Transmission Scheduling of P2P Real-Time Communication Based on Restless Multi-Armed Bandit

Ting Wu
Guangxi Normal University
Shikun Tian
Guangxi Normal University
Shengda Tang (tangsd911@163.com)
Guangxi Minzu University https://orcid.org/0000-0003-2802-1498

Research Article

Keywords: P2P real-time communication system, transmission scheduling strategy, restless multi-armed bandit, Whittle index

Posted Date: October 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1841963/v1

Transmission Scheduling of P2P Real-Time Communication Based on Restless Multi-Armed Bandit

Ting Wu, Shikun Tian, Shengda Tang

Abstract

In this paper, we consider a transmission scheduling problem of the point-to-point (P2P) real-time communication over fading channel. Specifically, we assume each transmission task with a strict time delay arrives into the system randomly, we consider the transmission scheduling strategy under the time delay constraint of the communication system so as to maximize the total discounted expected reward of the system. The communication model is formulated in the framework of the Markov decision process (MDP), and since the curse of dimensionality of the proposed MDP architecture, the transmission scheduling problem is analyzed based on the restless multi-armed bandit (RMAB) process. We show the existence of the Whittle index of the P2P transmission communication, and based on which, the closed solution of Whittle index based transmission scheduling strategy is obtained. Finally, the numerical results are given to verify the availability of the Whittle index based transmission scheduling algorithm, and it is also shown that our proposed transmission scheduling algorithm is with low computational complexity, and it reduces the computational time greatly in comparison to the MDP method.

The work described in this paper was supported by National Natural Science Foundation of China (No. 61761008), Natural Science Foundation of Guangxi (No. 2018JJA170024), and Innovation Project of Guangxi Graduate Education (No. YJSCXP202105).

T. Wu is with the School of Mathematics and Statistics in Guangxi Normal University, Guilin, Guangxi, P. R. China. (email: wuting038766@stu.gxnu.edu.cn).

S. Tian is with the School of Mathematics and Statistics in Guangxi Normal University, Guilin, Guangxi, P. R. China. (email: tianshikunlike1@163.com).

S. Tang is with the School of Mathematics and Statistics in Guangxi Normal University and also with Center for Applied Mathematics of Guangxi (Guangxi Normal University), Guilin, Guangxi, P. R. China. (email: tangsd911@163.com).

Keywords: P2P real-time communication system, transmission scheduling strategy, restless multi-armed bandit, Whittle index.

I. Introduction

With the increase in popularity of the internet communication and explosive growth of data interaction, the demand for high-speed communication links becomes increasing greatly and the types of network information are more and more complex [1]. To overcome these challenges, scholars have made a great effort to solve the related communication scheduling problems, and the relevant available technologies include digital subscriber line conversion (DSL) [2], point-to-point (P2P) wireless communication [3], cable hybrid system [4], satellite link [5], [6], etc. Among them, the P2P communication has great advantages in long-distance real-time communication because of its wireless transmission characteristics. The convenience, easy implementation and high security of P2P communication makes it increasingly popular in real-time wireless communication.

Recently, there has been an increasing research interests being directed to P2P real-time communication networks. For example, in [7] a network design method is proposed to meet the requirements of remote P2P dense wavelength division multiplexing (DWDM) link for long-distance communication. To ensure accurate tracking and stable connection of P2P communication, the authors of [8] propose a low-cost solution for the tracking antenna in millimeter wave band in 5G communication. For the P2P link of integrated satellite aviation network, which is considered as one of the key driving factors of the vision of the 6G wireless network, the literature [6] divides the space network to study the possible properties of the connection link (i.e. radio-frequency or free space optics). In [9], the authors consider the transmission scheduling problem and investigate the real-time communication strategy under the constraints of random packet arrival, packet loss and heterogeneous deadline. Considering the real-time transmission in fading channel, the reference [10] studies the P2P wireless transmission problem when channel state information (CSI) cannot be observed, and studies it by modeling the scheduling problem as partially observable Markov decision process (POMDP). Based on the finite state Markov channel model and ARQ retransmission protocol, the authors of [11] consider the real-time transmission strategy of data packets from users to base stations, and the threshold structure of the transmission strategy is proved. A review of recent results and theoretical approaches related

to P2P wireless communication system can be found in [12] and the references therein.

