Supporting Hierarchical Users in a Random Multiple Access System

Faeze Heydaryan and Jie Luo
ECE Dept., Colorado State Univ., Fort Collins, CO 80523
Email: {faezeh66, rockey}@colostate.edu

***Abstract*—Classical random access protocols support prioritized user groups by allowing a high priority user to transmit with a relatively high probability. However, when a high priority user competes for the channel with a large number of low priority users, transmission success probability of the high priority user can still diminish to zero. In this paper, a distributed medium access control (MAC) framework is proposed to support hierarchical user groups in a random multiple access system. A hierarchical user structure, such as a primary-secondary user structure, differs from a priority user structure in the following senses. First, when the number of primary users is small, the MAC framework guarantees that channel availability should stay above a pre-determined threshold no matter how many secondary users are competing for the channel. Second, when the number of primary users is large, the MAC framework drives transmission probabilities of the secondary users to zero but does not reject channel access to the primary users. These properties are achieved in a distributed environment without direct message exchange between users, without knowledge on the number of users, and without knowing whether each transmission should belong to a primary or a secondary user.**

I. Introduction

Diversity of wireless devices and applications often requires wireless networks to provide differentiated services to users in the sense of supporting user groups with different priority levels. Take the enhanced distributed coordination function (EDCF) in 802.11e for example. Users (or traffics) in 802.11e EDCF can be assigned to four different priority levels with different adaptation schemes on their backoff windows. A high priority user generally maintains a backoff window smaller in size than that of a low priority user. This leads to a higher transmission probability for the high priority users and therefore gives their packets an advantage in getting through the shared wireless channel. However, such a priority user structure is "soft" in the sense that, when the system has a large number of low priority users whose transmission activities cause significant contention, the packet success probability of a high priority user can still be driven close to zero.

Recent trend of dynamic spectrum access (DSA) created the new demand of supporting a "hard" hierarchical user structure in wireless systems [1] [2]. Take channel sharing with the primary-secondary user structure for example. It is expected that secondary users should access the channel only if they can guarantee no disturbance to communication activities of the primary users. Existing DSA literature often assumes that secondary users should be able to identify whether an existing transmission should belong to a primary user or not. Disruptive interference is often avoided with online coordinations between primary and secondary users or within the secondary user group. There is little discussion on how to support hierarchical user groups in a random access environment, where users do not exchange information with each other directly, and where packet collision is part of the natural transmission outcomes.

Classical distributed MAC protocols such as 802.11 DCF and 802.11e EDCF assume each link layer user can only determine whether to transmit a packet or not. They also assume a simple link layer channel model such as the collision channel. Recently, a new channel coding theory was proposed in [3] and [4] for distributed communication at the physical layer. The coding theory allows each physical layer user (transmitter) to prepare an ensemble of channel codes, and to choose an arbitrary one to encode its message. While code ensembles of the users are assumed to be known, actual coding decisions are not shared among the users or with the receiver. The receiver, on the other hand, should either decode the messages of interest or report collision, depending on whether a pre-determined error probability requirement can be met. Fundamental limit of the system was characterized using a distributed channel capacity region defined in the vector space of the coding decisions of the users. The distributed capacity region was shown to coincide with the classical Shannon capacity region in a sense explained in [4]. The new channel coding theory provided the basic physical layer support for an enhancement to the physical-link layer interface [4], which allows each link layer user to be equipped with multiple transmission options. These options correspond to different codes at the physical layer, possibly representing different communication settings such as different transmission power and rate combinations. The interface enhancement enables data link layer protocols to exploit advanced wireless communication adaptations through the navigation of different transmission options. It also enables the modeling of a wide range of general but realistic link layer channels that can be derived from the physical layer channel and coding details of the packets.

With the new channel coding theory, a distributed MAC framework was proposed in [5] to support a general link layer channel model and to support multiple transmission options

This work was supported by the National Science Foundation under Grant CNS-1618960. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

at each link layer user [4]. In this paper, we further extend the MAC framework of [5] to support hierarchical primary-secondary user groups in a time-slotted random multiple access system. While the MAC algorithm can incorporate a general link layer channel model, we still assume a single transmission option at each user. Further extensions that equip each user with multiple transmission options can be found in [6].

II. A Stochastic Approximation Framework

Consider a wireless multiple access network with K users (transmitters) and a common receiver. Among the users, K_p of them are labeled as "primary" users and K_s of them are labeled as "secondary" users. $K = K_p + K_s$. Each user only knows its own label, but does not know the labels of other users, as well as the values of K_p, K_s and K. As we will explain later, users will adopt the same type of MAC algorithms, but with users of different labels setting their parameters differently. Other than such a difference, we assume that the users are homogeneous.

