \mathbb{V}$ can be obtained from

$$[\mathbf{N}(v)]_{y,y} = -\sum_{y' \neq y} [\mathbf{N}(v)]_{y,y'} \quad (29)$$

where $[\mathbf{N}(v)]_{y,y'}$ denotes the element at row y and column y' of matrix $\mathbf{N}(v)$

The rate transition matrix $\mathbf{N}(v)$ can be transformed into an equivalent probability transition matrix $\mathbf{M}(v)$ by using uniformization method. This probability transition matrix can be obtained from

$$\mathbf{M}(v) = \frac{\mathbf{N}(v)}{\nu} + \mathbf{I} \quad (30)$$

for $v \in \mathbb{V}$ where $\mathbf{I}$ is an identity matrix and

$$\nu \geq \min_{y,v} \left(\left| [\mathbf{N}(v)]_{y,y} \right| \right). \quad (31)$$

In other words, ν is greater than or equal to the absolute value of the minimum diagonal element in $\mathbf{N}(v)$.

$$\mathbf{N}(v=1)=\begin{bmatrix} \mathbf{N}_B(0,0) & \mathbf{N}_B(0,1) & & & \\ \mathbf{N}_B(1,0) & \mathbf{N}_B(1,1) & \mathbf{N}_B(1,2) & & \\ \ddots & \ddots & \ddots & & \\ & \mathbf{N}_B(c_B^*-1,c_B^*-2) & \mathbf{N}_B(c_B^*-1,c_B^*-1) & \mathbf{N}_B(c_B^*-1,c_B^*) \\ & & \mathbf{N}_B(c_B^*,c_B^*-1) & \mathbf{N}_B(c_B^*,c_B^*) \end{bmatrix} \tag{26}$$

$$\mathbf{N}_B(c_B,c_B)=\begin{bmatrix} \mathbf{N}_{B-1}(0,0) & \mathbf{N}_{B-1}(0,1) & & & \\ \mathbf{N}_{B-1}(1,0) & \mathbf{N}_{B-1}(1,1) & \mathbf{N}_{B-1}(1,2) & & \\ \ddots & \ddots & \ddots & & \\ & \mathbf{N}_{B-1}(c_{B-1}^*-1,c_{B-1}^*-2) & \mathbf{N}_{B-1}(c_{B-1}^*-1,c_{B-1}^*-1) & \mathbf{N}_{B-1}(c_{B-1}^*-1,c_{B-1}^*) \\ & & \mathbf{N}_{B-1}(c_{B-1}^*,c_{B-1}^*-1) & \mathbf{N}_{B-1}(c_{B-1}^*,c_{B-1}^*) \end{bmatrix} \tag{28}$$

D. CMDP Formulation: Objective and Constraints

The objective of admission control is to maximize the long-term revenue of the router while satisfying the connection-level QoS requirements, namely, the target connection blocking probability and the target average per-connection throughput for all traffic classes. The long-term revenue, connection blocking probability, and per-connection throughput of traffic class b are defined, respectively, as follows:

$$\mathscr{J}_{R,b} = \lim_{t\to\infty} \sup \frac{1}{t}\sum_{t'=1}^{t} E\left(\mathscr{R}_b(\mathcal{C}(t'),\mathcal{V}(t'))\right), \tag{32}$$

$$\mathscr{J}_{B,b} = \lim_{t\to\infty} \sup \frac{1}{t}\sum_{t'=1}^{t} E\left(\mathscr{B}_b(\mathcal{C}(t'),\mathcal{V}(t'))\right) \tag{33}$$

$$\mathscr{J}_{Th,b} = \lim_{t\to\infty} \sup \frac{1}{t}\sum_{t'=1}^{t} E\left(\mathscr{T}_b(\mathcal{C}(t'),\mathcal{V}(t'))\right). \tag{34}$$

Here $\mathcal{C}(t') \in \Omega$ and $\mathcal{V}(t') \in \mathbb{V}$ denote, respectively, the state and action at time t'. $\mathscr{R}_b(c,v)$, $\mathscr{B}_b(c,v)$, and $\mathscr{T}_b(c,v)$ for $c \in \Omega$ and $v \in \mathbb{V}$ denote the immediate revenue, connection blocking probability, and per-connection throughput of traffic class b, respectively, which are functions of composite state c and action v. Note that the composite state c is defined as $c = (c_1,\ldots,c_b,\ldots,c_B)$. The element of this composite state c is the realization of the variables defined in the state space Ω, i.e., the number of ongoing connections for the different traffic classes.

