IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 17, 2020, accepted April 11, 2020, date of publication April 20, 2020, date of current version May 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2988671

Non-Stationary Bandit Strategy for Rate Adaptation With Delayed Feedback

YAPENG ZHAO[1,2], HUA QIAN[2], (Senior Member, IEEE), KAI KANG[2], (Member, IEEE), AND YANLIANG JIN[1]

[1]School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China
[2]Shanghai Advanced Research Institute, China Academy of Sciences, Shanghai 201210, China

Corresponding author: Kai Kang (kangk@sari.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61671436, and in part by the Science and Technology Commission Foundation of Shanghai under Grant 18511103502.

ABSTRACT Rate adaptation is an efficient mechanism to utilize the channel capacity by adjusting the modulation and coding scheme in a dynamic wireless environment. The channel feedback, such as acknowledgment/negative acknowledgment (ACK/NACK) messages or the channel measurement such as received signal strength indicator (RSSI) can be applied to the rate adaptation. Existing rate adaptation algorithms are mainly driven by heuristics. They can not achieve satisfactory transmission rates in the time-varying environment. In this paper, we focus on the rate adaptation problem in a time-division duplex (TDD) system. A multi-armed bandit (MAB) strategy is applied to learn the changes of the channel condition from both RSSI and ACK/NACK signals. A discounted upper confidence bound based rate adaptation (DUCB-RA) algorithm is proposed. We show that the performance of the proposed algorithm is converged to the optimal with mathematical proofs. Simulation results demonstrate that the proposed algorithm can adapt to the time-varying channel and achieve better transmission throughput compared to existing rate adaptation algorithms.

INDEX TERMS Throughput, rate adaptation, wireless communication, time-division duplex, multi-armed bandit.

I. INTRODUCTION

Comparing to wired communication systems, wireless communication systems suffer from time-varying channels caused by channel fading or interference [1]–[4]. These stochastic effects become more severe when the environment changes, for example, the movements of mobile stations [5], [6]. Rate adaptation (RA) is necessary to meet the channel changes by adjusting the modulation and coding schemes of the transmitter.

In order to determine appropriate transmission rate, the channel condition needs to be evaluated. The channel state information (CSI) that contains all information about the channel properties is the most accurate measure of the channel. However, complete CSI feedback is costly and usually infrequent [7]. Other performance metrics such as signal-to-noise ratio (SNR), received signal strength indicator (RSSI), and acknowledgement/negative acknowledgement (ACK/NACK) can be used to select appropriate transmission rates [8], [9]. Based on the selection of channel condition evaluation metrics, the RA schemes can be classified as frame-level and measurement-based schemes [10].

Frame-level RA schemes determine the transmission rate of current packet from the knowledge of previous transmissions. Such knowledge is available in the form of the ACK/NACK signals. Frame-level RA schemes usually can not respond to channel variations that occur on short timescales. For comparison, measurement-based schemes can respond to fast channel variation. These schemes determine the transmission rate based on the channel measurements. The channel measurements include SNR, RSSI, etc. The mapping between the rate and the channel measurement can change when the channel condition changes [11].

Auto rate fallback (ARF) adjusted the transmission rate intuitively: it decreased the transmission rate by one gear when missing 2 ACKs consecutively and increased the transmission rate by one gear when receiving 10 ACKs consecutively [8]. Adaptive ARF (AARF) increased the time interval between attempts at a higher rate when encountering successive probe failures [12]. ARF and AARF classified channel

The associate editor coordinating the review of this manuscript and approving it for publication was Lei Guo.

conditions as either "good" or "bad" based on received ACK signals, and adjusted the rate accordingly. This binary classification was not efficient in converging to the optimal rate, especially when facing a large collection of available rates or a rapidly varying channel. Minstrel utilized a mechanism called multi-rate retry chain to update the rate adaptation strategy [13]. The retry chain consisted of four different rates and the corresponding number of attempts. Minstrel allocated 90% transmission for the normal transmission, and the rest 10% transmission for probing other rates. SampleRate attempted the available rate in the order from the highest one to the lowest one [14]. SampleRate allocated a fixed percentage of packets to probe other rates. Due to the fixed exploration ratio, SampleRate and Minstrel algorithms could not adapt to the fast time-varying channel quickly, and their performance degraded in the static channel.

The SNR-guided rate adaptation (SGRA) algorithm set up the relationship between SNR and frame delivery ratio (FDR), and utilized forced probes to calibrate such relationship in a real-world channel [9]. SGRA ignored the fact that the mapping from SNR to the transmission rate was not deterministic in practical channel conditions. In [15] and [16], the authors acknowledged that RSSI or SNR alone did not accurately capture the changes in wireless channels.

Besides, The ACK/NACK signals are always delayed feedbacks of previous transmissions [17], while the impact of the delay is ignored in existing work. The delayed feedback may cause incorrect rate selection thus degrade the overall throughput.

To address the above issues, reinforcement learning algorithms can be appropriate tools. All RA strategies have to trade-off between exploitation and exploration in the dynamic wireless environment. It is straightforward to map the exploitation and exploration phases of the multi-armed bandit (MAB) algorithm into the RA framework and it is easy to control the switches between the exploitation and exploration phases [18]. In addition, in the RA problem, the channel state transition does not depend on the rate selection (action). The action reward does not depend on the previous channel state either. Thus, MAB is sufficient and effective to model the RA problem.

In this paper, we model the RA problem as a MAB problem. A discounted upper confidence bound based rate adaptation (DUCB-RA) algorithm is proposed. Our contributions can be summarized as follows: First, both of RSSI and ACK/NACK signals are adopted to determine appropriate transmission rates, while major existing work only deals with one of them. Secondly, we treat the ACK/NACK signals as delayed responses to the quality of the transmission, which is typical in most wireless systems such as the long term evolution (LTE) system [17]. Thus, our model is more accurate than other works. Thirdly, we model the RA adaptation as a MAB problem. We show theoretically that the proposed algorithm is asymptotically optimal.

The rest of the paper is organized as follows. Section II introduces the system model. Section III presents the proposed DUCB-RA algorithm, and Section IV gives a theoretical analysis to the proposed algorithm's performance. Section V discusses the RA performance under various simulation scenarios. Finally Section VI concludes the paper.

