Abstract—This paper considers cross-layer medium access control (MAC) protocol design in wireless networks. Taking a mutually interactive MAC-PHY perspective, we aim to design an MAC protocol that is in favor of the PHY layer information transmission, and the improved PHY, in turn, can improve the MAC performance. Motivated by the fact that as long as good channel estimation can be achieved, advanced signal processing does allow effective signal separation given that the multiuser interference is limited to a certain degree, we propose a novel MAC protocol, named hybrid ALOHA, which allows collision-free channel estimation and makes it possible for simultaneous multiuser transmission. Comparing with traditional ALOHA, there are more than one pilot subslots in each hybrid ALOHA slot. Each user randomly selects a pilot subslot for training sequence transmission. Therefore, it is possible for different users to transmit their training sequences over non-overlapping pilot subslots and achieving collision-free channel estimation. Relying mainly on the general multipacket reception (MPR) model, in this paper, quantitative analysis is conducted for the proposed hybrid ALOHA protocol in terms of throughput, stability as well as delay. It is observed that significant performance improvement can be achieved in comparison with the traditional ALOHA protocol based either on the collision model or the MPR model.

Index Terms—Cross-layer design, hybrid ALOHA, MPR model, throughput, stability region, delay performance.

I. INTRODUCTION

Over the past decades, studies on medium access control (MAC) have largely relied on the “collision” model: a packet is successfully received if and only if there are no concurrent transmissions [1]–[3]. MAC protocols such as traditional ALOHA [1], [4], the tree algorithms [5], the window random-access algorithm [6] and a number of adaptive MAC schemes [7]–[9], have all been developed by assuming that the physical (PHY) layer is characterized as an idealized collision model [10]. Although it is convenient for analysis, the collision model fails to represent all the characteristics of the PHY layer. As pointed out by Tong et al., in [11], the model is optimistic as it ignores the channel effects such as fading and noise on reception, whereas the model is also pessimistic since it overlooks the possibility that the packets may be successfully decoded even in the presence of simultaneous multiuser transmissions.

Going beyond the collision model and assuming a multiuser PHY layer, Ghez, Verdú and Schwartz proposed the multipacket reception (MPR) model in [12], [13], where the reception of packets is characterized by conditional probabilities instead of deterministic failures or successes. The MPR model provides a more realistic interface between the PHY layer and the MAC layer, and propels the research on cross-layer MAC design. In [14]–[17], the impact of MPR on the performance of slotted ALOHA is studied by investigating the capture model, which, being a subclass of the MPR, assumes that at most one user’s packet can be successfully decoded when multiple users transmit at the same time. More examples on protocol design for networks in the presence of capture can be found in [18]–[21]. Based on the general MPR model beyond the capture channel, in [10] and [22], a multi-queue service room (MQSR) protocol and a simpler dynamic queue protocol have been developed to adaptively grant channel access to users, respectively.

One limitation with the MPR model of Ghez et al. is that it assumes a symmetric model with indistinguishable users. To differentiate users in a multimedia networks where each user may have his own rate, in [23], Naware, Mergen and Tong introduce a more general asymmetric MPR model as the interface between the MAC layer and the PHY layer. Based on the generalized MPR model, they investigate the stability and delay performance of slotted ALOHA. Their study leads to an interesting result: the burden at the MAC layer can be reduced significantly if we can ensure a reasonably good PHY layer.

As is well known, being an important characteristic that directly influences the quality of the PHY layer, the channel state information (CSI) plays a critical role in signal detection and estimation. The more accurate the CSI estimation is, the more likely that the signal can be recovered at the receiver, and hence the stronger the MPR capability. There has been a line of work studying the effect of CSI on resource allocation, e.g., [24]–[27]. Assumptions are usually made that either transmitters or receivers can track the channel state, and the knowledge of such information is employed to improve the network capacity. In other words, existing works have been focused on improving the MAC performance by exploiting the PHY layer transmission capability. Here, we turn to examine the effect of the MAC protocol design on the PHY layer performance, and study how the mutual MAC-PHY interaction will influence the overall system performance. More specifically, as illustrated in Fig. 1, taking a mutually interactive MAC-PHY perspective, we aim to design an MAC protocol that is in favor of the PHY layer information transmission, and the improved PHY layer, in turn, can improve the performance in terms of throughput, stability and delay.

In this paper, based on the fact that as long as good channel estimation can be achieved, advanced signal processing does allow effective signal separation given that the multiuser interference is limited to a certain degree, we propose a hybrid ALOHA protocol for more efficient random access control. Comparing with conventional ALOHA, in the hybrid ALOHA...
The rest of the paper is organized as follows. In Section II, the system model and the proposed hybrid ALOHA are presented. In Section III, the throughput is analyzed based on hybrid ALOHA. In Section IV, the stability region is first derived for the two-user case, then the sufficient condition for stability is obtained for the general finite-user system. The delay behavior is studied in Section V. An illustrative example is presented in Section VI and we conclude in Section VII.

II. THE HYBRID ALOHA PROTOCOL

A. System Model

Consider a wireless network with a set $\mathcal{N}$ of users, $\mathcal{N} = \{1, 2, \cdots, N\}$, communicating with a common access point. Each user is equipped with an infinite buffer for storing arriving and backlogged packets. The packet arrival processes are assumed to be independent from user to user. The channel is slotted in time, with slot period larger than the packet length. When the buffer of the $i$th ($i \in \mathcal{N}$) user is nonempty, he tosses a coin to determine whether to transmit in the current slot or not. The probability of transmission of the $i$th user at each slot is assumed to be $p_i$. Packets are assumed to be of equal size for all users and composed of two parts: the first part is the training sequence for channel estimation, and the second part is the information data. The length of the training sequence is typically much smaller than that of the information data. The arrivals of the $i$th user are assumed to be independent and identically distributed (i.i.d) Bernoulli random variables from slot to slot, with the average number of arrivals being $\lambda_i$ packets per slot. If this arrival rate is measured in packets per unit time, we then use $\Lambda_i$ to denote it. Throughout the paper, we use the slot length of traditional ALOHA as the reference time unit and denote it as $1$. Note that the slot length is protocol-dependent and may not always be 1.