Wireless data traffic is exponentially increasing and such trend is expected to continue in the coming years. In fact, most commercial radio-communication infrastructures, including AM/FM and high-definition TV broadcasting, as well as GPS, satellite, cellular, and WiFi communications, are limited to the relatively narrow portions of the spectrum between 300 MHz and 3 GHz where electromagnetic propagation conditions are more favorable and low-cost semiconductor technologies are easily available [13]. However, the above frequency range is excessively crowded. For this reason, great efforts have been made to develop wireless communication systems operating at mm-wave frequency, in which larger bandwidth is available in order to satisfy future capacity requirements [12]. Due to objective factors such as path, atmosphere, scattering, refraction, etc., most actual channels are fading [14], [15]. Fading is one of the main characteristics of the channel, which affects the channel capacity and communication [16]. Therefore, considering the P2P real-time communication over fading channel can be more suitable for practical application scenarios.

Compared with the above literatures, the main contributions of our work are as follows.

- In this paper, we consider the transmission scheduling problem of the P2P real-time communication system over the fading channel. Specifically, CSI is added to the system model to supplement the neglect of CSI related effects in literature [9]. Considering the influence of CSI on P2P real-time communication system, the wireless transmission system is modeled as a finite state Markov decision process (FSMDP).
- By introducing Lagrange multiplication factor, we reformulate the P2P transmission system as a restless multi-armed bandit (RMAB) process. Based on these, the indexability of the transmission scheduling problem is proved, and the closed expression of the Whittle index is obtained. Furthermore, the closed expression of Whittle index transmission strategy for P2P real-time communication is given, which extends the threshold expression of the threshold structure of the optimal transmission scheduling strategy in reference [11].
- The simulation results have been done to show the advantages of the proposed algorithm. Compared with the traditional algorithm of MDP, the Whittle index based transmission scheduling algorithm proposed in this paper can effectively reduce the calculation cost and time without losing the accuracy. In addition, compared with other general strategies, the strategy obtained by this algorithm is obviously better.

The rest of this paper is organized as follows. Section II describes the P2P real-time communication model. Section III formulates the transmission scheduling problem under the framework of the MDP. Section IV proves the indexability of P2P real-time communication system, and the closed expression of the Whittle index based transmission scheduling strategy is developed. Section V gives the numerical results, and finally, the conclusion is presented in Section VI.

II. System Model and Problem Formulation

This paper mainly considers the transmission scheduling problem of the multi-user P2P real-time communication over the fading channel, and the communication model is described as follows. The main notations used in this paper are shown in Table I.

TABLE I

Main notations of P2P communication system model

N	Number of P2P pairs in communication system
Z	Maximum channel capacity
X_t	Channel state at time t
M	Total number of channel state
$\mathcal{C}$	Channel state space
γ_{mn}	Transit probability from γ_m to γ_n
B_i	Size of each transmission task of the i-th P2P pair
L_i	Time delay of each transmission task of the i-th P2P pair
$p(c_m)$	Transmission success probability when channel state is c_m
Q_i	Arrival probability of the new task for the i-th P2P
S_t	State of the system at time t
S_0	Initial state of the system
$\mathcal{S}_i$	State set of the i-th P2P pair
$s_{i,t}$	State of the i-th P2P pair at time t
a_i	Action of P2P pair i taken by the system
a_i^*	Optimal action of P2P pair i taken by the system
g	adopted transmission strategy
$R(S_t, A_t)$	Reward obtained by the system at time t
β	Discount factor.
$f(b_{i,r})$	Penalty function when the residual packet is $b_{i,r}$
w	Subsidy of system when action is not transmission
$w^*(s_{i,t})$	Whittle index of the state $s_{i,t}$ of the i-th P2P pair

A. System Model

In this paper, we consider a discrete-time communication system. The time is divided into equal slots, and we denote the time slot set by $\mathcal{H} \triangleq \{0, 1, 2, \ldots\}$.

As described in Fig. 1, we consider a typical P2P wireless real-time communication sytem, in which there are N P2P pairs, the transmission point (TP) and the receiving point (RP) of each P2P pair can communicate directly over a wireless channel.

1) Channel Model: Similar to [17], we first introduce a finite state Markov chain which is denoted by $X = \{X_t, t \in \mathcal{H}\}$ to describe the fading channel. More specifically, the fading channel is assumed to be divided into M non overlapping intervals, each interval is mapped into a channel state, and we denote the channel state space as $\mathcal{C} \triangleq \{c_1, c_2, \ldots, c_M\}$. Assuming that in each time slot the channel state does not vary, while it can change in different time slots. Let the transition probability matrix of X be $\mathbf{\Gamma} = [\gamma_{mn}]_{M\times M}$, where γ_{mn} represents the probability of making a transition from state c_m to state c_n, i.e.,

$$\gamma_{mn} = P(X_{t+1} = c_n | X_t = c_m).$$

We refer to X as the Markov fading channel model.