Let time be slotted with each slot equaling the length of one packet. Assume that each user is equipped with one transmission option and an idling option, and is backlogged with a saturated message queue. At the beginning of each time slot t, a user determines whether to idle or to transmit a packet according to an associated probability parameter. Transmission decisions of the users are made individually. They are shared neither among the users nor with the receiver. The probability parameter associated to user k, $k = 1, \ldots, K$, is denoted by $p_k(t)$ for time slot t. At the end of each time slot t, based on available channel feedback, which we will discuss later, each user k derives a target transmission probability $\tilde{p}_k(t)$. User k then updates its transmission probability by

$$p_k(t+1) = p_k(t) + \alpha(t)(\tilde{p}_k(t) - p_k(t)), \quad (1)$$

where $\alpha(t) > 0$ is a step size parameter of time slot t. Let $\boldsymbol{p}(t) = [p_1(t), p_2(t), \ldots, p_K(t)]^T$ denote a K-length vector that consists of the transmission probabilities of all users in time slot t. Let $\tilde{\boldsymbol{p}}(t) = [\tilde{p}_1(t), \tilde{p}_2(t), \ldots, \tilde{p}_K(t)]^T$ denote the corresponding target vector. According to (1), $\boldsymbol{p}(t)$ is updated by

$$\boldsymbol{p}(t+1) = \boldsymbol{p}(t) + \alpha(t)(\tilde{\boldsymbol{p}}(t) - \boldsymbol{p}(t)). \quad (2)$$

Probability adaptation given in (2) falls into the stochastic approximation framework [7] [8], where the target probability vector $\tilde{\boldsymbol{p}}(t)$ is often calculated from noisy estimates of certain system variables, e.g., channel idling probability.

Define $\hat{\boldsymbol{p}}(t) = [\hat{p}_1(t), \hat{p}_2(t), \ldots, \hat{p}_K(t)]^T$ as the "theoretical value" of $\tilde{\boldsymbol{p}}(t)$ under the assumption that there is no measurement noise and no feedback error in time slot t. Let $E_t[\tilde{\boldsymbol{p}}(t)]$ be the conditional expectation of $\tilde{\boldsymbol{p}}(t)$ given the system state at the beginning of time slot t. The difference between $E_t[\tilde{\boldsymbol{p}}(t)]$ and $\hat{\boldsymbol{p}}(t)$ is defined as the bias in the target probability vector calculation, denoted by $\boldsymbol{g}(t)$.

$$\boldsymbol{g}(t) = E_t[\tilde{\boldsymbol{p}}(t)] - \hat{\boldsymbol{p}}(t). \quad (3)$$

We assume that, given the communication channel, both $\hat{\boldsymbol{p}}(t) = \hat{\boldsymbol{p}}(\boldsymbol{p}(t))$ and $\boldsymbol{g}(t) = \boldsymbol{g}(\boldsymbol{p}(t))$ should only be functions of $\boldsymbol{p}(t)$, which is the transmission probability vector in time slot t.

The following two conditions are typically required for the convergence of a stochastic approximation algorithm [4] [5].

Condition 1: (Mean and Bias) There exists a constant $K_m > 0$ and a bounding sequence $0 \leq \beta(t) \leq 1$, such that $\|\boldsymbol{g}(\boldsymbol{p}(t))\| \leq K_m \beta(t)$, where $\|.\|$ denotes the second order norm. We assume that $\beta(t)$ is controllable in the sense that one can design protocols to ensure $\beta(t) \leq \epsilon$ for any chosen $\epsilon > 0$ and for large enough t.

Condition 2: (Lipschitz Continuity) There exists a constant $K_l > 0$, such that $\|\hat{\boldsymbol{p}}(\boldsymbol{p}_a) - \hat{\boldsymbol{p}}(\boldsymbol{p}_b)\| \leq K_l \|\boldsymbol{p}_a - \boldsymbol{p}_b\|$, for all $\boldsymbol{p}_a, \boldsymbol{p}_b$.

According to stochastic approximation theory [7] [8], if the above two conditions are satisfied, and values of the step size sequence $\alpha(t)$ and the bounding sequence $\beta(t)$ are small enough, then trajectory of the transmission probability vector $\boldsymbol{p}(t)$ under distributed adaptation algorithm given in (2) can be approximated by the following associated ordinary differential equation (ODE) in a sense explained in [7] [8],

$$\frac{d\boldsymbol{p}(t)}{dt} = -[\boldsymbol{p}(t) - \hat{\boldsymbol{p}}(t)], \quad (4)$$

where we used t again to denote the continuous time variable. Because all entries of $\boldsymbol{p}(t)$ and $\hat{\boldsymbol{p}}(t)$ stay in the range of $[0, 1]$, any equilibrium $\boldsymbol{p}^* = [p_1^*, \ldots, p_K^*]^T$ of the associated ODE must satisfy

$$\boldsymbol{p}^* = \hat{\boldsymbol{p}}(\boldsymbol{p}^*). \quad (5)$$

Suppose that the associated ODE given in (4) has a unique solution at $\boldsymbol{p}^*$, then under mild conditions on $\alpha(t)$ and $\beta(t)$, $\boldsymbol{p}(t)$ should converge to $\boldsymbol{p}^*$ in senses explained in [5, Theorem 1] and [5, Theorem 2]. With these convergence results, the key question is how to design a distributed MAC algorithm to satisfy Conditions 1 and 2 and to place the equilibrium of the associated ODE at a desired point.