II. SYSTEM MODEL

For a wireless communication system, the capacity of an additive white Gaussian noise (AWGN) channel depends on the channel bandwidth and SNR. Thus the mapping from SNR to the optimal transmission rate is fixed. When taking the channel fading into consideration, such relationship is no longer fixed. The relation between the optimal rate and SNR can be highly dynamic when the channel condition changes. In addition, it is not trivial to obtain a reliable estimate of the SNR of a link. Many radio interfaces only provide RSSI as an uncalibrated SNR estimate.

In this work, we consider a time-division duplex (TDD) based system. Fig. 1 depicts the simplified transmission between the transmitter and the receiver. At time slot 1, the transmitter sends packet to the receiver. The receiver reports ACK/NACK signal to the transmitter. The transmit process and receive process occur alternatively. In this system, we assume that the channel condition is "slow-varying". In other words, the channel condition does not change within a packet transmission, while it can change from packet to packet.

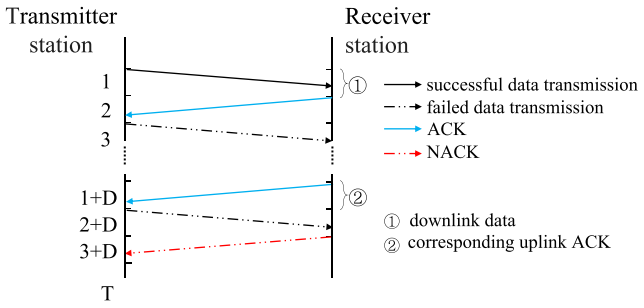


FIGURE 1. Packet transmission in TDD system.

For each packet, the transmitter needs to select a rate from the set of available rates $\{R_k\}_{k=1}^{K}$. In this set, we assign a larger index to a larger rate, i.e., R_1 is the lowest rate, and R_K is the largest one.

In the RA problem, performance metrics that measures the channel conditions can be considered. The RSSI can be measured when a packet is received. The RSSI of the received signal can be a measure of the SNR of the transmit channel considering the reciprocity of the TDD system. The ACK/NACK signal, usually provided as the control information, can be obtained by the transmitter as an indicator.

The ACK/NACK signals, in general, are delayed response to the quality of the previous transmissions. For example, in time division LTE (TD-LTE) system, the ACK/NACK responds to the transmitted signal 3 time slots ago [17]. Without loss of generality, we assume the ACK/NACK signals are delayed for D time slots, where $D \geq 3$, and D is odd due to

the nature of the TDD system. If the ACK signal is received, the transmission is successful. If the NACK signal is received or the ACK/NACK signal is lost, the transmission fails. Let $\theta_{t,k}$ denote the probability of a successful transmission in the time slot t with rate R_k.

Next, we consider the channel conditions. According the variation of the channels, we classify the channels into two categories: stationary channel environment, non-stationary channel environment.

A. STATIONARY CHANNEL ENVIRONMENT

In stationary channel environment, the channel is considered to be static. A stationary channel implies that $\theta_{t,k}$ does not evolve over time. In stationary channel environment, a larger rate may incur more errors, thus may have lower successful transmission probability than a smaller rate, i.e., $\theta_{t,k} \leq \theta_{t,k-1}$.

B. NON-STATIONARY CHANNEL ENVIRONMENT

In practice, channel conditions are always non-stationary, i.e., $\theta_{t,k}$ may evolve over time. Due to the movement of users or objects, the wireless communication system suffers from fading effects. The RSSI varies over time.

In addition, if we consider the changes of the environment, the mapping between the RSSI and the optimal transmission rate may also changes.

We assume that the RSSI can be divided into discrete L levels $\{S_l\}_{l=1}^{L}$ based on the fading status, and the channel states is also discretized into M states $\{C_m\}_{m=1}^{M}$ based on the noise level to simplify the discussion. Furthermore, we assume that the channel state becomes worse from C_M to C_1. In a worse channel, higher RSSI is required to achieve the same rate. The channel state transitions can be modeled as a hidden Markov model (HMM) [19]. The transition probability from channel state i to state j is defined as

$$P_{i\to j} = Pr(C_t = j | C_{t-1} = i). \tag{1}$$

In this work, we assume the channel changes slowly, meaning the channel state transition only occurs between adjacent channel states. Clearly, when the channel state changes, the probability of successful transmission $\theta_{t,k}$ may also change.

In order to determine the appropriate transmission rate, RA algorithms can be applied. For example, The LA algorithm compares the average RSSI RSS_{avg} with the threshold Th_k for rate R_k to select the appropriate rate [20]. The average RSSI RSS_{avg} is defined as

$$RSS_{avg} = (1 - a_1)RSS'_{avg} + a_1 RSS_i, \tag{2}$$

where RSS'_{avg} is the average RSSI of previous time slot, RSS_i is the RSSI value observed in time slot i, and $a_1 \in [0, 1]$ is a time decaying factor. When the transmission fails, the LA algorithm updates the threshold Th_k associated with the rate as

$$Th_k = (1 - a_2)Th_k + a_2 RSS_i, \tag{3}$$

where $a_2 \in [0, 1]$ is the weight of RSS_i value observed in time slot i.

RSSI is not a perfect indicator for the channel condition. When noise or interference changes, the RSSI may stay unchanged. In this case, the LA algorithm may be stuck at a low rate.

The enhanced history-aware robust rate adaptation algorithm (HA-RRAA) proposed in [21] selects a new rate for the packet transmissions in the next adaptive time window T_R. HA-RRAA observes the ACK/NACK information in the time window, and calculates the packet loss ratio of these frames. It decreases the rate to the next lower one if the loss ratio is greater than a threshold, or increases to the next higher one if the loss ratio is smaller than another threshold.

RSSI and ACK/NACK provide information about the channel condition from different perspectives. RSSI is an estimate of SNR, which gives a direct suggestion to the transmission rate. ACK/NACK signals record authentic channel response of each packet transmission. Conventional RA algorithms generally performed adaptation based on one of these signals. In our proposed algorithm, we take both RSSI and ACK/NACK signals as inputs. The proposed algorithm is discussed in detail in the next section.