The Markov channel model describes the evolution of channels with different fading effects. The binary symmetric channel (BSC) and the Gilbert Elliot model are two special cases of the Markov channel model described above [17]–[19].

Throughout this paper, we suppose the current state of the channel is $X_t = c_m$ and the next state of the channel is $X_{t+1} = c_n$ to simplify the notation.

2) Communication Model: In the P2P real-time communication system, the TP sends the data task to the RP within a given time delay over the fading channel. We assume that, for each P2P pair, at most one packet of the data task can be transmitted in each time slot. The RP can obtain the CSI of the wireless communication system and transmit it to the TP and base station through a separate feedback channel. The base station schedules the data transmission at the TP according to CSI, so as to achieve better scheduling decision and improve the performance of wireless communication system. For the i-th P2P pair, $i \in \mathcal{N} \triangleq \{1, 2, \ldots, N\}$, we denote the size of the k-th incoming transmission task (data) by B_i^k, and its time delay by L_i^k. That is, the TP of the i-th P2P pair needs to transmit B_i data packets within L_i time slots. Without loss of generality, we let $B_i^k = B_i$, $L_i^k = L_i$ for all $k \in \{1, 2, \ldots\}$. Due to the limitation of channel

resources, we assume that, in each time slot, at most $Z(\leq N)$ P2P pairs can communicate with each other simultaneously. Therefore, in each time slot the system should decide that which P2P pairs should communicate over the channel according to the current state.

Since the packet transmission is affected by the channel state, we let $p(c_m)$ be the probability that the transmission is successful when the channel state is c_m, when the transmission fails, the packet is allowed to be retransmitted. For each transmission task, when its time delay ends, the residual task still waiting for transmission in the data buffer will be lost, and the new transmission task will arrive with some probability. We denote the arrival probability of the new task by $Q_i \in [0, 1]$ for the i-th P2P.

At time $t \in \mathcal{H}$, we denote the number of remaining packets in the i-th P2P pair by $B_{i,t}$, and the corresponding residual transmission time by $L_{i,t}$, respectively. It is easy to see that $B_{i,t} \in \mathcal{B} \triangleq \{0, 1, \ldots, B_i\}$ and $L_{i,t} \in \mathcal{L} \triangleq \{0, 1, \ldots, L_i\}$. It should be noted that $B_{i,t+L_{i,t}}$ means the remaining data task at the end of the transmission deadline. We refer to $s_{i,t} \triangleq [L_{i,t}, B_{i,t}]$ as the state of i-th P2P pair, and $S_t \triangleq [X_t, s_{1,t}, s_{2,t}, \ldots, s_{N,t}]$ as the state of the P2P communication system at time t, respectively.

Now, we obtain our P2P real-time communication system model

$$S = \{S_t, t \in \mathcal{H}\} \tag{1}$$

with state space $\mathcal{S} \triangleq \mathcal{C} \times \mathcal{B} \times \mathcal{L}$. The stochastic process S describes the evolution of the P2P real-time communication system with random arrival data task over the fading channel.

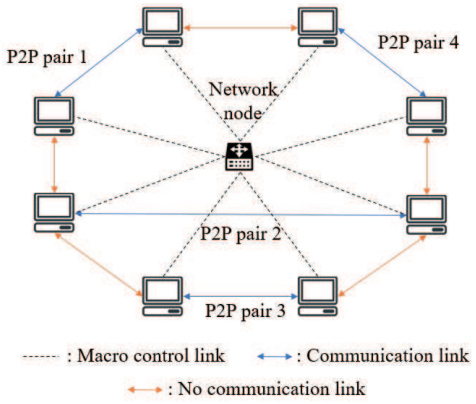

Fig. 1. Diagram of the P2P real-time communication system.

B. Transmission scheduling Problem

Let a_i be the action taken by the i-th P2P pair. Specifically, at each decision time epoch, let $a_i = 1$ represent that the TP of the i-th P2P pair transmits the data packet, and $a_i = 0$ represent that the TP does not transmit the data packet. Denote

$$\mathcal{A} = \Big\{ [a_1, a_2, \ldots, a_N] \Big| \sum_{i=1}^{N} a_i \leq Z \Big\},$$

and $\mathcal{A}$ is referred to as the action space of the P2P communication system. The condition $\sum_{i=1}^{N} a_i \leq Z$ indicates that at each time slot, at most Z P2P pairs are allowed to transmit their data packets due to the limitation of the channel resource.