If κ_b denotes the weight corresponding to revenue gained from traffic class b, the immediate revenue of the router can be obtained as a linear function of the corresponding number of ongoing connections as

$$\mathscr{R}_b(c,v) = c_b\kappa_b \tag{35}$$

for $c = (c_1,\ldots,c_b,\ldots,c_B)$. The immediate connection blocking probability can be obtained from

$$\mathscr{B}_b(c,v) = \begin{cases} \frac{\alpha_b}{\nu}, & (c_b+1 > c_b^*) \vee (v_b = 0) \\ 0, & \text{otherwise} \end{cases} \tag{36}$$

where α_b is the connection arrival rate of traffic class b, $\vee$ is the logical OR operator, and ν is obtained from (31). The immediate per-connection throughput can be calculated as $\mathscr{T}_b(c,v) = Th_b(c)$, where $Th_b(c)$ can be obtained from (23) given c_b ongoing connections.

The admission control policy is denoted by ρ for which $v = \rho(c)$ for $c \in \Omega$ and $v \in \mathbb{V}$. Again, we consider a randomized policy with probability distribution $\chi(\rho(c))$ for which $\sum_{\rho(c)\in\mathbb{V}} \chi(\rho(c)) = 1$. The CMDP formulation for admission control can now be expressed as follows:

$$\text{Max:} \quad \sum_{b=1}^{B} \mathscr{J}_{R,b}(\rho) \tag{37}$$

$$\text{Subj to:} \quad \mathscr{J}_{B,b}(\rho) \le B_b^{max} \tag{38}$$

$$\mathscr{J}_{Th,b}(\rho) \ge Th_b^{min} \tag{39}$$

for $b = \{1,\ldots,B\}$. B_b^{max} and Th_b^{min} are the maximum connection blocking probability and minimum per-connection throughput threshold, respectively. Note that the long-term revenue, connection blocking probability, and per-connection throughput here are defined as functions of policy ρ. Since the long-term admission control depends on throughput performance, the MIMO channel parameters implicitly affect the optimal admission control policy.

E. Optimal Policy for Admission Control

The solution of the CMDP formulation for admission control is referred to as the optimal policy ρ^* which maximizes the revenue of the wireless router while at the same time satisfies the connection-level QoS requirements. To obtain the optimal policy ρ^*, the CMDP formulation in (37)-(39) can be transformed into an equivalent linear programming problem. The optimal solution of the linear programming problem is denoted by δ^*. Let $\delta(c,v)$ denote the steady state probability that action v is taken when the state is c. The optimization problem corresponding to the CMDP formulation for admission control can be expressed as follows:

$$\text{Max:} \quad \sum_{c\in\Omega}\sum_{v\in\mathbb{V}}\sum_{b=1}^{B} \mathscr{R}_b(c,v)\delta(c,v) \tag{40}$$

$$\text{Subj to:} \quad \sum_{c\in\Omega}\sum_{v\in\mathbb{V}} \mathscr{B}_b(c,v)\delta(c,v) \le B_b^{max}, \quad \forall b \tag{41}$$

$$\sum_{c\in\Omega}\sum_{v\in\mathbb{V}} \mathscr{T}_b(c,v)\delta(c,v) \ge Th_b^{min}, \quad \forall b \tag{42}$$

$$\sum_{v\in\mathbb{V}} \delta(c',v) = \sum_{c\in\Omega}\sum_{v\in\mathbb{V}} M(c'|c,v)\delta(c,v) \tag{43}$$

$$\sum_{c\in\Omega}\sum_{v\in\mathbb{V}} \delta(c,v) = 1, \quad \delta(c,v) \ge 0 \tag{44}$$

for $c' \in \Omega$, where $M(c'|c,v)$ is the probability that the state changes from c to c' when action v is taken. This probability is the element of matrix $\mathbf{M}(v)$ (e.g., defined in (30)). The objective and the constraints defined in (40), (41), and (42) correspond to those in (37), (38) and (39), respectively.