III. DISCOUNTED UPPER CONFIDENCE BOUND BASED RATE ADAPTATION ALGORITHM

In this section, we present a discounted upper confidence bound based rate adaptation (DUCB-RA) algorithm that adopts both RSSI and ACK/NACK information. The rate selection problem can be modeled as a MAB problem. Under each RSSI level S_l, there are K possible transmission rates R_k. The K transmission rates can be considered as arms in the MAB problem. A total of L MAB problems can be formulated. When a transmission of rate R_k at RSSI level S_l occurs, the specific arm of the MAB problem is pulled. A reward that evaluates the performance of the current selection can be calculated when the corresponding ACK/NACK signal is received. Next transmission rate can be determined based on the RSSI measurement and the updated reward estimations. The overall algorithm is detailed as follows.

To begin with, for RSSI level S_l, an initial guess of the transmission rate can be obtained. For example, such guess can be obtained during the random access phase. Then, each rate R_k is assigned with initial performance estimate $r_{0,k}$, which is the instantaneous reward obtained of each transmission. If the transmission rate is smaller than the initial guess, the instantaneous reward is assigned as $r_{0,k} = R_k/R_K$. If the transmission rate is larger than the initial guess, $r_{0,k} = 0$. The accumulated estimated reward r_k is initialized by the instantaneous reward, or $r_k = r_{0,k}$.

In the subsequent transmissions, the estimated reward r_k is updated based on the ACK/NACK feedbacks. The instantaneous reward of selected rate R_k is given by $r_{t,k} = ACK_t R_k/R_K$, where $ACK_t = 1$ when an ACK signal is received at time slot t, or $ACK_t = 0$ otherwise. As we discussed earlier, the ACK/NACK feedback is a delayed version.

The ACK/NACK received in time slot t gives a quantified evaluation of the transmission with rate R_k occurred in time slot $(t - D)$. Since R_K is the largest rate in the available set. It is clear that the instantaneous reward is bounded with $r_{t,k} \in [0, 1]$. Based on the recorded historical reward and the instantaneous reward $r_{t,k}$, the estimated reward r_k at time slot t can be updated as

$$r_k = \frac{1}{N_{t,k}} \sum_{i=1+D}^{t} \gamma^{t-i} r_{i,k} \delta_{i-D,k}, \tag{4}$$

where γ is a forgetting factor, $\delta_{t,k}$ is an indicator parameter. In (4), $N_{t,k}$ is the discounted number chosen times of R_k, which is denoted as

$$N_{t,k} = \sum_{i=0}^{t} \gamma^{t-i} \delta_{i,k}. \tag{5}$$

When the channel is non-stationary, the parameter γ in (4) is used to limit the influence of outdated observation. The parameter $\delta_{t,k}$ indicates which rate is selected at time slot t, which is defined as

$$\delta_{t,k} = \begin{cases} 1, & \text{if } R_k \text{ is selected at time slot } t, \\ 0, & \text{if other rate is selected at time slot } t. \end{cases} \tag{6}$$

Since the initial guess of the rate is performed for the system, the parameter $\delta_{t,k}$ can initialized by setting $\delta_{0,k} = 1$.

The estimated reward r_k is updated after the ACK/NACK signal is fed back. If there is no ACK/NACK signals received, r_k can still be updated by assuming a NACK signal is received. In this case, however the current estimate of RSSI can not be obtained. The transmitter maintains the previous estimation of RSSI to perform the rate selection.

For the next transmission, the rate associated with the maximum estimated reward r_k is selected for transmission. The rate selection mechanism is given by

$$k^* = \arg\max_{k \in K} r_k. \tag{7}$$

When the channel condition changes, (7) may not be able to track the optimal rate. For example, when the channel condition becomes better, the above RA algorithms will continue to select the previous rate as it yields the best estimated reward. The new best rate is not chosen due to lack of exploration. A bias term $c_{t,k}$ can be introduced to increase the probability of exploring other rates. With exploration, more rates can be probed with updated performance estimates. The decision of rate adaptation is more appropriate, especially in a time-varying channel. Based on the one-sided confidence interval derived from the Chernoff-Hoeffding bound [22], [23], $c_{t,k}$ is set as

$$c_{t,k} = 2\sqrt{\xi \log(n_t / N_{t,k})}, \tag{8}$$

where ξ is an adjustable parameter to control the exploration. In (8), n_t represents the total discounted number of chosen times of all the R_k, which is denoted as

$$n_t = \sum_{k=1}^{K} N_{t,k}. \tag{9}$$

Thus the new rate selection mechanism is given by

$$k^* = \arg\max_{k \in K} r_k + c_{t,k}. \tag{10}$$

With the bias term $c_{t,k}$ expressed in (8), the new rate selection mechanism suggests that if a rate is less selected, a larger bias is applied to the estimated reward. The probability of been selected is increased. The parameter ξ in (8) can be applied to control the exploration ratio. In a fast-varying channel, more exploration is needed. A larger ξ can be applied. Otherwise, a smaller ξ can be applied.

The proposed DUCB-RA algorithm is summarized in Algorithm 1. The transmitter selects the rate according to the initial channel interaction at the beginning, then updates the estimated r_k based on the ACK/NACK feedbacks. The rate of the next transmission can be determined with (10). With the proposed DUCB-RA algorithm, the transmitter can track the time-varying channel through both RSSI and ACK/NACK information and provide appropriate rate for transmission.

Algorithm 1 DUCB-RA Algorithm

Input: $\{R_k\}_{k=1}^{K}$, current RSSI S_l, estimated reward r_k.
for $t = 2, 4, \cdots, T$
  Obtain the ACK/NACK signal.
  Update the estimated reward r_k associated with S_l using (4).
  if no ACK/NACK signal **then**
    Keep the previous RSSI as S_l for the next rate selection.
  else
    Measure current RSSI as S_l for the next rate selection.
  end if
  Select rate for the next time slot $(t + 1)$ using (10) according to S_l.
end for

IV. PERFORMANCE ANALYSIS

In this section, we study the performance of the proposed DUCB-RA algorithm. We show theocratically that the regret performance achieves a sub-linear order. In other words, the proposed algorithm is asymptotically optimal.