In each time slot, the receiver makes message recovery and collision detection decisions for the data packets without knowing the transmission status of the users. The receiver also assumes the existence of a virtual packet, and makes a decision on whether virtual packet reception should be regarded as successful or not. Only one virtual packet is assumed in each time slot, and virtual packets assumed in different time slots are identical. As explained in [4] [5], a virtual packet is an assumed packet that is not physically transmitted by any user. Virtual packet detection essentially checks whether current operation point of the channel is located inside its fundamental limit with a pre-determined margin. General considerations on virtual packet design and detection are explained in [4, Section 4.2] and will not be discussed in this paper. We assume that the receiver should estimate the success probability of the virtual packet, denoted by $q_v(t)$ for time slot t, and feed it back to the users. $q_v(t)$ should then be used by each user to derive its target transmission probability.

As introduced in [5], we model the link layer multiple access channel using two sets of parameters, both can be theoretically derived from the physical layer channel model and coding details of the packets. The first parameter set, $\{C_{rj}\}$

for $j \geq 0$, is termed the "real channel parameter set." C_{rj} denotes the conditional success probability of a real packet, should it be transmitted in parallel with j other real packets. The second parameter set, $\{C_{vj}\}$ for $j \geq 0$, is termed the "virtual channel parameter set." C_{vj} denotes the success probability of the virtual packet should it be transmitted in parallel with j real packets. We assume that $C_{vj} \geq C_{v(j+1)}$ for all $j \geq 0$, meaning that the virtual packet should not get a better chance to go through the channel if the number of parallel real packet transmissions increases. Both $\{C_{rj}\}$ and $\{C_{vj}\}$ are assumed to be known at the users as well as at the receiver.

III. Supporting A Hierarchical User Structure

Let the system have K_p primary users and K_s secondary users, where the values of K_p and K_s are unknown to the users. Let $K = K_p + K_s$. We term q_v the "channel contention measure", which is the success probability of the virtual packet. If all users have the same transmission probability p, q_v can be calculated by

$$q_v(p, K) = \sum_{j=0}^{K} \binom{K}{j} p^j (1-p)^{K-j} C_{vj}. \quad (6)$$

Without global information, a primary (secondary) user assumes that the system should only contain an unknown number of primary (secondary) users. Upon receiving the feedback of q_v, each primary (secondary) user should obtain an estimated number of users, denoted by $\hat{K}$, and then use $\hat{K}$ to determine the corresponding transmission probability target. Such an operation requires each primary (secondary) user to design two key functions, denoted by $p_p^*(\hat{K})$ and $q_{vp}^*(\hat{K})$ ($p_s^*(\hat{K})$ and $q_{vs}^*(\hat{K})$), both are functions of $\hat{K}$. $p_p^*(\hat{K})$ ($p_s^*(\hat{K})$) represents the designed transmission probability of the users if the multiple access system only contains $\hat{K}$ primary (secondary) users. $q_{vp}^*(\hat{K})$ ($q_{vs}^*(\hat{K})$), on the other hand, represents the "theoretical channel contention measure" if the multiple access system only contains $\hat{K}$ primary (secondary) users, and all users have the same transmission probability of $p_p^*(\hat{K})$ ($p_s^*(\hat{K})$). Both functions should be designed for all values of $\hat{K} \geq 0$, including both integer and non-integer values. Define $K_{p\min}$ ($K_{s\min}$) as the maximum $\hat{K}$ that maximizes $p_{vp}^*(\hat{K})$ ($p_{vs}^*(\hat{K})$).

Given the two key functions, the distributed MAC algorithm should operate as follows.