The probability distribution of optimal randomized policy ρ^* can be obtained from

$$\chi(v = \rho^*(c)) = \frac{\delta^*(c, v)}{\sum_{v' \in \mathbb{V}} \delta^*(c, v')}. \tag{45}$$

F. Connection-level QoS Performance Measures

The connection-level performance measures for admission control can be obtained based on the steady state probabilities when the optimal policy ρ^* is applied. The steady state probability for the system to be in state c is denoted by $q_{\rho^*}(c)$ for $c \in \Omega$, which can be obtained by solving the following set of equations: $\mathbf{q}_{\rho^*}\mathbf{M}(\rho^*) = \mathbf{q}_{\rho^*}$ and $\mathbf{q}_{\rho^*}\mathbf{1} = 1$, where $\mathbf{q}_{\rho^*} = \left[\begin{array}{ccc} \cdots & q_{\rho^*}((c_1, \ldots, c_b, \ldots, c_B)) & \cdots \end{array}\right]$. $\mathbf{M}(\rho^*)$ is the probability transition matrix when the optimal randomized policy $\rho^*(c)$ for admission control is applied.

The *average number of ongoing connections* for traffic class b can be obtained from

$$\overline{c}_b = \sum_{c_b=1}^{c_b^*} c_b \left(\sum_{\forall c_{b'}} q_{\rho^*}((\ldots, c_{b'}, \ldots)) \right) \tag{46}$$

for $b' \neq b$. The *connection blocking probability* can be obtained from

$$Bl_b = \sum_{c=(c_1,\ldots,c_b,\ldots,c_B)} \left(\sum_{v \in \mathbb{V}} \mathscr{B}_b(c, v) q_{\rho^*}(c) \chi(v) \right). \tag{47}$$

The *per-connection throughput* can be obtained from

$$Thr_b = \sum_{c=(c_1,\ldots,c_b,\ldots,c_B)} \left(\sum_{v \in \mathbb{V}} Th_b(c) q_{\rho^*}(c) \chi(v) \right) \tag{48}$$

where $Th_b(c)$ is obtained from (23) given the number of ongoing connections of the different traffic classes.

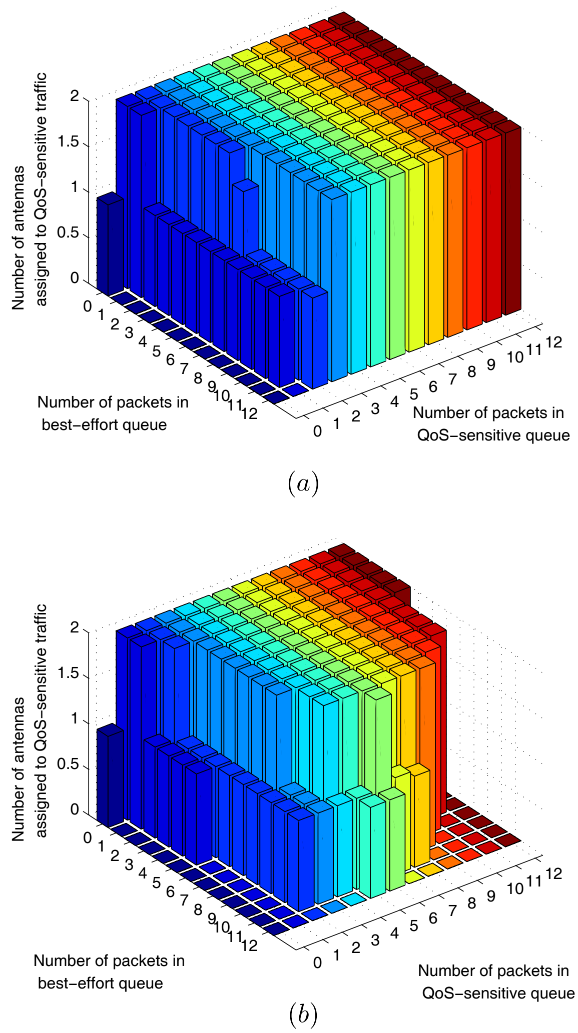


Fig. 3. Optimal policy for antenna assignment for (a) $D_1^{max} = 1.5$ and (b) $D_1^{max} = 2.1$ given the number of packets in queues for QoS-sensitive and best-effort traffic.