Ideally, if a rate adaptation mechanism has perfect knowledge of the current channel condition, an appropriate rate that maximizes the throughput can be selected. Regret is introduced to measure the performance loss between the proposed algorithm and the omniscient RA mechanism with perfect knowledge of the channel condition. The accumulative regret R_T is introduced to measure the performance loss of the proposed algorithm compared to the omniscient RA mechanism. The R_T is the gap between the accumulative reward obtained from the proposed algorithm and that of the omniscient RA mechanism. We show next that with properly selected parameters, the regret performance of proposed DUCB-RA algorithm achieves a sub-linear order. The sub-linear growth

of the accumulative regret demonstrates that the proposed algorithm is asymptotically optimal [22].

We first analysis the performance of the algorithm assuming that RSSI is constant and ACK/NACK signals are fed back timely.

Let $\mathcal{T}_s$ represents the time slots that transmission occurs and $\mathcal{T}_r$ represents the time slots that the reception occurs. We have $\mathcal{T}_s = \{2k-1; 2k-1 \leq T, k \in \mathbb{N}\}$, $\mathcal{T}_r = \{2k; 2k \leq T, k \in \mathbb{N}\}$. In the subsequent notations in this section, $t \in \mathcal{T}_s$ if there is no extra explanation.

The expectation of R_T is given by

$$\mathbf{E}[R_T] = \mathbf{E}\left[\sum_{t=1}^{T}\left(r_{t,k_t^*} - r_{t,k}\right)\delta_{t,k}\mathbf{1}_{\{\mu_{t,k}<\mu_{t,k_t^*}\}}\right], \tag{11}$$

where $\mathbf{E}[\cdot]$ denotes the expectation, $\mathbf{1}_{\{\cdot\}}$ is an indicator function, which is 1 if the statement in the parentheses is true, or 0 otherwise. In (11), $\mu_{t,k} = \theta_{t,k}R_k/R_K$ is the expected reward of choosing R_k. Let k_t^* denote the optimal rate in time slot t. If $\mu_{t,k} < \mu_{t,k_t^*}$, R_k is not the optimal choice. The term $\delta_{t,k}\mathbf{1}_{\{\mu_{t,k}<\mu_{t,k_t^*}\}}$ in (11) represents the case that the sub-optimal rate is selected in time slot t. The term $\left(r_{t,k_t^*} - r_{t,k}\right)$ represents the reward loss due to the selection of the sub-optimal rate.

Let $M_T(k)$ denote the number of times of choosing the sub-optimal rate R_k in the first T time slots, which is expressed as

$$M_T(k) = \sum_{t=1}^{T}\delta_{t,k}\mathbf{1}_{\{k\neq k_t^*\}}. \tag{12}$$

Since $r_{t,k} \in [0, 1]$ and $r_{t,k_t^*} > r_{t,k}$, we have

$$0 < r_{t,k_t^*} - r_{t,k} < 1, \tag{13}$$

substituting (12) and (13) into (11), we have

$$\begin{aligned}\mathbf{E}[R_T] &= \mathbf{E}\left[\sum_{t=1}^{T}\left(r_{t,k_t^*} - r_{t,k}\right)\delta_{t,k}\mathbf{1}_{\{\mu_{t,k}<\mu_{t,k_t^*}\}}\right]\\ &\leq \mathbf{E}\left[\sum_{t=1}^{T}\delta_{t,k}\mathbf{1}_{\{\mu_{t,k}<\mu_{t,k_t^*}\}}\right]\\ &= \mathbf{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{K}\delta_{t,k}\mathbf{1}_{\{k\neq k_t^*\}}\right]\\ &= \sum_{k=1}^{K}\mathbf{E}[M_T(k)].\end{aligned} \tag{14}$$

According to (14), if we find the upper bound of the expectation of the number of times that suboptimal rates are selected, we can establish the upper bound of the expected regret.

In (12), $M_T(k)$ can be further divided into two parts by a control parameter Λ for the subsequent proof.

$$\begin{aligned}M_T(k) = &\underbrace{\sum_{t=1}^{T}\delta_{t,k}\mathbf{1}_{\{k\neq k_t^*, N_{t,k}<\Lambda\}}}_{①}\\ &+\underbrace{\sum_{t=1}^{T}\delta_{t,k}\mathbf{1}_{\{k\neq k_t^*, N_{t,k}\geq\Lambda\}}}_{②}.\end{aligned} \tag{15}$$

According to the Lemma 1 in [23], the first part ① in (15) is upper-bounded by

$$\sum_{t=1}^{T}\delta_{t,k}\mathbf{1}_{\{k\neq k_t^*, N_{t,k}<\Lambda\}} \leq \lceil T(1-\gamma)/2\rceil\Lambda\gamma^{-\frac{1}{1-\gamma}}, \tag{16}$$

where the notation $\lceil\cdot\rceil$ means the round up process.

If the channel state changes, the estimated reward for R_k can be poor for $D(\gamma)$ rounds. As shown in [23], $D(\gamma)$ is given by

$$D(\gamma) = \log((1-\gamma)\xi\log n_K)/\log(\gamma), \tag{17}$$

where n_K is calculated according to (9). The natural logarithm base e is omitted in this paper.

Let T' denotes the rest rounds except $D(\gamma)$ rounds, the second part ② in (15) is bounded by

$$\begin{aligned}\sum_{t=1}^{T}\delta_{t,k}\mathbf{1}_{\{k\neq k_t^*, N_{t,k}\geq\Lambda\}} \leq\ &\Upsilon D(\gamma)\\ &+\sum_{t\in T'}\underbrace{\delta_{t,k}\mathbf{1}_{\{k\neq k_t^*, N_{t,k}\geq\Lambda\}}}_{③},\end{aligned} \tag{18}$$

where Υ denotes the number of times of channel changes within T.