Distributed MAC Algorithm:

1) Each user initializes its transmission probability.
2) Let $Q > 0$ be a pre-determined integer. Over an interval of Q time slots, the receiver measures the success probability of the virtual packet, denoted by q_v, and feeds q_v back to all users.
3) Upon receiving q_v, each primary (secondary) user should derive an estimate of the number of users $\hat{K}$ by solving the following equation.

$$q_{vp}^*(\hat{K}) = q_v \qquad (q_{vs}^*(\hat{K}) = q_v). \quad (7)$$

If a $\hat{K}$ satisfying (7) cannot be found, a primary (secondary) user should set $\hat{K} = K_{p\min}$ ($\hat{K} = K_{s\min}$) if $q_v > \max_{\hat{K}} q_{vp}^*(\hat{K})$ ($q_v > \max_{\hat{K}} q_{vs}^*(\hat{K})$), or it should set $\hat{K} = \infty$ otherwise.
4) Each primary (secondary) user, say user k_p (k_s), updates its transmission probability by

$$p_{k_p} = (1-\alpha)p_{k_p} + \alpha p_p^*(\hat{K}),$$
$$(p_{k_s} = (1-\alpha)p_{k_s} + \alpha p_s^*(\hat{K}),) \quad (8)$$

where α is the step size parameter for user k_p (k_s).
5) The process is repeated from Step 2 till transmission probabilities of all users converge.

We will now introduce the specific design of the two key functions. Let us first focus on a primary user. We assume that primary users intend to maximize a symmetric network utility, denoted by $U_p(\hat{K}, p_p, \{C_{rj}\})$, under the assumption that the system contains $\hat{K}$ homogeneous primary users and all users have the same transmission probability p_p. Let x_p^* be obtained from the following asymptotic utility optimization.

$$x_p^* = \arg\max_x \lim_{\hat{K}\to\infty} U_p\left(\hat{K}, \frac{x}{\hat{K}}, \{C_{rj}\}\right). \quad (9)$$

Based on the utility optimization objective, a primary user should design its desired transmission probability function $p_p^*(\hat{K})$ as

$$p_p^*(\hat{K}) = \frac{x_p^*}{\max\{\hat{K}, \hat{K}_{p\min}\} + b_p}, \quad (10)$$

where $\hat{K}_{p\min}$ and b_p are design parameters whose values should be determined by following the guideline given in [5]. Particularly, $\hat{K}_{p\min}$ should take an integer value slightly less than x_p^*, and b_p should be chosen to satisfy $b_p > \max\{1, x_p^* - \gamma_{p\epsilon_v}\}$, where $\gamma_{p\epsilon_v}$ is a parameter defined in [5, Theorem 4]. We skip the detailed definition of $\gamma_{p\epsilon_v}$ here because usually a good design should yield $x_p^* - \gamma_{p\epsilon_v} < 1$, and hence the effective constraint on b_p should be simplified to $b_p > 1$ [5, Theorem 5].

With the $p_p^*(\hat{K})$ function given by (10), for integer-valued $\hat{K}$, $q_{vp}^*(\hat{K})$ should equal the actual channel contention measure when the estimated number of users is accurate and when all users have the same transmission probability of $p_p^*(\hat{K})$. In other words, we should have

$$q_{vp}^*(\hat{K}) = q_v(p_p^*(\hat{K}), \hat{K}), \quad (11)$$

where $q_v(p_p^*(\hat{K}), \hat{K})$ can be further calculated using (6). For non-integer-valued $\hat{K}$, the "theoretical channel contention measure" function $q_{vp}^*(\hat{K})$ should be designed using the following linear interpolation approach [5].

$$q_{vp}^*(\hat{K}) = \frac{p_p^*(\hat{K}) - p_p^*(\lfloor\hat{K}\rfloor + 1)}{p_p^*(\lfloor\hat{K}\rfloor) - p_p^*(\lfloor\hat{K}\rfloor + 1)} q_v(p_p^*(\hat{K}), \lfloor\hat{K}\rfloor)$$
$$+ \frac{p_p^*(\lfloor\hat{K}\rfloor) - p_p^*(\hat{K})}{p_p^*(\lfloor\hat{K}\rfloor) - p_p^*(\lfloor\hat{K}\rfloor + 1)} q_v(p_p^*(\hat{K}), \lfloor\hat{K}\rfloor + 1), \quad (12)$$

where $\lfloor\hat{K}\rfloor$ represents the largest integer below $\hat{K}$.

When $p_p^*(\hat{K})$ and $q_{vp}^*(\hat{K})$ functions are designed according to (10), (11), and (12), we have the following two monotonicity properties. On one hand, given K, the channel contention measure function $q_v(p_p^*(\hat{K}), K)$ is monotonically non-decreasing in $\hat{K}$ [5, Theorem 3]. On the other hand, the theoretical channel contention measure function $q_{vp}^*(\hat{K})$ is monotonically non-increasing in $\hat{K}$ and is strictly decreasing in $\hat{K}$ for $\hat{K} \geq \hat{K}_{p\min}$ [5, Theorem 4]. The basic considerations behind the design of $p_p^*(\hat{K})$ and $q_{vp}^*(\hat{K})$ functions can be briefly explained as follows [5]. Under the assumption that the system only contains primary users, setting $p_p^*(\hat{K})$ at $p_p^*(\hat{K}) \approx \frac{x_p^*}{\hat{K}}$ is asymptotically optimal (or close to optimal for large $\hat{K}$) in terms of symmetric utility maximization. It is also a general observation that setting $p_p^*(\hat{K})$ at $p_p^*(\hat{K}) \approx \frac{x_p^*}{\hat{K}}$ should be not far from optimal for all $\hat{K}$ values and for most of the utility functions of interest. If we term $p_p^*(\hat{K}) \approx \frac{x_p^*}{\hat{K}}$ the ideal solution, then the proposed design given in (10) should be close to ideal, with the necessary revisions to achieve the desired monotonicity properties required for the convergence proof of the distributed MAC algorithm [5, Theorem 5].