VI. Performance Evaluation

A. Parameter Setting

For numerical results, we consider a router with two transmit antennas and two queues (i.e., $B = 2$) in which $b = 1$ and $b = 2$ correspond to QoS-sensitive and best-effort traffic, respectively. The size of each queue (X_1, X_2) is 12 packets. The set of actions for antenna assignment is denoted by $\mathbb{U} = \{1, \ldots, 4\}$. The number of modulation modes is $K = 2$. The packet size is $L = 255$ bits and block codes are used for forward error correction in which errors up to $e = 2$ bits can be corrected. QoS-sensitive traffic has the maximum delay tolerance of $D_1^{max} = 1.5$ time slots, while best-effort traffic has no strict delay requirement. For antenna assignment, the weights corresponding to packet dropping probability are $w_1 = w_2 = 1$. The largest eigenvalue of the channel matrix is assumed to be 1. The target BER is 10^{-4}. The average SNR for the two subchannels are $\overline{\gamma}_1 = \overline{\gamma}_2 = 10$ dB where $\frac{P}{\sigma^2} = 20$. The packet arrivals for each connection follow a Poission process with average rate of $\overline{a}_1 = \overline{a}_2 = 0.2$ packets/time slot. The average connection arrival rate is $\alpha_1 = \alpha_2 = 0.2$ connection/min for both QoS-sensitive and best-effort traffic. The average connection holding times are $1/\theta_1 = 10$ and $1/\theta_2 = 15$ min for QoS-sensitive and best-effort connections, respectively. The maximum tolerable packet dropping probabilities for both QoS-sensitive and best-effort traffic are $L_1^{max} = L_2^{max} = 0.1$. For admission control, the weights corresponding to revenues from QoS-sensitive and best-effort traffic are $\kappa_1 = 1$ and $\kappa_2 = 0.2$, respectively. The maximum number of connections for both QoS-sensitive and best-effort traffic are $c_1^* = c_2^* = 10$. The maximum connection blocking probability and the minimum average per-connection throughput are $B_1^{max} = B_2^{max} = 0.05$ and $Th_1^{min} = Th_2^{min} = 0.1994$ packets/time slot.

B. Numerical Results

The analytical framework for RRM can be used to quantitatively evaluate both the packet-level and connection-level QoS performances and their interactions under wide variations in different system parameters. For brevity of the paper, however, only representative results are presented.

1) Optimal Policy for Antenna Assignment: Figs. 3(a)-3(b) show the number of antennas allocated to QoS-sensitive traffic

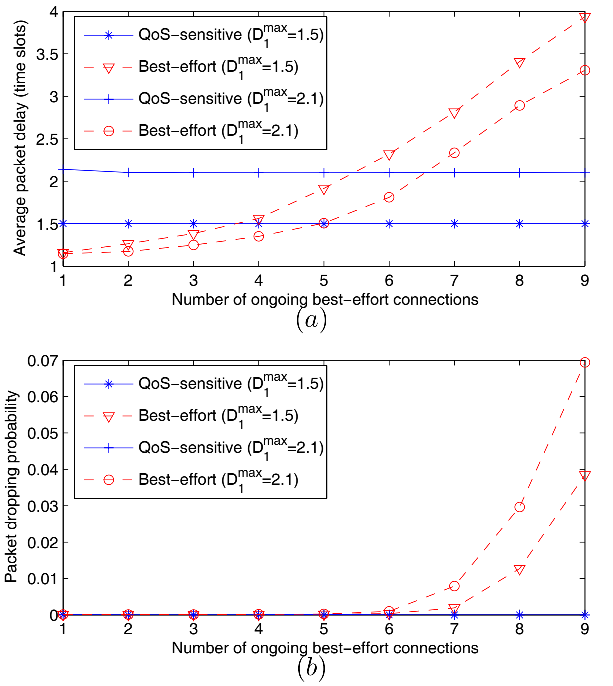


Fig. 4. (a) Average packet delay and (b) packet dropping probability of QoS-sensitive and best-effort traffic ($c_1 = 3$).