We adopt the general MPR model in [23] where the multiuser PHY layer is characterized by a set of conditional probabilities. For any subset $\mathcal{S} \subseteq \mathcal{N}$ of users transmitting in a slot, the marginal probability of successfully receiving packets from users in $\mathcal{R} \subseteq \mathcal{S}$, given that users in $\mathcal{S}$ transmit, is defined as

$$q_{\mathcal{R}\mathcal{S}} = \sum_{\mathcal{U}: \mathcal{R} \subseteq \mathcal{U} \subseteq \mathcal{S}} q_{\mathcal{U}, \mathcal{S}},$$

with

$$q_{\mathcal{U}, \mathcal{S}} = Pr\{\text{only packets from } \mathcal{U} \text{ are successfully received } | \text{ users in } \mathcal{S} \text{ transmit}\}, \mathcal{U} \subseteq \mathcal{S}.$$  \hspace{1cm} (2)

Clearly, $q_{\mathcal{R}\mathcal{S}}$ depends on the quality of the PHY layer.

Assume that at the end of each slot, the receiver gives an instantaneous feedback of all the packets that were successfully received to all the users. The users remove successful packets from their buffers while unsuccessful packets are retained. Let $N_{t+1}^i$ denote the queue length of the $i$th user at the beginning of time slot $t$, the queue evolution function for the $i$th ($i \in \{1, 2, \cdots, N\}$) queue is given by [23], [28]

$$N_{t+1}^i = [N_t^i - Y_t^i]^+ + \beta_t^i,$$

where $\beta_t^i$ is the number of arrivals during the $t$th slot to the $i$th user with $E(\beta_t^i) = \lambda_t < \infty$, $Y_t^i$ is the Bernoulli random variable.
variable denoting the number of departures from queue \(i\) in time slot \(t\), and \([x]^+ = \max(0,x)\).

### B. Hybrid ALOHA

The proposed hybrid ALOHA protocol aims at improving MPR capability by allowing conditional collision-free channel estimation and simultaneous transmission. In hybrid ALOHA, each slot contains the data subslot and multiple pilot subslots, and each user can randomly select a pilot subslot to transmit his/her training sequence. In other words, idle pilot subslot(s) are introduced to make it possible for different users to transmit their training sequences at non-overlapping subslots, whereby collision-free channel estimation could be achieved. If the physical layer can accommodate \(M\) users, (i.e., given reasonably accurate channel estimation, the user’s packet can be successfully decoded if and only if there are no more than \(M\) simultaneous users), then the hybrid ALOHA slot has \(M\) pilot subslots, this implies that \(M-1\) idle sections are inserted to each traditional ALOHA slot.

![Fig. 2. Illustration of the hybrid ALOHA slot structure for \(M = 2\). Each user randomly selects a pilot subslot for training sequence transmission at the beginning of each slot.](image)

Fig. 2 illustrates the slot structure of hybrid ALOHA in the case of \(M = 2\), which implies that successful transmission is possible only when there are no more than two simultaneous users. Each slot has \(M+1=3\) subslots. The preceding two subslots, each having a length of \(\tau\), are the “pilot subslots” reserved for training sequences. When a user is involved in a transmission, we assume that the selection of the pilot subslots is of equal probability. As shown in Fig. 2, when two users transmit over the same time slot, then the probability that they will transmit their training sequences over non-overlapping pilot subslots is 0.5. The information data is always transmitted in the “data subslot”. We assume that the length of the data subslot is \(1-\tau\) with \(\tau \ll 1\). Recall that the length of the traditional ALOHA slot is used as the reference time unit, denoted as 1, which consists of a training subslot of length \(\tau\) and a data subslot of length \(1-\tau\).

In general, our analysis in this paper is based on the MPR model as described in Section II.A. At the same time, for results illustration and simplification of the analysis, we also introduce a simplified model: Given that there are no more than \(M\) users transmitting simultaneously in one slot, users who experience collision in the pilot subslots fail in transmission, whereas those who have collision-free channel estimation survive. In the following sections, this simplified model is referred to as the **Simplistic Assumption**. It is mainly used in the stability region analysis for simplification of the derivation.

### III. THROUGHPUT CALCULATION

As a figure of merit, the throughput of the system, which is a function of \(\tau\) and denoted here as \(\nu(\tau)\), is calculated below. The throughput is determined by the average traffic successfully getting through the channel, i.e., \(\nu(\tau) = E[T_i]\), where \(E[\cdot]\) is the expectation operator and \(T_i\) denotes the number of packets successfully getting through the channel during slot \(t\). Let \(\Lambda\) be the overall arrival rate to the system, in the unit of **number of packets per unit time**. That is, \(\Lambda = \sum_{i=1}^{N} \Lambda_i\), where \(\Lambda_i\) is the arrival rate of the \(i\)th user measured in packets per unit time. The average traffic per slot is \(\nu(\tau) = \Lambda(M\tau + 1 - \tau)\). Let \(A_S\) be the event that users in \(S\) transmit, and \(A_{R_S}\) the event that only users in \(R\) succeed given that users in \(S\) transmit. The system throughput under the general MPR model is given by

\[
\nu(\tau) = \sum_{S \subseteq \mathcal{N}} E[T_i | A_S] \cdot Pr\{A_S\}
\]

\[
= \sum_{S \subseteq \mathcal{N}} \sum_{R \subseteq S} E[T_i | A_{R_S}] \cdot Pr\{A_{R_S}\} \cdot Pr\{A_S\}
\]

\[
= \sum_{S \subseteq \mathcal{N}} \sum_{R \subseteq S} R(\tau) \frac{|R|}{|S|} \cdot q_{R,S} \cdot Pr\{A_S\},
\]

(4)

Here \(|R|\) and \(|S|\) denote the number of elements in sets \(R\) and \(S\), respectively, and it follows from (2) that \(Pr\{A_{R_S}\} = q_{R,S}\).

For the symmetric MPR model where users are indistinguishable, a more tractable form can be obtained. Denote \(A_K\) as the event that \(K\) users transmit in one slot and denote \(A_K^i\) as the event that exactly \(i\) out of \(K\) users succeed in transmission, then

\[
\nu(\tau) = \sum_{K=1}^{M} E[T_i | A_K] \cdot Pr\{A_K\}
\]

\[
= R(\tau) \sum_{K=1}^{M} \sum_{i=1}^{K} \frac{i}{K} \cdot Pr\{A_K^i | A_K\} Pr\{A_K\},
\]

(5)

where \(Pr\{A_K^i | A_K\}\) is determined by the MPR capability of the system, and \(Pr\{A_K\}\) relies on the packet arrival model. We have assumed that the packet arrivals for the users follow the Bernoulli processes, and the overall arrival rate is \(\Lambda\). If the number of users in the system is sufficiently large, the binomial distribution can be closely approximated using the Poisson distribution [29]. The closed-form expression of \(Pr\{A_K\}\) can be obtained by assuming that the packets arrival follows the Poisson distribution with parameter \(R(\tau) = \Lambda(M\tau + 1 - \tau)\).