We further divide the term ③ in (18) into two parts: $\delta_{t,k}\mathbf{1}_{\{k\neq k_t^*\}}$ and $N_{t,k} \geq \Lambda$. The term ③ is satisfied as the intersection of the above two terms. The first part holds in the following cases. First, $\mu_{t,k}$ and μ_{t,k_t^*} are close to each other, the bias term $c_{t,k}$ can not discriminate them. Thus (10) may choose the sub-optimal rate. Second, the estimated reward $r_{k_t^*}$ of the optimal rate is under-estimated. Third, the estimated reward r_k of the sub-optimal rate is over-estimated. These cases can be summarized as

$$\begin{cases}\mu_{t,k_t^*} - c_{t,k} \leq \mu_{t,k} + c_{t,k}, & \text{(19a)}\\ r_{k_t^*} \leq \mu_{t,k_t^*} - c_{t,k_t^*}, & \text{(19b)}\\ r_k \geq \mu_{t,k} + c_{t,k}. & \text{(19c)}\end{cases}$$

The union of the above three terms is the necessary and sufficient conditions for $\delta_{t,k}\mathbf{1}_{\{k\neq k_t^*\}}$. Combing (19) with $N_{t,k} \geq \Lambda$, we can get the requirement of the term ③. Next, we analysis (19a) to (19c) in detail.

Let $\Delta\mu$ represent the minimum gap between $\mu_{t,k}$ and μ_{t,k_t^*}. If $c_{t,k} \leq \frac{\Delta\mu}{2}$, (19a) never occurs. Substituting $c_{t,k} \leq \frac{\Delta\mu}{2}$, we obtain the value of the control parameter $\Lambda = 16\xi\log n_t/\Delta\mu^2$.

For terms (19b) and (19c), the proof in [23] shows that the probabilities of (19b) and (19c) are equal and are bounded by

$$P_r \leq (1-\gamma)^{-1} - K + \left\lceil \frac{\log \frac{1}{1-\gamma}}{\log(1+\eta)} \right\rceil \frac{T(1-\gamma)}{1-\gamma^{1/(1-\gamma)}}. \quad (20)$$

Combining (16), (18), (19) and (20), we conclude

$$\mathbf{E}[M_T(k)] \leq \mathrm{C}_1 T(1-\gamma)\log\frac{1}{1-\gamma} + \mathrm{C}_2\frac{\Upsilon}{1-\gamma}\log\frac{1}{1-\gamma} - 1, \quad (21)$$

where

$$\mathrm{C}_1 = \frac{16\sqrt{2}\xi}{\gamma^{1/(1-\gamma)}(\Delta\mu)^2} + \frac{4}{\left(1-\frac{1}{\mathrm{e}}\right)\log(1+4\sqrt{1-1/2\xi})}, \quad (22)$$

$$\mathrm{C}_2 = \frac{\gamma-1}{\log(1-\gamma)\log\gamma}\log\left((1-\gamma)\xi\log n_K\right). \quad (23)$$

The bound (21) is associate with T and Υ. Setting $\gamma = 1-(\sqrt{2\Upsilon/T})/4$ [22], we have

$$\lim_{T\to\infty}\frac{\mathbf{E}[M_T(k)]}{\sqrt{T\Upsilon}\log T} = 0. \quad (24)$$

The growth rate $\sqrt{T}\log T$ is lower than the linear growth rate T. We know that the growth rate of $\mathbf{E}[M_T(k)]$ achieves the sub-linear order. From (14), we conclude that the growth rate of $\mathbf{E}[R_T]$ also achieves the sub-linear order.

The above proofs are conducted under the assumption that the RSSI remains the same. When the RSSI changes, RSSI may vary when channel state remains the same. The worst case is that every RSSI level is excited in all Υ channel state changes. Thus the expected regret is bounded by

$$\mathbf{E}[R_T] \leq L\sum_{k=1}^{K}\mathbf{E}[M_T(k)]. \quad (25)$$

which has the same sub-linear growth characteristic as the fixed RSSI case.

Next, we consider the performance loss due to the delayed feedback. We first consider a stationary MAB problem which the reward is fixed delayed by one time slot. On other words, there is another selection between the action selection and the reward reception. The intercalary rate selection does not receive any guidance as a result of the delayed reward [24]. In a stationary MAB problem, the only performance loss comes from the possible suboptimal action selection caused by the first delayed feedback. ACK/NACK signals are delayed D time slots in the system model we formulate in Section II. The delayed feedbacks cause $(D-1)/2$ times rate selection being unguided. The expectation of the accumulative regret $\mathbf{E}[R'_T]$ in the delayed model is bounded by

$$\mathbf{E}[R'_T] \leq \mathbf{E}[R_T] + \tfrac{D-1}{2}. \quad (26)$$

Next, we consider the changes in the RSSI. Since there are L MAB problems, i,e., RSSI is varying over time, the additional loss is multiplied by L at most. We have

$$\mathbf{E}[R'_T] \leq L(\mathbf{E}[R_T] + \frac{D-1}{2}). \quad (27)$$

We further consider the performance loss caused by channel state changes, The previous obtained knowledge is outdated if the channel state is changed. The RA mechanism needs learn the new condition. Thus extra loss caused by delayed feedback need to be included whenever the channel state is changed. We have

$$\begin{aligned}\mathbf{E}[R'_T] &\leq L(\mathbf{E}[R_T] + \frac{D-1}{2} + \Upsilon) \\ &= L(\sum_{k=1}^{K}\mathbf{E}[M_T(k)] + \frac{D-1}{2} + \Upsilon). \end{aligned} \quad (28)$$

From (28), we conclude that $\mathbf{E}[R'_T]$ also achieves the sub-linear order, thus the proposed algorithm is asymptotically optimal.

The theoretical analysis sets the upper bound of the performance loss of proposed DUCB-RA algorithm in the time-varying channel. In section V, we provide simulation results to demonstrate the performance of the proposed algorithm.

V. SIMULATION RESULTS

In this section, we present numerical results of the proposed DUCB-RA algorithm. A system with 15 different transmission rates $\{R_k\}_{k=1}^{15}$ is considered. Channel states are divided into 10 states $\{C_m\}_{m=1}^{10}$ depending on the interference and noise level. The path-loss model and Rician fading are introduced to simulate the time-varying RSSI. The RSSI is quantized into 35 levels $\{S_l\}_{l=1}^{35}$.