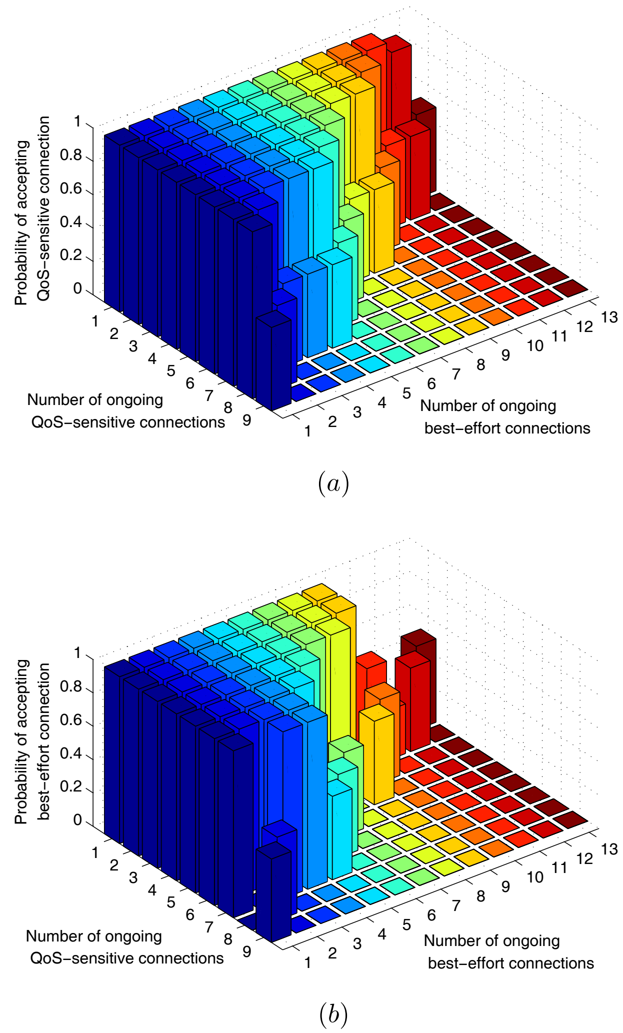


Fig. 5. Probability of accepting QoS-sensitive and best-effort connections obtained from optimal policy for admission control.

(for the optimal policy) given the number of packets in the queues of both QoS-sensitive and best-effort traffic. With small delay requirement (i.e., $D_1^{max} = 1.5$ in Fig. 3(a)), we observe that the optimal policy allocates all of the antennas to the QoS-sensitive traffic if the corresponding queue is not empty. However, if the number of packets in the queue for QoS-sensitive traffic is small (i.e., $x_1 = 1, 2$), some antennas are allocated to best-effort traffic so that the packet dropping probability is minimized. If the queue for QoS-sensitive traffic is empty, all antennas are allocated to the best-effort traffic instead. The optimal antenna assignment policy adapts to the QoS requirement of the QoS-sensitive traffic. For example, when the delay requirement of the QoS-sensitive traffic is larger (i.e., $D_1^{max} = 2.1$ in Fig. 3(b)), the number of allocated antennas to QoS-sensitive traffic becomes smaller.

The effect of adaptation (which is inherent in the optimal policy) on average packet delay and packet dropping performances are shown in Figs. 4(a)-4(b) as the number of ongoing best-effort connections varies. The optimal policy maintains the average packet delay for QoS-sensitive traffic below the maximum tolerable limits (i.e., $D_1^{max} = 1.5$ and $D_1^{max} = 2.1$ time slots). However, as the number of ongoing best-effort connections increases, the packet-level performances of best-effort traffic degrade. In order to avoid this performance degradation, admission control would be required.

2) Optimal Policy for CAC: Figs. 5(a)-5(b) show the probability of accepting QoS-sensitive and best-effort connections given the number of ongoing connections. This probability is obtained from the optimal policy for admission control. As expected, when the number of ongoing connections of both traffic classes is small, the admission controller accepts incoming connections for both QoS-sensitive and best-effort traffic. However, when the number of ongoing connections is large, the admission controller prioritizes QoS-sensitive connections over best-effort connections. This can be observed from the higher acceptance probability at the same number of ongoing connections. Note that outside the admissible region, the probabilities of accepting the QoS-sensitive and best-effort connection are zeros.