The result is summarized in the following proposition.

**Proposition 1**: Under the symmetric MPR model, as well as the Poisson approximation with parameter \(R(\tau) = \Lambda(M\tau + 1 - \tau)\), the throughput \(\nu(\tau)\) of hybrid ALOHA, measured in packets per hybrid ALOHA slot, is given by

\[
\nu(\tau) = R(\tau) \sum_{K=1}^{M} \sum_{i=1}^{K} \frac{i}{K} \cdot Pr\{A_K^i | A_K\} \frac{R(\tau)^{K-1} e^{-R(\tau)}}{(K-1)!}.
\]

(6)

The expression of \(\nu(\tau)\) can be further simplified under the **Simplistic Assumption**. In fact, when \(M = 1\), Eq. (6) becomes

\[
\nu(\tau) = R(\tau) e^{-R(\tau)},
\]

which corresponds to the traditional slotted ALOHA system. For \(M = 2\), assuming users are symmetric, it is then easy to know that \(Pr\{A_1 | A_1\} = 1\), \(Pr\{A_2 | A_1\} = 0\) and \(Pr\{A_2 | A_2\} = 1/2\). Thus,

\[
\nu(\tau) = \left[ R(\tau)^2/2 + R(\tau) \right] e^{-R(\tau)}.
\]

(7)
Similarly, when $M = 3$, we have $\nu(\tau) = R(\tau)e^{-R(\tau)}[1 + \frac{2}{3}R(\tau) + \frac{2}{9}R(\tau)^2]$.

**Remark 1**: Clearly, the throughput of the system is a function of $\tau$. Increasing $\tau$ allows a longer training sequence, thereby resulting in better channel estimation. More specifically, as will be illustrated in Section VI, the mean-squared error of channel estimation is inversely proportional to $\tau$. Hence, increasing $\tau$ will result in improved MPR capability, which is desirable to improve the overall system performance. On the other hand, however, a larger $\tau$ also implies longer idle periods, leading to spectral inefficiency. Therefore, a tradeoff should be made on the design of $\tau$ for different applications. Please refer to Section VI for a specific example.

Fig. 3 compares the throughput of Hybrid ALOHA (under the Simplistic Assumption) for $M = 1, 2, 3$ with that of traditional ALOHA. Here we choose $\tau = 0.05$. The reference time unit is 1, which is the length of one slot in the case of $M = 1$, corresponding to the traditional ALOHA scheme. Note that in Proposition 1, the throughput is measured in the unit of “packets” per “hybrid ALOHA slot”, with slot length $[1 + (M - 1)\tau]$. For throughput comparison with traditional ALOHA, we need to normalize the throughput of hybrid ALOHA by the slot length. In other words, the throughput is normalized to “number of packets per unit-time”, which is $\nu(\tau)/[1 + (M - 1)\tau]$. It is shown that when $M = 2$, the proposed scheme has a 46% gain over traditional ALOHA ($M = 1$) in terms of maximum throughput, and the gain doubles for the $M = 3$ case. That is, the throughput tends to increase as $M$ increases. At the point where the maximum throughput is achieved, the proposed protocol has an expected number of attempting transmissions approximately 1.4 and 1.95 for $M = 2$ and $M = 3$, respectively, much larger than that of traditional slotted ALOHA which takes on the value 1. This can be intuitively understood as: the larger the value of $M$, the stronger the MPR capability. As will be seen in Section VI, due to the small length of $\tau$, the cost of extra bandwidth can be compensated for by the improved MPR capability.

Throughput comparison of the two protocols beyond the Simplistic Assumption is illustrated in Section VI through a practical example.

Without loss of generality, in what follows, the stability regions and the delay performance of the hybrid ALOHA system are derived for the specific case of $M = 2$, that is, at most two simultaneous users can transmit in one slot in order for successful reception.

**IV. STABILITY FOR HYBRID ALOHA**

**Definition 1**: [28] A multidimensional stochastic process $N^t = (N^t_1, \ldots, N^t_N)$ is said to be stable if for $x \in \mathbb{N}^N$ the following holds

$$\lim_{t \to \infty} P_r\{N^t < x\} = F(x) \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1,$$

where $F(x)$ is the limiting distribution function.

For an $N$-user slotted ALOHA system, the stability region ($\mathcal{S}_{ALOHA}$) is defined as the set of arrival rate $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_N]$ for which there exists a transmission probability vector $p = [p_1, p_2, \ldots, p_N]$ such that the queues in the system are stable. In literature, studies on the stability conditions of ALOHA systems were mainly based on the collision model, see [28], [30]–[33], for example. In [23], the stability analysis of slotted ALOHA based on the general MPR model is investigated. In this section, under the Simplistic Assumption, we first analyze the stability region of hybrid ALOHA for the two-user case and then derive the sufficient condition for stability of hybrid ALOHA in the case of $N > 2$.

**A. Stability Region for $N = 2$**

When $N = 2$, suppose the arrival rates for the two users are $\lambda_1$ and $\lambda_2$ (packets per slot), and their transmission probabilities are $p_1$ and $p_2$, respectively. Adopting the stochastic dominance method used in [23], we can obtain the following result.

**Lemma 1**: Under the Simplistic Assumption, for a fixed transmission probability vector $p = [p_1, p_2]$, the stability region of hybrid ALOHA is given by

$$\lambda_1 \leq p_1 - \frac{p_1\lambda_2}{2 - p_1} \quad \text{for} \quad \lambda_2 \leq p_2 - p_1p_2/2,$$

and

$$\lambda_2 \leq p_2 - \frac{p_2\lambda_1}{2 - p_2} \quad \text{for} \quad \lambda_1 \leq p_1 - p_1p_2/2,$$

where $\lambda_1$ and $\lambda_2$ are the arrival rates for the two users in the unit of packets per slot.

The proof can be obtained following the lines in [23] and is omitted here.

Let $\mathcal{S}_{ALOHA}(p)$ be the stability region of hybrid ALOHA for a fixed $p$, then the overall stability region of hybrid ALOHA can be characterized as

$$\mathcal{S}_{ALOHA} = \bigcup_{p \in [0,1]^N} \mathcal{S}_{ALOHA}(p).$$

By taking the closure of all the regions as $p$ varies over $[0,1]^2$, we can obtain the stability region of the proposed scheme. In [23], an alternative approach is used to find the stability region.
by solving a corresponding constraint optimization problem. The approach is adopted here and we obtain the following result.