Next, let us switch focus to a secondary user. A secondary user should also design two key functions, denoted respectively by $p_s^*(\hat{K})$ and $q_{vs}^*(\hat{K})$, both are functions of $\hat{K}$ which represents the estimated number of users under the assumption that the system should only contain secondary users. Differs from the design of a primary user, $p_s^*(\hat{K})$ and $q_{vs}^*(\hat{K})$ functions need to be designed to enforce the hierarchical user structure. We will show later that, a hierarchical user structure can be achieved by raising the tail of the $q_{vs}^*(\hat{K})$ function above a pre-determined threshold $\underline{q_v}$, i.e. by imposing the following constraint.

$$\lim_{\hat{K}\to\infty} q_{vs}^*(\hat{K}) \geq \underline{q_v}. \tag{13}$$

The value of $\underline{q_v}$ should be chosen according to the quality of service requirement of the primary users. To satisfy the constraint, let x_s^* be obtained by solving the following equation.

$$\lim_{\hat{K}\to\infty} q_v\left(\frac{x_s^*}{\hat{K}}, \hat{K}\right) = \underline{q_v}, \tag{14}$$

where $q_v\left(\frac{x_s^*}{\hat{K}}, \hat{K}\right)$ is further defined in (6). A secondary user should design its desired transmission probability function $p_s^*(\hat{K})$ as

$$p_s^*(\hat{K}) = \frac{x_s^*}{\max\{\hat{K}, \hat{K}_{s\min}\} + b_s}, \tag{15}$$

where $\hat{K}_{s\min}$ and b_s are design parameters whose values should be determined by following the guideline given in [5]. Particularly, $\hat{K}_{s\min}$ should take an integer value slightly less than x_s^*, and b_s should satisfy an effective constraint of $b_s > 1$ [5, Theorem 4]. Similar to the design of a primary user, a secondary user should design the "theoretical channel contention measure" function $q_{vs}^*(\hat{K})$ as

$$q_{vs}^*(\hat{K}) = \frac{p_s^*(\hat{K}) - p_s^*(\lfloor\hat{K}\rfloor + 1)}{p_s^*(\lfloor\hat{K}\rfloor) - p_s^*(\lfloor\hat{K}\rfloor + 1)} q_v(p_s^*(\hat{K}), \lfloor\hat{K}\rfloor) + \frac{p_s^*(\lfloor\hat{K}\rfloor) - p_s^*(\hat{K})}{p_s^*(\lfloor\hat{K}\rfloor) - p_s^*(\lfloor\hat{K}\rfloor + 1)} q_v(p_s^*(\hat{K}), \lfloor\hat{K}\rfloor + 1), \tag{16}$$

where $\hat{K}$ can take both integer and non-integer values.

According to [5, Theorem 3], given K, $q_v(p_s^*(\hat{K}), K)$ should be monotonically non-decreasing in $\hat{K}$. Meanwhile, according to [5, Theorem 4], $q_{vs}^*(\hat{K})$ should be monotonically non-increasing in $\hat{K}$ and should be strictly decreasing in $\hat{K}$ for $\hat{K} \geq \hat{K}_{s\min}$. Due to (14), we have $q_{vs}^*(\hat{K}) \geq \underline{q_v}$ for all $\hat{K}$.

With the above design, the distributed MAC algorithm supports the hierarchical primary-secondary user structure in the following sense.

Theorem 1: Let K_p be the number of primary users in the system. The value of K_p is unknown to the users as well as to the receiver. With the proposed MAC algorithm, the system should possess a unique equilibrium. Let channel contention measure at the equilibrium be denoted by q_v. On one hand, if K_p is small such that $q_{vp}^*(K_p) \geq \underline{q_v}$, then $q_v \geq \underline{q_v}$ must hold at the equilibrium. On the other hand, if K_p is large such that $q_{vp}^*(K_p) < \underline{q_v}$, then transmission probabilities of the secondary users should equal zero at the equilibrium.

Proof: According to stochastic approximation theory [7] [8], the system should have at least one equilibrium.