Figs. 6(a)-6(b) show the connection-level QoS performances for QoS-sensitive and best-effort traffic with optimal admission control and with *simple* admission control (i.e., based on admissible region only) when optimal antenna assignment policy is used. Since QoS-sensitive traffic has a maximum tolerable delay requirement, each connection requires more radio resources (i.e., transmit antennas). Therefore, with the simple admission control method, more best-effort connections can be accepted at the cost of increased connection blocking probability for QoS-sensitive traffic. However, with optimal admission control, connection blocking probability of QoS-sensitive traffic is much lower than that of best-effort traffic. Also, the connection blocking probability for best-effort traffic is bounded at 0.05 which is the maximum tolerable limit (Fig. 6(a)) and the average per-connection throughput does

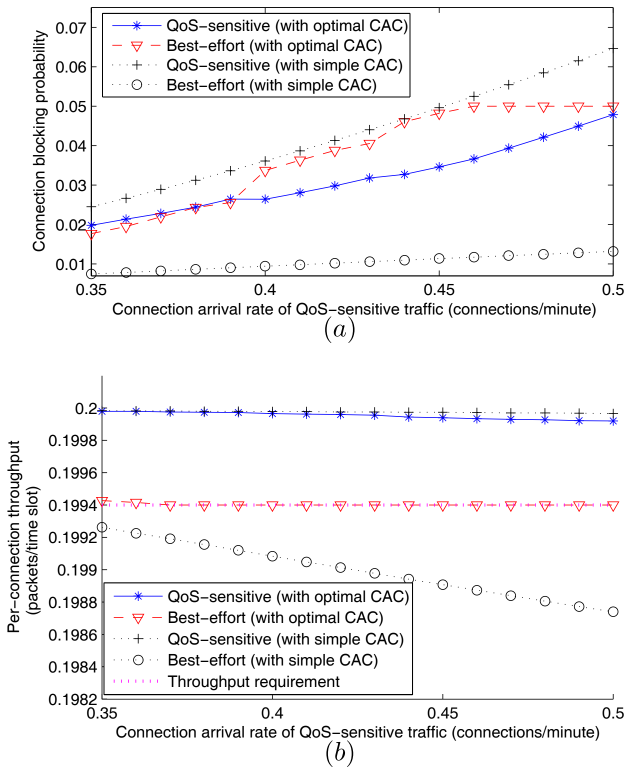


Fig. 6. (a) Connection blocking probability and (b) per-connection throughput under different connection arrival rate of QoS-sensitive traffic (for $\alpha_2 = 0.2$).

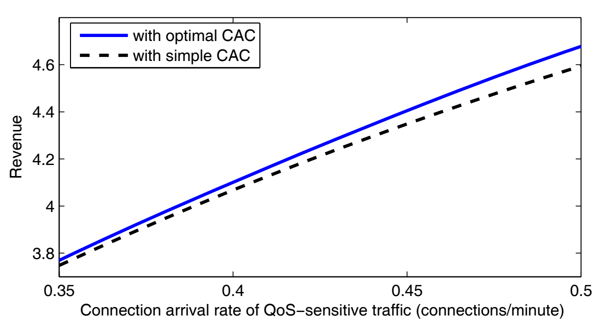


Fig. 7. Revenue of the router from QoS-sensitive and best-effort connections.

not fall below the minimum requirement of $Th_2^{min} = 0.1994$ packet/time slot (Fig. 6(b)). With the *simple* CAC scheme, the QoS requirements for the different traffic cannot be met. The optimal CAC policy, however, guarantees connection-level QoS while maximizing the revenue of the router at the same time (Fig. 7).

VII. Conclusion

We have presented a radio resource management framework for multiclass QoS provisioning in MIMO-enabled wireless routers. This framework consists of optimal antenna assignment to provide packet-level QoS and optimal admission control to provide connection-level QoS for different traffic classes. The optimal policies for antenna assignment and admission control have been obtained based on constrained Markov decision process models. Performance evaluation results have revealed the efficacy of the proposed framework. This framework can be extended to consider adaptive power allocation along with antenna assignment and the case of correlated MIMO channels. Also, extending the framework for a MIMO-OFDM system will be useful.

Acknowledgment

This work was supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the MKE (Ministry of Knowledge Economy), Korea under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment)(IITA-2009-C1090-0902-0005). This work was done in the Centre for Multimedia and Network Technology (CeMNet) of the School of Computer Engineering, Nanyang Technological University, Singapore.