**Proposition 2:** Under the Simplistic Assumption, the stability region of hybrid ALOHA coincides with that of a TDMA scheme which is characterized by the region below

$$\mathcal{S}_{H-ALOHA} = \{ (\lambda_1, \lambda_2) : (\lambda_1, \lambda_2) \geq (0, 0), (\lambda_1, \lambda_2) \text{ lies below } \lambda_1 + \lambda_2 = 1, \ 0 \leq \lambda_1 \leq 1 \},$$

where $\lambda_1$ and $\lambda_2$ are the arrival rates for the two users in the unit of packets per slot.

**Remark 2:** Beyond the Simplistic Assumption, in the case when the collision of the training sequences does not necessarily lead to reception failure(s), stronger MPR capability could be achieved. In such cases, the stability region of hybrid ALOHA can be characterized by Lemma 2 in [23], which is likely to outperform that of TDMA schemes. More details can be found in Section VI.

**B. Sufficient Condition of Stability for $N > 2$**

In the case of $N > 2$, stability regions are difficult to obtain for general ALOHA systems even under the collision model. However, in literature, stability conditions for a fixed transmission probability vector $p$ have been developed under various models. Szpanski [28] derived sufficient and necessary conditions for stability of the ALOHA system with the collision model. In [23], Naware, Mergen and Tong provided the sufficient condition for stability of the ALOHA system under the generalized MPR model. The stochastic dominance [31] and the Loynes Theorem [34] were the keys to these derivations, which, associated with the Simplistic Assumption, are adopted here to derive the sufficient condition for stability of hybrid ALOHA.

We construct a modified system as follows. Let $\mathcal{P} = \{ \mathcal{S}, \mathcal{U} \}$ be a partition of $\mathcal{N}$ such that users in $\mathcal{S} \neq \mathcal{N}$ work exactly in the same manner as in the original system, while users in $\mathcal{U}$ persistently transmit dummy packets even if their queues are empty. Users in $\mathcal{U}$ are called persistent and those in $\mathcal{S}$ nonpersistent. Such a modified system is denoted by $\Theta^\mathcal{P}$. Let $\mathbf{N}_p = (\mathbf{N}_p^\mathcal{S}, \mathbf{N}_p^\mathcal{U})$ denote the queue lengths in $\Theta^\mathcal{P}$ and it can be proved that $\mathbf{N}_p$ stochastically dominates $\mathbf{N}$ of the original system provided that the initial conditions are identical [30], [35].

By the construction above, the process $\mathbf{N}_p^\mathcal{S}$ is an $|S|$-dimensional Markov chain that mimics the behavior of the original system. Note that the system consisting of users in $\mathcal{S}$ forms a smaller copy of the original system with modified reception probabilities. Induction arguments can then be applied to establish the stability conditions. We further assume that the Markov chain $\mathbf{N}_p^\mathcal{S}$ is stationary and ergodic, and denote the stationary version as $\mathbf{N}_p^\mathcal{S}$ to indicate that the process starts from the stationary distribution.

Let $Y_i^t(\mathcal{P})$ be the departure process from the $i$th queue in the dominant system $\Theta^\mathcal{P}$. Given that $Y_i^t(\mathcal{P})$ is stationary, we denote $P_{\text{succ}}(\mathcal{P}) = E[Y_i^t(\mathcal{P})]$ the probability of a successful transmission from the $i$th user in system $\Theta^\mathcal{P}$, which, under the Simplistic Assumption, is given in (12) at the bottom of this page. In (12), $p_{i,j}$ is the transmission probability of the training sequence at subslot $j$ for user $i$, hence the transmission probability of user $i$ in one slot is given by $p_i = \sum_{j=1}^{2} p_{i,j}$. Let $\chi(k) = 0$ for $k = 0$ and $\chi(k) = 1$ for $k > 0$. Then $\chi(\mathbf{N}_p^\mathcal{S}) = \chi(\mathbf{N}_p^\mathcal{U}) = \chi(\mathbf{N}_p^\mathcal{S})$ with $i_k \in \mathcal{S}$ for all $k = 1, \ldots, |S|$. The first term of the right-hand side of (12) represents the probability of successful transmission of user $i$ when no one else transmits. The second and third terms represent the probabilities of successful transmission of user $i$ when there is another user involved in transmission. More specifically, these two terms correspond to the cases when the other user belongs to $\mathcal{U}$ and $\mathcal{S}$, respectively.

Under the partition $\mathcal{P} = (\mathcal{S}, \mathcal{U})$, let $\mathcal{S}_\mathcal{U}$ and $\mathcal{S}_\mathcal{S}$ denote the stability regions for the whole system $\Theta^\mathcal{P}$ and the system

---

1A real random variable $X$ is said to stochastically dominate a real random variable $Y$ if $\forall z \in \mathbb{R}$, $Pr\{X > z\} \geq Pr\{Y > z\}$. This dominance is denoted by $X \succeq_Y Y$. 

---

![Fig. 4. Stability regions for different protocols in the two-user case.](image-url)
consisting of nonpersistent users $S$, respectively. Defining
\[ \mathcal{S}_N = \bigcup_{\mathcal{P}} \{ \lambda_N = \{\lambda_1, \cdots, \lambda_N\} : \lambda_k < P^k_{\text{succ}}(\mathcal{P}) \, \forall k \in \mathcal{U} \]n and $\lambda_S \in \mathcal{S}_S \},
(14)
then we have the following result.

**Proposition 3:** For a fixed transmission probability vector $p$, under the Simplistic Assumption, if $\lambda_N \in \mathcal{S}_N$, then the hybrid ALOHA system is stable, i.e., $\mathcal{S}_N \subseteq \mathcal{S}_{-\text{ALOHA}}(p)$.