We first show that the equilibrium must be unique. Assume that this is not true. Let the system contain two equilibria, whose corresponding channel contention measures equal q_v and $\tilde{q}_v$, respectively. Without loss of generality, we assume that $q_v < \tilde{q}_v$. Assume that, at the first equilibrium corresponding to channel contention measure q_v, the number of users estimated by the primary users and by the secondary users equal respectively $\hat{K}_p$ and $\hat{K}_s$. At the other equilibrium corresponding to channel contention measure $\tilde{q}_v$, let the estimates equal $\tilde{K}_p$ and $\tilde{K}_s$, respectively. Consequently, we have

$$\begin{aligned} q_v &= q_{vp}^*(\hat{K}_p) = q_{vs}^*(\hat{K}_s), \\ \tilde{q}_v &= q_{vp}^*(\tilde{K}_p) = q_{vs}^*(\tilde{K}_s). \end{aligned} \tag{17}$$

Because $q_v < \tilde{q}_v$, due to the fact that $q_{vp}^*(\hat{K})$ and $q_{vs}^*(\hat{K})$ functions are non-increasing in $\hat{K}$ [5, Theorem 4], (17) implies that $\hat{K}_p \geq \tilde{K}_p$ and $\hat{K}_s \geq \tilde{K}_s$. This consequently implies that $p_p^*(\hat{K}_p) \leq p_p^*(\tilde{K}_p)$ and $p_s^*(\hat{K}_p) \leq p_s^*(\tilde{K}_p)$. However, if each user at the first equilibrium should transmit at a probability no higher than the corresponding probability at the other equilibrium, we must have $q_v \geq \tilde{q}_v$, which contradicts the assumption that $q_v < \tilde{q}_v$. Therefore, equilibrium of the system must be unique.

Let q_v be the channel contention measure at the unique equilibrium. Next, we prove the following statement, which is equivalent to the conclusion of the theorem. That is, if $q_v < \underline{q_v}$, we must have $q_{vp}^*(K_p) < \underline{q_v}$. Otherwise if $q_v \geq \underline{q_v}$, we must have $q_{vp}^*(K_p) \geq \underline{q_v}$.

According to the proposed MAC algorithm, if $q_v < \underline{q_v} \leq q_{vs}^*(\hat{K})$ for all $\hat{K}$, we must have $\hat{K}_s = \infty$ and all secondary users should have zero transmission probability at the equilibrium. Consequently, the system becomes equivalent to one with homogeneous (primary) users, as analyzed in [5]. According to [5, Theorem 5], we should have

$$q_v = q_v(p_p^*(K_p), K_p) = q_{vp}^*(K_p). \tag{18}$$

In other words, primary users should obtain a correct estimate of the number of users $\hat{K}_p = K_p$. This implies that $q^*_{vp}(K_p) = q^*_{vp}(\hat{K}_p) = q_v < \underline{q_v}$.

If $q_v \geq \underline{q_v}$, on the other hand, we have $q_v = q^*_{vp}(\hat{K}_p)$. In this case, $\hat{K}_s < \infty$, meaning that secondary users should transmit with a positive probability. Now assume that we force all secondary users to exit the system. This action should help increasing the value of q_v at the new equilibrium. We know that, without the secondary users, contention measure of the new system equilibrium should equal $q_v(p^*_p(K_p), K_p) = q^*_{vp}(K_p)$. Consequently, we must have

$$\underline{q_v} \leq q_v \leq q_v(p^*_p(K_p), K_p) = q^*_{vp}(K_p). \tag{19}$$

$\Diamond$

IV. Simulation Results

The example presented in this section is extended from [5, Example 2]. Consider a random multiple access system over a simple fading channel. In each time slot, with a probability of 0.3, the channel can support no more than $M_1 = 4$ parallel real packet transmissions, and with a probability of 0.7, the channel can support no more than $M_2 = 6$ parallel real packet transmissions. In this case, the real channel parameter set $\{C_{rj}\}$ is given by $C_{rj} = 1$ for $j < 4$, $C_{rj} = 0.7$ for $4 \leq j < 6$, and $C_{rj} = 0$ for $j \geq 6$. We design the virtual packet to have the same coding details of a real packet. Consequently, the virtual channel parameter set $\{C_{vj}\}$ is identical to the real channel parameter set, i.e., $C_{vj} = C_{rj}$ for all $j \geq 0$.

Assume that the primary users intend to maximize the symmetric throughput weighted by a transmission energy cost of of $E = 0.3$, under the assumption that the system only contains primary users. With the number of users being K and all users transmitting with the same probability p, the utility function of the primary users is given by

$$U(K, p, \{C_{rj}\}) = -EKp + \sum_{j=0}^{K-1} K \binom{K-1}{j} p^{j+1}(1-p)^{K-1-j} C_{rj}. \tag{20}$$

As explained in [5, Example 2], with this utility function, primary users can design their theoretical transmission probability function as $p^*_p(\hat{K}) = \frac{3.29}{\max\{\hat{K},3\}+1.01}$, which implies $x^*_p = 3.29$, $\hat{K}_{p\,\min} = 3$, and $b_p = 1.01$. Note that the value of x^*_p is determined using (9). Such a design is near optimal in terms of maximizing the chosen utility when the system only contains homogeneous primary users.