In our experiment, the transmitter sends packets to the receiver from time slot 1. For each packet transmission, the transmitter selects the rate from the set $\{R_k\}_{k=1}^{15}$ based on the the RSSI evaluation and estimated reward.

We would like to examine the performance of DUCB-RA compared with the other RA algorithms: ARF, LA, Minstrel, and HA-RRAA in both stationary radio environments and non-stationary radio environments. In the simulation, the multi-rate retry chain in Minstrel is updated when every 10 data packets is sent to match the simulation condition of the DUCB-RA algorithm. The initial threshold of LA is set according to the best estimate of the initial channel condition.

We first consider a static channel condition. Both the channel state and the RSSI are constant. The simulations are conducted with different states and RSSIs. Since the channel is static, we set the forgetting factor γ to 1, meaning all previous samples are counted, and ξ to 0.1 since there is no need to explore much.

The performances of above algorithms under different channel settings are recorded in Table 1. In this table, the throughputs are normalized by the optimal throughput under selected channel conditions.

TABLE 1. Normalized throughput in different static channel conditions.

Channel condition		Normalized throughput				
		ARF	LA	Minstrel	HA-RRAA	DUCB-RA
Channel 1	$\frac{C_5}{S_{15}}$	0.89	0.99	0.93	0.98	0.98
Channel 2	$\frac{C_5}{S_{20}}$	0.89	0.99	0.94	0.98	0.98
Channel 3	$\frac{C_{10}}{S_{35}}$	0.98	0.99	0.96	1	0.99

From Table 1, we observe that all algorithms achieve satisfactory performance in all cases. Due to the fixed exploration ratio, ARF and Minstrel algorithms probe other rates frequently in the stationary channel condition. Their performance degrades in such case. Comparing the performance in channel 3 to other channel conditions, we observe that the performance of all RA algorithms becomes better. In this case, the channel condition is the best and the optimal rate selection is the highest rate. There are no higher rates for exploration. The potential performance loss caused by the exploration is reduced.

Let us look into the rate selections in different time slot in one of the simulations. Fig. 2 shows the actual rate that the transmitter chooses with ARF and DUCB-RA algorithm in a fixed channel condition (channel 2 in Table 1). The optimal rate is selected by the omniscient strategy with perfect knowledge of the channel condition. It can be observed that the proposed DUCB-RA algorithm always pick the optimal rate except several probes. On the other hand, ARF, which is driven by heuristics, probes the higher rate frequently in the stationary channel condition, resulting considerable performance loss.

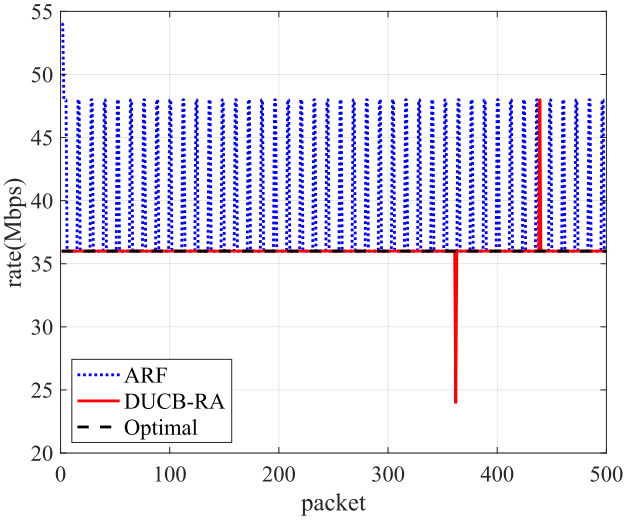


FIGURE 2. The rate selection in the static channel.

Next, we study the RA performance in a time-varying channel. In this simulation, 500 packets are sent within $T = 1000$. The channel state transitions are generated by the HMM model with neighborhood transition probability of 0.04. There are 9 channel state transitions in the specific realization. To track the time-varying channel, the parameters γ and ξ are set to 0.95 and 0.65 respectively. The obtained throughputs of different RA algorithms are shown in Fig. 3. From this figure, we observe that the proposed DUCB-RA algorithm can track the time-varying channel better than other algorithms and provide the best average throughput among all RA algorithms. From this figure, we also observe that all RA algorithms can sense the degradation of the channel better than sense the improvement of the channel. It is natural as the loss of packet is easy to observe. On the other hand, when the channel condition improves, appropriated exploration mechanism is needed to track the changes of the channel.

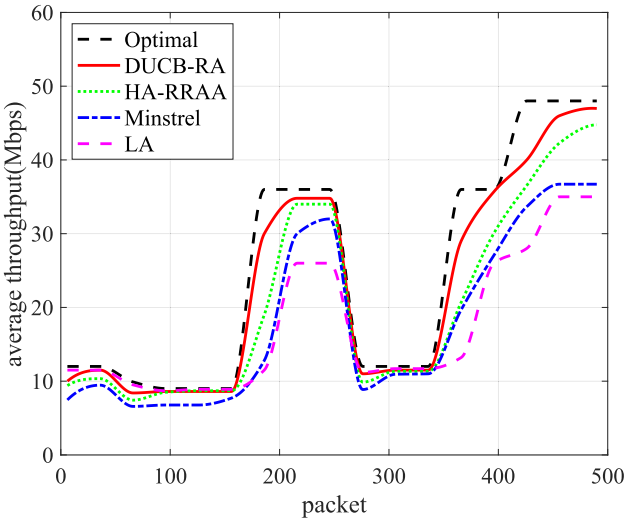


FIGURE 3. Throughput in a time-varying channel.