Here is the sketch of the proof. Suppose $N \in \{1, 2\}$ and $\mathcal{P} = \{ S, U \} = \{\{2\}, \{1\}\}$. If $p_{1,1} = p_{1,2} = \frac{1}{2}p_1$, then from (12) we have
\[ P^2_{\text{succ}}(\mathcal{P}) = p_2(1 - p_1) + (p_2, p_1, p_2, p_1, p_1) = p_2 - p_1 p_2 / 2, \]
\[ P^1_{\text{succ}}(\mathcal{P}) = p_1 \left[ (1 - p_2) Pr(\chi(N^0_2) = 1) + \frac{1}{2} p_1 p_2 Pr(\chi(N^0_2) = 1) \right]. \]
(16)
Let $p_{2,2} = p_1 p_2 / 2$, then $\lambda_2 < P^2_{\text{succ}}(\mathcal{P})$, the probability $Pr(\chi(N^0_2) \geq 1)$ is $\lambda_2 / u_2$. We have
\[ P^1_{\text{succ}}(\mathcal{P}) = p_1 \left[ (1 - \frac{\lambda_2}{u_2}) + (1 - p_2) \frac{\lambda_2}{u_2} + \frac{1}{2} p_1 p_2 \frac{\lambda_2}{u_2} \right] = p_1 - \frac{p_1 \lambda_2}{2} - p_1. \]
(17)
Hence we obtain the stability condition identical to (9). Similarly, condition (10) can be obtained by performing the partition as $\mathcal{P} = \{ S, U \} = \{\{1\}, \{2\}\}$. The union of the regions under these two partitions characterizes the stability region given in Lemma 1. This proves Proposition 3 for the case of $N = 2$. For $N > 2$, mathematical induction can then be applied to prove the validity of Proposition 3. Please refer to [23], [28] for details.

V. DELAY PERFORMANCE FOR $N = 2$

We consider to study the delay performance of hybrid ALOHA in this section. Based on the collision model, the “exact” delays in the two-user slotted ALOHA systems with symmetric and asymmetric arrival rates and transmission probabilities are calculated in [36] and [37], respectively. In [23], the average delay is computed for the capture model in a symmetrical two-user system. It is hard to get the exact delay for $N > 2$, but bounds on the average delay have been obtained in [30], [35], [38] for the collision model, and in [39] for the capture model.

In this section, we derive the upper and lower bounds for the symmetric two-user system based on the MPR model. Let $\lambda_1 = \lambda_2 = \lambda$ be the arrival rate measured in packet per hybrid ALOHA slot and $p$ the transmission probability of both users. If we denote $q_{\{1\}} = q_{\{2\}} = a$, $q_{\{1, 2\}} = q_{\{2, 1\}} = b$ and $q_{\{1, 2\}} = c$, the bounds of the delay are given by the following proposition.

**Proposition 4:** Consider a symmetric two-user system. If the system is stable, i.e., $\lambda < p a + p^2 (b + c - a)$, the lower and upper bounds of the average delay for either user are given by
\[ D_{\text{Lower}} = \frac{1}{a} \left[ a(1 - \lambda) + p(b + c - a)(1 - \lambda / 2) \right], \]
(18)
and
\[ D_{\text{Upper}} = D_{\text{Lower}} + \frac{p^3 c [a - (b + c)]}{2 a r [p a + p^2 (b + c - a) - \lambda]} \].
(19)

**Proof:** Please refer to the appendix.

When the arrival rate $\lambda$ is low (e.g., $\lambda < 0.1$), the given upper bound is quite loose, as illustrated in Fig. 5. However, if the MPR capability of the system is strong enough, for small $\lambda$’s we can approximate the desired average delay with the lower bound. Intuitively, if the arrival rate $\lambda$ is small and the MPR capability is strong enough to handle all the packet deliveries, then the probability that the two queues are empty at the steady state tends to be 1, which, known from the appendix, is equivalent to $\lim_{t \rightarrow \infty} E(1|N_1^0 = 0, N_2^0 = 0) = 1$. Therefore, $\lim_{t \rightarrow \infty} E(1|N_1^0 > 0, N_2^0 > 0) = 0$, and the analysis in the appendix shows that the two bounds merge and become the exact average delay.

For moderate $\lambda$’s, if the MPR capability is relatively strong with $b+c$ getting close to $a$, or if the transmission probability $p$ is relatively small, the two bounds get very close to each other, as shown in Fig. 5(a) and Fig. 5(b). In these cases, the actual delay of the system can be roughly determined, e.g., by taking the average of the two bounds. Note that when the MPR model reduces to the capture model, $q_{\{1, 2\}} = c = 0$, the two delay bounds merge, and we obtain the “exact” delay which coincides with the result in [23].

VI. AN ILLUSTRATIVE EXAMPLE

In this section, a practical example is provided to illustrate the performance of hybrid ALOHA. Beyond the Simplistic
Assumption, the influences of noise, user interference and channel errors are all taken into consideration for signal detection.

A. System Set-up

Consider a two-user system. Both users communicate with the base station which employs a $P$-element linear antenna array. The signals of the two users arrive at the array with spatial angles $\theta = [\theta_1, \theta_2]$ with respect to the array normal. Suppose the array elements are uniformly spaced with a distance of one wavelength. The array response matrix $V(\theta)$ is then given by

$$V(\theta) = \begin{bmatrix} e^{j2\pi \cos \theta_1}, & \cdots, & e^{j2\pi (P-1) \cos \theta_1} \\ e^{j2\pi \cos \theta_2}, & \cdots, & e^{j2\pi (P-1) \cos \theta_2} \end{bmatrix}^T.$$  

The received signal $y$ at the base station is given by

$$y = V(\theta)Hs + n,$$  

where $H = \text{diag}[h_1, h_2]$ is the diagonal matrix containing the channel fading parameters for the two users. For simplicity, we assume flat fading channels, with $h_i$ ($i = 1, 2$) being a zero-mean complex Gaussian random variable of unit variance, i.e., $E[h_i] = 0$ and $E[|h_i|^2] = 1$. Let $s = [s_1, s_2]^T$, where $s_1 = 1$ or $-1$ denotes the symbol transmitted by user $i$ in a particular symbol interval (no intersymbol interference is considered in this example). $n$ is the additive white Gaussian noise with zero mean and variance $\sigma_n^2 I_P$, and the noise is assumed to be independent of the information sequence.

We assume that the channel fading is independent for two users. Furthermore, the fading is i.i.d from slot to slot yet remains unchanged during each slot. At the base station, the pilot-aided channel estimate $\hat{h}_i$ differs from the actual channel $h_i$ by an independent complex Gaussian random variable $\Delta h_i$, i.e.,

$$\hat{h}_i = h_i + \Delta h_i, \quad i = 1, 2,$$  

where $\Delta h_i$ are i.i.d from slot to slot with zero mean and variance $\sigma_i^2$. Thus $\hat{h}_i's$ are zero-mean complex Gaussian random variables with variance $\sigma_i^2 = 1 + \sigma_i^2$. The actual channel $h_i$ can be written in terms of $\hat{h}_i$ as [40]

$$h_i = \rho \hat{h}_i + \Delta h_i,$$  

where $\rho = 1/(1 + \sigma_i^2)$, $\hat{h}_i's$ are i.i.d complex Gaussian random variables with zero mean and variance $\sigma^2 = \sigma_i^2/(1 + \sigma_i^2)$. $E[h_i\hat{h}_i^*] = 0, \forall i, j = 1, 2$.