Next, let us consider the secondary users. Assume that the hierarchical user structure requires $\underline{q_v}$ to be set at $\underline{q_v} = 0.88$. This gives $x^*_s = 2.655$ according to (14). Consequently, we can design the theoretical transmission probability function of the secondary users as $p^*_s(\hat{K}) = \frac{2.655}{\max\{\hat{K},2\}+1.01}$. This implies that $b_s = 1.01$ and $K_{s\,\min} = 2$.

In Figure 1, we plotted the theoretical channel contention measure functions $q^*_{vp}(\hat{K})$ for primary users and $q^*_{vs}(\hat{K})$ for secondary users. These functions are calculated using (11), (12),

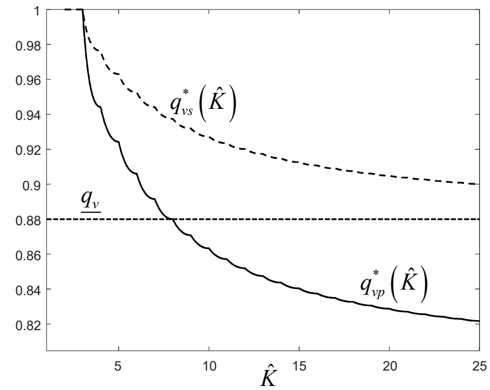

Fig. 1. Theoretical channel contention measure functions for primary and secondary users.

and (16). Note that $q^*_{vs}(\hat{K}) \geq \underline{q_v} = 0.88$. It can be seen that, the key idea of supporting the hierarchical user structure is to raise the tail of the $q^*_{vs}(\hat{K})$ function for the secondary users above $\underline{q_v}$, such that aggregated impact of the secondary users on the channel contention measure is well controlled no matter how many secondary users compete for the channel.

In Figure 2, we plotted channel contention measure of the system at its unique equilibrium as a function of the number of primary users K_p and the number of secondary users K_s. The figure shows that, when the number of primary users is

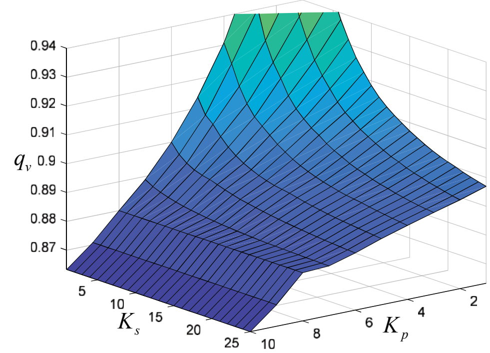

Fig. 2. Channel contention measure as a function of the numbers of primary and secondary users.

small $K_p \leq 7$, we have $q^*_{vp}(K_p) > \underline{q_v} = 0.88$. In this case, secondary users can access the channel. But the system keeps the channel contention measure above $\underline{q_v} = 0.88$ irrespective of the number of secondary users. When the number of primary users is large $K_p > 7$, on the other hand, we have $q^*_{vp}(K_p) < \underline{q_v}$. In this case transmission probabilities of the secondary users are kept at zero, and therefore q_v is not affected by the number of secondary users.

Next, we assume that the transmission probabilities of all users are initialized at 0. In each time slot, a user randomly determines whether to transmit a packet or not. The receiver uses an exponential moving average approach to measure q_v. More specifically, q_v is initialized at $q_v = 1$. In each time slot, an indicator variable $I_v \in \{0, 1\}$ is used to represent the success/failure status of the virtual packet reception. q_v is then updated as $q_v = (1 - \frac{1}{300})q_v + \frac{1}{300}I_v$, and is fed back to the transmitters at the end of each time slot. With the updated q_v, each user adapts its transmission probability according to the MAC algorithm proposed in Section III with a constant step size of $\alpha = 0.05$.