Let's further study how the RA algorithms adapt to the time-varying channel by showing their real-time rate selection activities in one realization of the simulations. In Fig. 4, we show the real-time rate selection results of the proposed

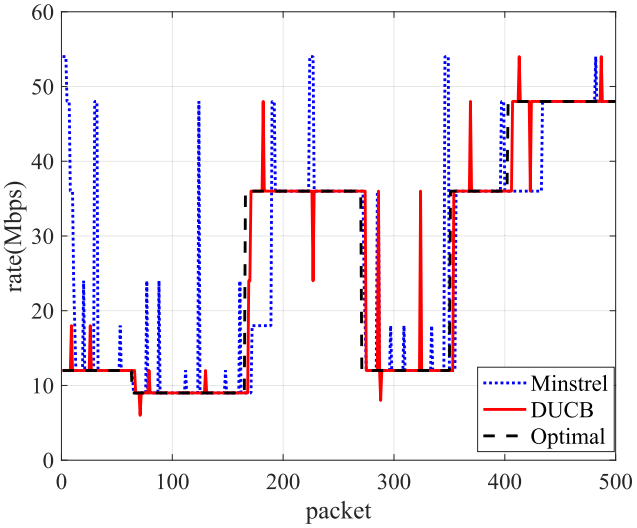


FIGURE 4. The rate selection in a time-varying channel.

algorithm and the Minstrel algorithm in the above simulation. From Fig. 4, we observe that the proposed DUCB-RA algorithm can track the time-varying channel better than the Minstrel algorithm. When the variation of current channel is gentle, DUCB-RA algorithm does not explore other rates as aggressively as Minstrel. Minstrel random selects a rate in exploration transmissions which is higher than or equal to the previous best one. The random probe mechanism causes the performance loss in the stationary channel, and is inefficient when the channel condition improves fast.

In the third case, we study the impact of delayed feedback. We compare the simulation results with the cases that ACK/NACK signals are fed back timely. The normalized throughputs are shown in Table 2 for different RA algorithms. We observe that comparing to the timely fed-back ACK/NACK, all RA algorithms are getting worse if the ACK/NACK signal are delayed. The performance degradation with delayed feedback is small for Minstrel and the proposed DUCB-RA algorithms, suggesting that the rate adaptation algorithms of these two algorithms are robust against feedback delays.

TABLE 2. Normalized throughput in timely and delayed feedback cases.

Channel setting	Normalized throughput				
	ARF	LA	Minstrel	HA-RRAA	DUCB-RA
Delayed feedback	0.82	0.74	0.85	0.87	0.92
Timely feedback	0.85	0.79	0.86	0.89	0.93

In the last simulation, we study the regret performance of the proposed DUCB-RA algorithm. The accumulated regret of the proposed algorithm in different channel conditions are shown in Fig. 5. In this simulation, channel 1 is the static channel, channel 2 is the time-varying channel with timely feedback and channel 3 is the time-varying channel with delayed feedback. 2500 packets are sent within $T = 5000$. The channel state transitions are generated by the HMM model with neighborhood transition probability of 0.04. There are 100 channel state transitions in channel 2 and channel 3. The parameters settings are $\gamma = 1$, $\xi = 0.1$, $\gamma = 0.9$, $\xi = 0.7$ and $\gamma = 0.9$, $\xi = 0.7$ respectively in these three channel conditions. We observe that the regret performance of the proposed DUCB-RA algorithm achieves a sub-linear order in all three cases, indicating asymptotically optimal throughput.

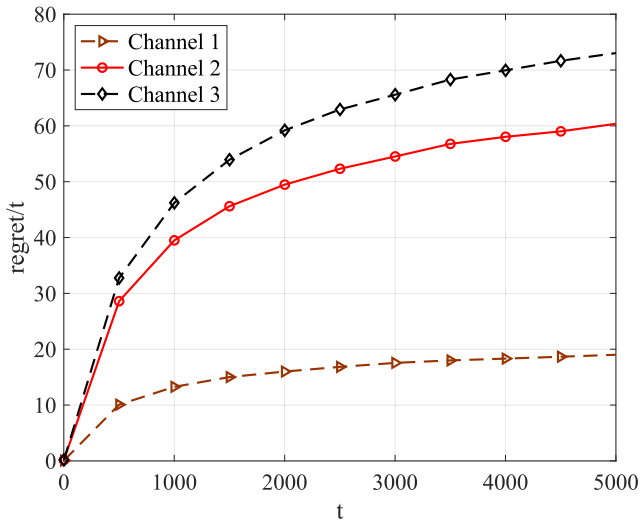


FIGURE 5. Regret performance in different channels.

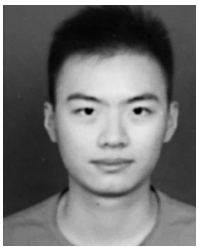
VI. CONCLUSION

In this paper, a robust rate adaptation algorithm DUCB-RA for time-varying channels is proposed. The rate selection problem is formulated as a non-stationary MAB problem. The proposed algorithm utilizes both the ACK/NACK and RSSI information to select appropriate rate. Specifically, the ACK/NACK signals are treated as delayed responses to the quality of the transmission, which is typical in practice yet is ignored in most studies. Thus our algorithm is more accurate when choosing appropriate rate. We show that the performance of the proposed algorithm is upper bounded by a sub-linear order with mathematical proofs. Simulation results demonstrate that the proposed algorithm can achieve better performance than existing rate adaptation schemes in both static and time-varying channel conditions.