Considering coherent reception, signal processing at the receiver produces the output

$$z = \hat{H}^*Wy = \hat{H}^*WV(\theta)Hs + \hat{H}^*Wn,$$  

where $*$ denotes the complex conjugate operator, $\hat{H} = \text{diag}[\hat{h}_1, \hat{h}_2]$ contains the estimated channel coefficients for the two users, $W$ represents the beamforming weight matrix. The $i$th row of $W$, $w_i$, represents the weight vector for the $i$th user.

If we denote the $i$th column of $V(\theta)$ as $v_i$, (23) can be rewritten as

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} w_1v_1h_1\hat{h}_1^* + w_2v_2h_2\hat{h}_2^* \\ w_2v_1h_1\hat{h}_1^* + w_2v_2h_2\hat{h}_2^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} \hat{h}_1^*w_1n \\ \hat{h}_2^*w_2n \end{bmatrix}.$$  

Assuming that the data source is i.i.d., and without loss of generality, considering the case when $s_1 = 1$ (results for $s_1 = -1$ can be similarly obtained), from (22) and (24), it is easy to see

$$E[z_i|\hat{h}_i] = w_i v_i \rho |\hat{h}_i|^2,$$  

$$\text{Var}[z_i|\hat{h}_i] = |\hat{h}_i|^2 (|w_i v_i|^2 \sigma^2 + |w_i v_2|^2 + w_i v_1^H \sigma_n^2).$$  

Given the estimate of the channel, the signal to interference plus noise ratio (SINR) is then given by

$$E[\text{SINR}|\hat{h}_i] = \frac{\rho^2 |\hat{h}_i|^2}{\sigma^2 + |w_i v_i|^2 |\hat{h}_i|^2 + w_i v_i^H \sigma_n^2}.$$  

If the interference from user 2 is not present, then only the noise accounts, and we have

$$E[\text{SINR}|\hat{h}_i] = \frac{\rho^2 |\hat{h}_i|^2}{\sigma^2 + w_i v_i^H \sigma_n^2}.$$
B. Evaluation of Channel Estimation Error

Suppose both users transmit simultaneously in one of the pilot subslots. The received signal is given by
\[ y = v_1 h_1 s_1 + v_2 h_2 s_2 + n. \] (29)
Regardless of the existence of user 2, the “brute force” estimation of \( h_1 \) is given by
\[ \hat{h}_1 = \frac{v_1^H y}{|v_1^H v_1|^2} = h_1 + \frac{v_1^H v_2 h_2 s_2}{|v_1^H v_1|^2} + \frac{v_1^H n}{|v_1^H v_1|^2}. \] (30)
If \( L(\tau) \) pilot symbols are transmitted, where \( L(\tau) \) is proportional to \( \tau \), then by averaging, we have the mean-squared error of user 1’s channel estimation as
\[ \sigma^2_e(\tau) = \frac{1}{L(\tau)} \left( \frac{|v_1^H v_1|^2}{|v_1^H v_1|^2} + \sigma^2_n \right). \] (31)
In the absence of user 2, the channel estimation error is only introduced by the noise, hence
\[ \sigma^2_e(\tau) = \frac{1}{L(\tau)} \sigma^2_n. \] (32)
C. Probabilities of Successful Signal Receptions

With the derivations above, we can then adopt the SINR threshold model [41] for packet reception. Under this model, user 1’s packet is assumed to be successfully received if \( E[SINR(\hat{h}_1)] > \beta \), where \( \beta \) is the predetermined threshold required by system QoS.
When only user 1 transmits, we have
\[ q_{1||1} = Pr \left\{ \frac{\rho_1^2 |\hat{h}_1|^2}{\sigma^2_1 + \frac{w_1 w_1^H}{|w_1 v_1|^2} \sigma^2_n} > \beta \right\}, \] (33)
where \( \rho_1^2 = 1/(1 + \bar{\sigma}^2_1) \), \( \bar{\sigma}^2_1 = \bar{\sigma}^2_e (1 + \bar{\sigma}^2_2) \), and \( \bar{\sigma}^2_e \) is the variance of the collision-free channel estimation error given by (32).
Since \( |\hat{h}_1|^2 \sim exp(1/\sigma^2) \), the above probability is given by
\[
q_{1||1} = \frac{1}{\sigma^2} \int_{\beta(\sigma^2 + \frac{w_1 w_1^H}{|w_1 v_1|^2} \sigma^2_n)}^{\infty} e^{-\frac{x}{\sigma^2}} dx
= \exp \left\{ -\beta [\bar{\sigma}^2_1 + (1 + \bar{\sigma}^2_2)] \left( \frac{w_1 w_1^H}{|w_1 v_1|^2} \right) \right\}. \] (34)
Similarly, the probability of user 1’s packet being successfully received when two users stagger the training sequences in two pilot subslots, \( q_{1||1,2} \), is given by
\[
q_{1||1,2} = Pr \left\{ \frac{\rho_1^2 |\hat{h}_1|^2}{\sigma^2_1 + \frac{w_1 w_1^H}{|w_1 v_1|^2} + \frac{w_1 w_1^H}{|w_1 v_1|^2} \sigma^2_n} > \beta \right\}
= \exp \left\{ -\beta [\bar{\sigma}^2_1 + (1 + \bar{\sigma}^2_2)] \left( \frac{w_1 w_1^H}{|w_1 v_1|^2} + \frac{w_1 w_1^H}{|w_1 v_1|^2} \sigma^2_n \right) \right\}. \] (35)
Beyond the Simplistic Assumption, there is a probability that both users succeed even if their training sequences collide in one pilot subslot. Let \( q_{1||1,2} \) correspond to the probability of user 1’s packet being successfully received when two users’ training sequences collide. Eq. (35) applies to obtain the value of \( q_{1||1,2} \). Nevertheless, the collision-free channel estimation error \( \bar{\sigma}^2_e \) should be replaced by the estimation error with collision, denoted by \( \bar{\sigma}^2_e \), which is calculated using (31). Other probabilities can be obtained through similar calculations, and we can now obtain the throughput, stability region and the delay bounds of the proposed protocol.
In the sequel, we assume that at the base station, the linear antenna array has \( P = 5 \) elements, and the impinging signals are relatively close with \( \theta = [54^\circ, 64^\circ] \). The front-end processing exploits matched filter with the beamforming weight matrix \( W = V^H (\theta) \). Furthermore, we assume the packets are of small size containing 40 symbols per packet, with a training length of 2 symbols corresponding to \( \tau = 0.05 \).
D. Throughput of Hybrid ALOHA with Imperfect Channel Estimation