We assume that the system experiences three stages. At the beginning in Stage one, the system has 4 primary users and 3 secondary users. The system enters Stage two after the 3000th time slot, when 12 more secondary users enter the system with their transmission probabilities initialized at 0. After the 6000th time slot, the system enters Stage three when 6 more primary users enter the system with their transmission probabilities initialized at 0. Convergence behavior in actual channel contention measure q_v is illustrated in Figure 3 together with the theoretical q_v at the corresponding equilibria of the three stages. The figure demonstrates that the system can quickly

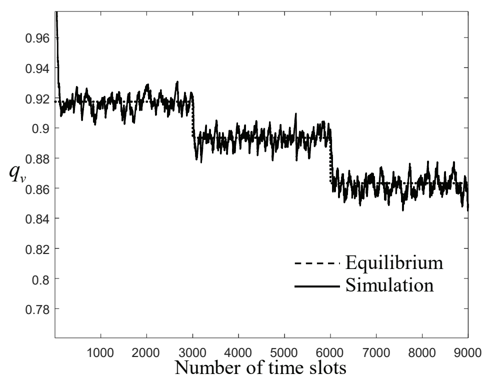

Fig. 3. Channel contention measure of the system through three stages.

adapt to changes in the number of users and keep the channel contention at the desired level. In Figure 4, we also illustrated the transmission probability targets calculated by the primary and the secondary users. Note that values of the simulated

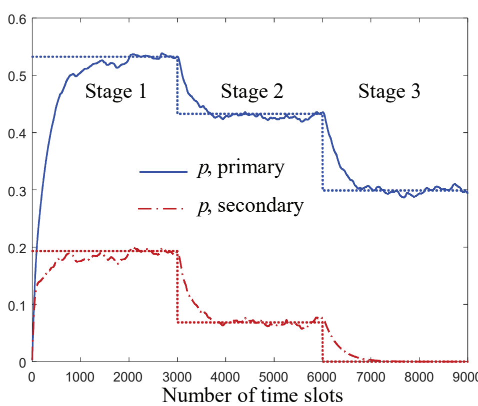

Fig. 4. Transmission probabilities of primary and secondary users through three stages. Dashed lines represent the corresponding values at the equilibrium.

variables presented in Figures 3 and 4 are calculated using the same exponential averaging approach explained above.

It can be seen that, in Stage 1 when a small number of primary and secondary users share the channel, the primary users transmit with a relatively higher probability. Consequently, a primary user gets an advantage over a secondary user in getting its packet through the channel. In Stage 2 when a large number of secondary users join the system, while their activities lead to lower transmission probabilities for both the primary users and the secondary users, channel contention measure of the system is maintained above $\underline{q_v} = 0.88$, which can be viewed as a quality of service guarantee to the primary users. In Stage 3 when a significant number of primary users enter the system, transmission probabilities of all secondary users are driven down to 0, leading the channel to be exclusively occupied by the primary users.

V. Conclusion

We proposed a distributed MAC framework to support hierarchical user groups in random multiple access systems. The MAC algorithms do not require direct message exchange among users. Users do not need to know the numbers of primary and secondary users in the system. Users also do not need the capability of identifying whether a transmitted packet should belong to a primary user or to a secondary user. The proposed MAC algorithm adapts the transmission scheme of each user by comparing the actual channel contention measure to a theoretical channel contention measure function. With the simple idea of raising the tail of the theoretical channel contention measure function for the secondary users to a predetermined threshold, aggregated impact of the secondary users on contention level of the channel is well controlled no matter how many secondary users compete for the channel. Simulation results showed that the proposed MAC algorithm can maintain the hierarchical user structure and can also be reasonably responsive to a dynamic environment with users joining/exiting the system.

References

[1] M. Khoshkholgh, K. Navaie, and H. Yanikomeroglu, *Access Strategies for Spectrum Sharing in Fading Environment: Overlay, Underlay, and Mixed,* IEEE Trans. Mobile Comp., Vol. 9, pp. 1780-1793, Dec. 2010.

[2] A. Domenico, E. Strinati, and M. Benedetto, *A Survey on MAC Strategies for Cognitive Radio Networks,* IEEE Commun. Surveys & Tutorials, Vol. 14, pp. 21-44, First Quarter, 2012.

[3] J. Luo, *A Generalized Channel Coding Theory for Distributed Communication,* IEEE Trans. Commun., Vol. 63, pp. 1043-1056, Apr. 2015.

[4] Y. Tang, F. Heydaryan, and J. Luo, *Distributed Coding in A Multiple Access Environment,* Foundations and Trends in Networking, Vol. 12, No. 4, pp. 260-412, Now Publishers, 2018.

[5] Y. Tang, F. Heydaryan, and J. Luo, *Distributed Multiple Access with A General Link Layer Channel,* Ad Hoc Networks, Vol. 85, pp. 120-130, Mar. 2019.

[6] F. Heydaryan and J. Luo, *Random Multiple Access with Hierarchical Users,* in IEEE Trans. Wireless Commun., doi: 10.1109/TWC.2021.3075451.

[7] H. Kushner and G. Yin, Stochastic Approximation Algorithms and Applications, New York: Springer-Verlag, 1997.

[8] V. Borkar and S. Meyn. The O.D.E Method for Convergence of Stochastic Approximation and Reinforcement Learning, SIAM J. Contrl. and Opt., Vol. 38, pp. 447-469, Jan. 2000.