References

[1] W. Ajib and D. Haccoun, "An overview of scheduling algorithms in MIMO-based fourth-generation wireless systems," *IEEE Network*, vol. 19, no. 5, pp. 43-48, Sept.-Oct. 2005.
[2] E. Perahia, "IEEE 802.11n development: history, process, and technology," *IEEE Commun. Mag.*, vol. 46, no. 7, pp. 48-55, July 2008.
[3] L. Garber, "Mobile WiMax: the next wireless battle ground," *IEEE Computer*, vol. 41, no. 6, pp. 16-18, June 2008.
[4] Z. Zhou, B. Vucetic, M. Dohler, and Y. Li, "MIMO systems with adaptive modulation," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1828-1842, Sept. 2005.
[5] Y.-J. Choi, J. Kim, and S. Bahk, "Downlink scheduling with fairness and optimal antenna assignment for MIMO cellular systems," in *Proc. IEEE Global Telecommun. Conf.*, vol. 5, pp. 3165-3169, Nov.-Dec. 2004.
[6] M. Torabzadeh and Y. Ji, "A near optimal antenna assignment for MIMO systems with low complexity," in *Proc. IEEE International Conf. Networks*, pp. 336-341, Nov. 2007.
[7] A. Ghassane and A. Sonia, "Adaptive scheduling for MIMO wireless networks: cross-layer approach and application to HSDPA," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 259-268, Jan. 2007.
[8] A. F. Dana, M. Sharif, A. Vakili, and B. Hassibi, "Differentiated rate scheduling for the down-link of cellular systems," *IEEE Trans. Commun.*, vol. 56, no. 10, pp. 1683-1693, Oct. 2008.
[9] J. Choi, "Power allocation for two different traffics in layered MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3942-3950, Oct. 2008.
[10] D. V. Djonin and V. Krishnamurthy, "Q-learning algorithms for constrained Markov decision processes with randomized monotone policies: application to MIMO transmission control," *IEEE Trans. Signal Processing*, vol. 55, pp. 5, part 2, pp. 2170-2181, May 2007.
[11] Y. Rong and Y. Hua, "Space-time power scheduling of MIMO links–fairness and QoS considerations," *IEEE J. Select. Topics Signal Processing*, vol. 2, no. 2, pp. 171-180, Apr. 2008.
[12] J. Xu, J. Kim, W. Paik, and J.-S. Seo, "Adaptive resource allocation algorithm with fairness for MIMO-OFDMA system," in *Proc. IEEE Veh. Technol. Conf.-Spring*, vol. 4, pp. 1585-1589, 2006.
[13] Y. Zhang and K. B. Letaief, "Adaptive resource allocation for multiaccess MIMO/OFDM systems with matched filtering," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1810-1816, Nov. 2005.
[14] M. Park, R. W. Heath, and S. M. Nettles Jr., "Improving throughput and fairness for MIMO ad hoc networks using antenna selection diversity," in *Proc. IEEE Global Telecommun. Conf.*, vol. 5, pp. 3363-3367, Nov.-Dec. 2004.
[15] B. Hamdaoui and P. Ramanathan, "Cross-layer optimized conditions for QoS support in multi-hop wireless networks with MIMO links," *IEEE J. Select. Areas Commun.*, vol. 25, no. 4, pp. 667-677, May 2007.
[16] A. Khattab, A. Sabharwal, and E. W. Knightly, "Fair randomized antenna allocation in asynchronous MIMO multi-hop networks," in *Proc. IEEE ICCCN'08*, St. Thomas, US, Aug. 2008.
[17] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," IETF, RFC 2475, Dec. 1998.
[18] D. Akselrod, A. Sinha, and T. Kirubarajan, "Hierarchical Markov decision processes based distributed data fusion and collaborative sensor management for multitarget multisensor tracking applications," in *Proc. IEEE International Conf. Systems, Man Cybernetics*, pp. 157-164, Oct. 2007.

[19] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746-1755, Sept. 2004.

[20] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia: Siam, 2000.

[21] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?" *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285-1294, July 2004.

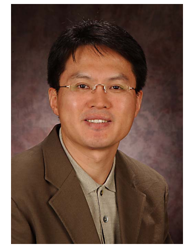
[22] K. J. Hole, H. Holm, and G. E. Oien, "Adaptive multidimensional coded modulation over flat fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 7, pp. 1153-1158, July 2007.

[23] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[24] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, Apr. 1994.