Under these settings, Fig. 6 presents the overall throughput of hybrid ALOHA with \( M = 2 \), as well as the throughput of traditional ALOHA based on both the collision model and the MPR model. Note that in the scheme of traditional ALOHA based on MPR model, simultaneous packet receptions are possible: successful signal decoding can be achieved as long as the SINR of the user’s signal is beyond the threshold required by QoS. As shown in (33)-(35), the probabilities of successful transmissions are related to \( \tau \) and are dependent on the noise power, interference level, and channel estimation accuracy.
Following Proposition 1, the normalized throughput of hybrid ALOHA can be evaluated through:
\[
\nu_{H\text{ALOHA}}(\tau) = \frac{R(\tau)e^{-R(\tau)}}{(1 + \tau)} \left\{ q_{1||1} + \left( \frac{q_{1||1,2}}{2} \right) \right\}. \] (36)
The throughput of traditional ALOHA based on the MPR model, \( \nu_{\text{ALOHA}}(\tau) \), has the same expression as (36), except that the corresponding probabilities would be different from that of hybrid ALOHA. The results presented in Fig. 6 show that hybrid ALOHA outperforms traditional ALOHA due to the improved MPR capability.

Fig. 6. Throughput comparison of different schemes under the specific system settings, \( SNR = 10 dB, \beta = 6 dB, \sigma^2_1 = 0.01, \sigma^2_2 = 0.06, p = 0.75 \).
As we have mentioned in Remark 1, the training length \( \tau \) should be designed according to specific system requirements. In this example, the optimal \( \tau \) maximizing the throughput of hybrid ALOHA \( \nu_{H\text{-ALOHA}}(\tau) \) can be obtained as:

\[
\tau_{\text{opt}} = \arg \max_{\tau \in \Omega} \{\nu_{H\text{-ALOHA}}(\tau)\},
\]

where \( \Omega = \{\tau \mid 0 < \tau < 1, L(\tau) \in \mathbb{N}\} \). \( L(\tau) \) is the number of training symbols defined in Section VI-B. The general closed-form expression of \( \tau_{\text{opt}} \) is difficult to find. However, the numerical result can be obtained through computer programming. In Fig. 7, the dashed line (red color) represents the optimal \( \tau \) against the arrival rates of each (symmetric) user under the system settings mentioned above.

In addition, an “equilibrium” curve (defined as the value of \( \tau \) such that \( \nu_{H\text{-ALOHA}}(\tau) = \nu_{\text{ALOHA}}(\tau) \)) is also presented in Fig. 7, below which the throughput of hybrid ALOHA is always higher than that of traditional ALOHA with the same training length \( \tau \). Beyond the equilibrium curve, the increment of \( \tau \) would degrade the spectral efficiency compared to that of the traditional ALOHA. Note that the staircase-like shapes in both curves reflect the constraint that the training sequence should contain integer number of symbols, i.e., \( L(\tau) \in \mathbb{N} \).

In Fig. 7, it is as expected that when arrival rates are very low, equilibrium values do not exist. This is due to the reason that rare collisions and insertion of idle sections make hybrid ALOHA inferior to traditional ALOHA.

**E. Stability Regions and Delay Bounds with Imperfect Channel Estimation**

Fig. 8 shows the stability regions for hybrid ALOHA under these settings. The arrival rates are in the unit of packets per unit time. For moderate arrival rates, hybrid ALOHA is shown to have the best performance due to the MPR improvement. On the other hand, in the cases when the arrival rate of one user is at the low end while that of the other user is at the high end, the stability region for hybrid ALOHA is contained in that of traditional ALOHA, and this degradation reflects the impact of idle sections. Refer to Fig. 4, it can be seen that within certain ranges, the stability region of hybrid ALOHA contains that of TDMA as well, even for the scheme without any guard time.

Fig. 9 illustrates the delays (bounds) of three different schemes as a function of the arrival rate for fixed transmission probability \( p = 0.75 \). Hybrid ALOHA results in the smallest delay among the three schemes for moderate rates, while the trend is reversed when the rates are small. The gains over traditional ALOHA with the MPR model seem relatively minor from the figure. One of the reasons is that the system contains only a small number of users \( (N = 2) \), which limits the possible improvement range. Much larger gap could be observed in systems with more users, which deserves further investigation.
VII. CONCLUSIONS

In this paper, taking a cross-layer (MAC and PHY) perspective, we proposed an efficient MAC protocol, named hybrid ALOHA, for wireless networks. In hybrid ALOHA, each slot has multiple pilot slots, from which each user can randomly selects one for training transmission. This design made it possible for collision-free channel estimation in multiple access environment. Quantitative analysis on the proposed protocol was carried out in terms of throughput, stability and delay. It was shown that significant performance improvement can be achieved in comparison with the traditional ALOHA protocol based either on the collision model or the MPR model.

APPENDIX

Proof of Proposition 4

Let $N_i^t$ denote the queue length of user $i$ ($i = 1, 2$) at time slot $t$. Let $\beta_i$ denote the number of packets that arrive at the $i$th user’s queue in time slot $t$. Let $F(x, y)$ be the moment generating function of the joint arrival process. Thus, for $|x| \leq 1$, $|y| \leq 1, t \in \mathbb{N}$

$$F(x, y) = E(x^{|G_1|} y^{|G_2|}) = (\lambda + \lambda\beta)(\lambda + \lambda\beta),$$  \hspace{1cm} (38)

where $\lambda = 1 - \lambda$. From the queue evolution equation (3), it can be seen that

$$E(x^{N_{i,1}} y^{N_{i,2}+1}) = F(x, y)\{E(1[N_1^0 = 0, N_2^0 = 0]) + (\frac{pa}{x} + 1 - pa)E(x^{N_1^1}1[N_1^0 > 0, N_2^0 = 0]) + (\frac{pa}{y} + 1 - pa)E(y^{N_2^1}1[N_1^0 = 0, N_2^0 > 0]) + \left[ (pa + p^2(b-a))\left(\frac{1}{x} + \frac{1}{y}\right) + \frac{p^2c}{xy} + 1 - 2(pa + p^2(b-a)) - p^2c \right] \times E(x^{N_1^1} y^{N_2^1}1[N_1^0 > 0, N_2^0 > 0]) \},$$  \hspace{1cm} (39)

where $1[\cdot]$ is the indicator function. According to Lemma 1 in [23], if $\lambda < pa + p^2(b + c - a)$, then the system is stable. Since $(N_1^t, N_2^t)$ is an irreducible, aperiodic Markov chain, stability is equivalent to existence of a unique stationary (limiting) distribution. Let $G(x, y)$ be the moment generating function of the joint stationary queue process, viz.,

$$G(x, y) = \lim_{t \to \infty} E(x^{N_1^t} y^{N_2^t}).$$  \hspace{1cm} (40)