REFERENCES

[1] X. Zhu, S. Chen, H. Hu, X. Su, and Y. Shi, "TDD-based mobile communication solutions for high-speed railway scenarios," *IEEE Wireless Commun.*, vol. 20, no. 6, pp. 22–29, Dec. 2013.
[2] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, Z. Zhang, E. Dutkiewicz, and T. Lv, "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2488–2524, 3rd Quart., 2019.
[3] T. Y. Arif, R. Munadi, and Fardian, "Evaluation of the Minstrel-HT rate adaptation algorithm in IEEE 802.11n WLANs," *Int. J. Simul., Syst. Sci. Technol.*, vol. 18, no. 1, pp. 1–7, Mar. 2017.
[4] L. Kriara and M. K. Marina, "SampleLite: A hybrid approach to 802.11n link adaptation," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 2, pp. 4–13, Apr. 2015.
[5] Z.-C. Dong, P.-Z. Fan, X.-F. Lei, and E. Panayirci, "Power and rate adaptation based on CSI and velocity variation for OFDM systems under doubly selective fading channels," *IEEE Access*, vol. 4, pp. 6833–6845, Oct. 2016.
[6] S. Li, M. Sun, Y.-C. Liang, B. Li, and C. Zhao, "Spectrum sensing for cognitive radios with unknown noise variance and time-variant fading channels," *IEEE Access*, vol. 5, pp. 21992–22003, Oct. 2017.
[7] G. Ku and J. M. Walsh, "Resource allocation and link adaptation in LTE and LTE advanced: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1605–1633, 3rd Quart., 2015.
[8] A. Kamerman and L. Monteban, "WaveLAN-II: A high-performance wireless LAN for the unlicensed band," *Bell Labs Tech. J.*, vol. 2, no. 3, pp. 118–133, Summer 1997.
[9] J. Zhang, K. Tan, J. Zhao, H. Wu, and Y. Zhang, "A practical SNR-guided rate adaptation," in *Proc. IEEE INFOCOM 27th Conf. Comput. Commun.*, Phoenix, AZ, USA, Apr. 2008, pp. 2083–2091.
[10] R. Combes, J. Ok, A. Proutiere, D. Yun, and Y. Yi, "Optimal rate sampling in 802.11 systems: Theory, design, and implementation," *IEEE Trans. Mobile Comput.*, vol. 18, no. 5, pp. 1145–1158, May 2019.
[11] J. Camp and E. Knightly, "Modulation rate adaptation in urban and vehicular environments: Cross-layer implementation and experimental evaluation," *IEEE/ACM Trans. Netw.*, vol. 18, no. 6, pp. 1949–1962, Dec. 2010.
[12] M. Lacage, M. H. Manshaei, and T. Turletti, "IEEE 802.11 rate adaptation: A practical approach," in *Proc. ACM Int. Symp. Modeling, Anal. Simul. Wireless Mobile Syst. (MSWiM)*, Venice, Italy, vol. 2004, pp. 126–134.

[13] D. Xia, J. Hart, and Q. Fu, "Evaluation of the minstrel rate adaptation algorithm in IEEE 802.11g WLANs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 2223–2228.
[14] J. C. Bicket, "Bit-rate selection in wireless networks," M.S. thesis, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2005.
[15] L. Deek, E. Garcia-Villegas, E. Belding, S.-J. Lee, and K. Almeroth, "Joint rate and channel width adaptation for 802.11 MIMO wireless networks," in *Proc. IEEE Int. Conf. Sens., Commun. Netw. (SECON)*, New Orleans, LA, USA, Jun. 2013, pp. 167–251.
[16] R. Crepaldi, J. Lee, R. Etkin, S.-J. Lee, and R. Kravets, "CSI-SF: Estimating wireless channel state using CSI sampling & fusion," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 154–162.
[17] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures*, document 36.213, 3rd Generation Partnership Project (3GPP), Specification (TS), Version 14.2.0, Apr. 2017. [Online]. Available: https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx
[18] R. S. Sutton and G. A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[19] P. Sadeghi, R. Kennedy, P. Rapajic, and R. Shams, "Finite-state Markov modeling of fading channels—A survey of principles and applications," *IEEE Signal Process. Mag.*, vol. 25, no. 5, pp. 57–80, Sep. 2008.
[20] J. P. Pavon and S. Chio, "Link adaptation strategy for IEEE 802.11 WLAN via received signal strength measurement," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Anchorage, AK, USA, May 2003, pp. 1108–1113.
[21] I. Pefkianakis, S. H. Y. Wong, H. Yang, S.-B. Lee, and S. Lu, "Toward history-aware robust 802.11 rate adaptation," *IEEE Trans. Mobile Comput.*, vol. 12, no. 3, pp. 502–515, Mar. 2013.
[22] L.-N. Zhang, X. Zuo, J.-W. Liu, W.-M. Li, and N. Ito, "Comments on finite-time analysis of the multiarmed bandit problem," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, Jul. 2019, pp. 235–256.
[23] A. Garivier and E. Moulines, "On upper-confidence bound policies for switching bandit problems," in *Proc. Springer Int. Conf. Algorithmic Learn. Theory (ALT)*, Espoo, Finland, 2011, pp. 174–188.
[24] P. Joulani, A. Gyorgy, and C. Szepesvári, "Online learning under delayed feedback," in *Proc. Int. Conf. Mach. Learn. (PMLR)*, Atlanta, GA, USA, 2013, pp. 1453–1461.



YAPENG ZHAO received the B.S. degree from the College of Communication Engineering, Jilin University, Changchun, China, in 2017. He is currently pursuing the M.S. degree in communication and information system from Shanghai University, Shanghai, China. His current research interest includes wireless communication systems.

HUA QIAN (Senior Member, IEEE) received the B.S. and M.S. degrees from Tsinghua University, in 1998 and 2000, respectively, and the Ph.D. degree from the Georgia Institute of Technology, in 2005. He is currently a Professor with the Shanghai Advanced Research Institute, Chinese Academy of Sciences. He is also an Adjunct Professor with Shanghai Tech University. He has over ten years of research and development experience in signal processing, wire-less communications, and ASIC design. He has coauthored two book chapters, published 100 SCI/EI indexed articles, and applied for 60 patents, including four granted U.S. patents. His current research interests include nonlinear signal processing, distributed signal processing, and system design of wireless communications.

KAI KANG (Member, IEEE) received the Ph.D. degree from Tsinghua University, in July 2007. He was a Postdoctoral Research Fellow with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, from May 2012 to October 2015. He is currently a Professor with the Shanghai Advanced Research Institute, Chinese Academy of Sciences. He has published more than 20 academic articles and applied for more than 30 invention patents on 5G and wireless communication systems. His research interests include 5G and Wi-Fi networks, signal processing in wireless communication systems, and so on.

YANLIANG JIN received the B.S. and M.S. degrees in electrical engineering from Xidian University, Xi'an, China, in 1997 and 2000, respectively, and the Ph.D. degree in communication and information system from Shanghai Jiaotong University, in 2005. He currently holds an Associate Professorship with the School of Communication and Information Engineering (SCIE), Shanghai University (SHU). His research interests include mobile ad hoc networks (MANETs), wireless sensor networks (WSNs), wireless multimedia sensor networks (WMSNs), wireless broadband access and signal processing. He has published more than 30 journal articles and conference papers.