Dusit Niyato (M'09) is currently an assistant professor in the School of Computer Engineering, at the Nanyang Technological University, Singapore. He obtained his Ph.D. in Electrical and Computer Engineering from the University of Manitoba, Canada in 2008. His research interests are in the area of radio resource management in cognitive radio networks and broadband wireless access networks.

Ekram Hossain (S'98-M'01-SM'06) is currently an Associate Professor in the Department of Electrical and Computer Engineering at University of Manitoba, Winnipeg, Canada. Dr. Hossain's current research interests include design, analysis, and optimization of wireless communication networks and cognitive radio systems. He is a co-author/co-editor for the books *Dynamic Spectrum Access and Management in Cognitive Radio Networks* (Cambridge University Press, 2009), *Heterogeneous Wireless Access Networks* (Springer, 2008), *Introduction to Network Simulator NS2* (Springer, 2008), *Cognitive Wireless Communication Networks* (Springer, 2007), and *Wireless Mesh Networks: Architectures and Protocols* (Springer, 2007). Dr. Hossain serves as an Editor for the IEEE Transactions on Mobile Computing, the IEEE Transactions on Wireless Communications, the IEEE Transactions on Vehicular Technology, IEEE Wireless Communications, IEEE Communications Surveys and Tutorials and several other international journals. He served as a guest editor for the special issues of IEEE Communications Magazine (Cross-Layer Protocol Engineering for Wireless Mobile Networks, Advances in Mobile Multimedia) and IEEE Wireless Communications (Radio Resource Management and Protocol Engineering for IEEE 802.16). He served as a technical program co-chair for the IEEE Globecom 2007, IEEE WCNC 2008, IEEE VTC 2008-Fall, and QShine 2008: International Conference on Heterogeneous Networking for Quality, Reliability, Security, and Robustness. Dr. Hossain served as the technical program chair for the workshops on "Cognitive Wireless Networks" (CWNets 2007) and "Wireless Networking for Intelligent Transportation Systems" (WiN-ITS 2007) held in conjunction with QShine 2007 during August 14-17, in Vancouver, Canada, and the First IEEE International Workshop on Cognitive Radio and Networks (CRNETS 2008) in conjunction with IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008). He served as the technical program co-chair for the Symposium on "Next Generation Mobile Networks" (NGMN'06), NGMN'07, NGMN'08, NGMN'09 held in conjunction with International Wireless Communications and Mobile Computing Conference (IWCMC'06), IWCMC'07, IWCMC'08, and IWCMC'09. Dr. Hossain has several research awards to his credit which include Lucent Technologies Research Award for contribution to IEEE International Conference on Personal Wireless Communications (ICPWC'97), and the Best Student-paper Award in IWCMC'06. He is a registered Professional Engineer (P.Eng.) in the Province of Manitoba, Canada.



Dong In Kim (S'89-M'91-SM'02) received the B.S. and M.S. degrees in Electronics Engineering from Seoul National University, Seoul, Korea, in 1980 and 1984, respectively, and the M.S. and Ph.D. degrees in Electrical Engineering from University of Southern California (USC), Los Angeles, in 1987 and 1990, respectively.

From 1984 to 1985, he was a Researcher with Korea Telecom Research Center, Seoul. From 1986 to 1988, he was a Korean Government Graduate Fellow in the Department of Electrical Engineering, USC. From 1991 to 2002, he was with the University of Seoul, Seoul, leading the Wireless Communications Research Group. From 2002 to 2007, he was a tenured Full Professor in the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. From 1999 to 2000, he was a Visiting Professor at the University of Victoria, Victoria, BC. Since 2007, he has been with Sungkyunkwan University (SKKU), Suwon, Korea, where he is a Professor and SKKU Fellow in the School of Information and Communication Engineering. Since 1988, he is engaged in the research activities in the areas of wideband wireless transmission and access. His current research interests include cooperative relaying and base station (BS) cooperation, multiuser cognitive radio networks, advanced transceiver design, and cross-layer design.

Dr. Kim was an Editor for the IEEE Journal on Selected Areas in Communications: Wireless Communications Series and also a Division Editor for the Journal of Communications and Networks. He is currently an Editor for Spread Spectrum Transmission and Access for the IEEE Transactions on Communications and an Area Editor for Transmission Technology III for the IEEE Transactions on Wireless Communications. He also serves as Co-Editor-in-Chief for the Journal of Communications and Networks.