Note that

$$G(0, 0) = \lim_{t \to \infty} E(1[N_1^t = 0, N_2^t = 0]),$$  \hspace{1cm} (41)

$$G(x, 0) - G(0, 0) = \lim_{t \to \infty} E(x^{N_1^t}1[N_1^0 > 0, N_2^0 = 0]),$$  \hspace{1cm} (42)

$$G(0, y) - G(0, 0) = \lim_{t \to \infty} E(y^{N_2^t}1[N_1^0 = 0, N_2^0 > 0]),$$  \hspace{1cm} (43)

$$G(x, y) + G(0, 0) - G(x, 0) - G(0, y) = \lim_{t \to \infty} E(x^{N_1^t} y^{N_2^t}1[N_1^0 > 0, N_2^0 > 0]).$$  \hspace{1cm} (44)

From (39), it follows that

$$G(x, y) = \left\{C(x, y)[G(x, y) + G(0, 0) - G(x, 0) - G(0, 0)] + B(x, y)[G(0, y) - G(0, 0)] + A(x, y)[G(x, y) - G(0, 0)] + G(0, 0) \right\} F(x, y),$$  \hspace{1cm} (45)

where

$$A(x, y) = \frac{pa}{x} + 1 - pa,$$

$$B(x, y) = \frac{pa}{y} + 1 - pa,$$

$$C(x, y) = [pa + p^2(b-a)] \left(\frac{1}{x} + \frac{1}{y}\right) + \frac{p^2c}{xy} + 1 - 2[pa + p^2(b-a)] - p^2c.$$  \hspace{1cm} (46)

Using the symmetry property of $G(x, y)$, we have $G(1, 1) = 1$ and $G(1, 0) = G(0, 1)$. In (45), let $y = 1$ and take derivative with respect to $x$ at both sides and then let $x = 1$, we find

$$(2\theta - pa)G(1, 0) - (\theta - pa)G(0, 0) = \theta - \lambda,$$  \hspace{1cm} (47)

where $\theta = pa + p^2(b + c - a)$. Furthermore, (45) can be rewritten as

$$G(x, y) = \{C(x, y)\left[G(0, 0) - G(x, 0) - G(0, y)\right] + B(x, y)\left[G(0, y) - G(0, 0)\right] + A(x, y)[G(x, 0) - G(0, 0)] + G(0, 0) \} \frac{F(x, y)}{1 - F(x, y)C(x, y)}.$$  \hspace{1cm} (48)

Let $G(x, y) \equiv dG(x, y)/dx$. In (48), letting $y = 1$, taking the derivative with respect to $x$ and then letting $x = 1$, we obtain

$$G(1, 1) = \frac{p^2(b + c - a)G(1, 0) + \lambda\theta}{\theta - \lambda}.$$  \hspace{1cm} (49)

Based on (48), a more tedious calculation will lead to the following result.

$$\frac{G(x, x)}{dx} \bigg|_{x=1} = 2\lambda + \frac{(2\theta - pa)G(1, 0)}{\theta - \lambda} + \frac{\lambda^2 + 2\lambda - 4\lambda\theta + p^2c}{2(\theta - \lambda)} - \frac{p^2c[G(1, 0) - \frac{1}{2}G(0, 0)]}{\theta - \lambda}.$$  \hspace{1cm} (50)

If we use the fact that

$$\frac{G(x, x)}{dx} \bigg|_{x=1} = G(1, 1) + G(2, 1) = 2G(1, 1),$$  \hspace{1cm} (51)

and associate with the result from (49), we can solve for $G(1, 0)$ and $G(1, 1)$ to obtain

$$G(1, 1) = \frac{2\lambda\theta - \lambda^2pa - \lambda^2\theta - p^2c[b+c-a][G(1, 1) + G(0, 0) - G(1, 0) - G(0, 1)]}{2a[\theta - \lambda]}.$$  \hspace{1cm} (52)

Since $G(1, 1)$ is equal to the mean queue length of the users, applying Little’s Theorem [42], we have the average delay $D$ as

$$D = \frac{G(1, 1)}{\lambda} = \frac{1}{a}\left[\frac{a(1 - \lambda) + p(b + c - a)(1 - \lambda/2)}{pa + p^2(b + c - a) - \lambda}\right] + \phi,$$  \hspace{1cm} (53)

where

$$\phi = \frac{-p^2c[b+c-a][G(1, 1) + G(0, 0) - G(1, 0) - G(0, 1)]}{2ar(pa + p^2(b + c - a) - \lambda)}.$$  \hspace{1cm} (54)

Since $G(1, 1) + G(0, 0) - G(1, 0) - G(0, 1) = \lim_{t \to \infty} E(1[N_1^t > 0, N_2^t > 0])$ as (44) implies, associated
with \( b + c - a < 0 \) and \( \lambda < pa + p^2(b + c - a) \), we then conclude that

\[
0 \leq \phi \leq \frac{-p c(b + c - a)}{2 a r [ p a + p^2(b + c - a) - \lambda ]}.
\]

Therefore, the average delay is lower bounded by

\[
D_{\text{Lower}} = \frac{1}{a} \left[ \frac{\lambda(1 - \lambda) + p(b + c - a)(1 - \lambda/2)}{pa + p^2(b + c - a) - \lambda} \right],
\]

and upper bounded by

\[
D_{\text{Upper}} = \frac{1}{a} \left[ \frac{\lambda(1 - \lambda) + p(b + c - a)(1 - \lambda/2)}{pa + p^2(b + c - a) - \lambda} \right] - \frac{p c(b + c - a)}{2 a r [ p a + p^2(b + c - a) - \lambda]}. \]

From the analysis above, we see that the bounds become the exact delay of the system when

\[
\begin{align*}
&\text{for } c = 0 \text{ or } \lim_{t \to \infty} E(1|N^3_1 > 0, N^3_2 > 0) = 0,
\end{align*}
\]

Acknowledgement

The authors would like to thank Prof. Lang Tong from Cornell University for insightful suggestions and helpful discussions. Our thanks also go to the anonymous reviewers for their valuable comments.

References