Hypothetically, when the data packet is successfully transmitted, the system can obtain reward $r > 0$, otherwise the reward is 0. Let the transmission cost be $h(c_m)$ when the channel state is c_m. Denote

$$g = \{A_t(S_t) | A_t(S_t) \in \mathcal{A}, S_t \in \mathcal{S}, t \in \mathcal{H}\},$$

and we call g transmission strategy adopted for the P2P communication system.

Let $R(S_t, A_t)$ be the reward obtained by the communication system at time t, then the total expected discount reward under strategy g is defined as

$$V_g(S_t) = \mathbb{E}\Big[\sum_{t=0}^{\infty} \beta^t R(S_t, A_t) \Big| S_0 \Big], \tag{2}$$

where $\beta \in (0, 1)$ is the discount factor, and $\mathbb{E}[\cdot|S_0]$ is the conditional expectation when the initial state is S_0.

Our main objective is to find a strategy $g^* = \{A_0^*, A_1^*, \ldots\}$ such that

$$V_{g^*}(S_t) \triangleq V(S_t) = \max_{g \in \mathcal{A}} \Big\{ \mathbb{E}\Big[\sum_{t=0}^{\infty} \beta^t R(S_t, A_t) | S_0 \Big] \Big\} \tag{3a}$$

$$s.t.$$

$$B_{i,t+L_{i,t}} \leq 0 \quad for\ i \in \mathcal{N},\ t \in \mathcal{H}. \tag{3b}$$

The constraint condition (3b) indicates that each transmission task should be completed at the end of its time delay. The action sequence g^* is called the optimal transmission strategy, and $V(S_0)$ is the maximum total expected discount reward of the P2P communication system when the initial state is given as $S_0 \in \mathcal{S}$.

Now, the transmission scheduling problem is formulated, and in the following we will explore its solution in detail.

III. MDP FRAMEWORK FOR SCHEDULING PROBLEM

In this section, the transmission scheduling problem is formulated in framework of the MDP.

To obtain the MDP framework of the scheduling problem, the fist challenge is to remove the constraint condition in the transmission scheduling problem. To do that, we introduce the penalty function [18]. Specifically, when the task is completed within the given delay time, the penalty function is set to 0, while the task is not completed within the delay time, the system will get a great penalty (negative return). In order to obtain the maximum benefit, the system will avoid the transmission scheduling strategies that cannot complete the transmission task within the given time. Thus, the delay constraint (3b) is satisfied.

Based on this idea, we reconstructs MDP from state space, action space, transition probability and reward function as follows.

1) State space: $\mathcal{S}$.

2) Action space: $\mathcal{A}$.

3) Transition probability: Given the state of the i-th P2P pair $s_{i,t} = [L_{i,t}, B_{i,t}]$ at time t, then the following cases will be occurred:

(i) Case 1: $a_i = 1$ and $L_{i,t} > 1$.

In this case, if the transmission is successful, then the next state is $s_{i,t+1} = [L_{i,t} - 1, B_{i,t}^+]$, where $B_{i,t}^+ \triangleq \max\{(B_{i,t} - a_i), 0\}$. If the transmission is unsuccessful, then the next state is $s_{i,t+1} = [L_{i,t} - 1, B_{i,t}]$.

(ii) Case 2: $a_i = 0$ and $L_{i,t} > 1$.

In this case, the next state is $s_{i,t+1} = [L_{i,t} - 1, B_{i,t}]$.

(iii) Case 3: $L_{i,t} \leq 1$.

In this case, the residual task should be removed in the next time slot due to the end of the transmission time. Thus, the next state becomes $s_{i,t+1} = [L_i, B_i]$ with probability Q_i, or $s_{i,t+1} = [0, 0]$ with probability $(1 - Q_i)$.

Then, given $s_{i,t} = [L_{i,t}, B_{i,t}]$ and $X_t = c_m$, we can build the transition probability of the i-th P2P pair as follows.

When $L_{i,t} \geq 1$ and $B_{i,t} \geq 1$,

$$P(s_{i,t+1}|s_{i,t}, a_i) = \begin{cases} p(c_m) & s_{i,t+1} = [L_{i,t} - 1, B_{i,t}^+], \\ \bar{p}(c_m) & s_{i,t+1} = [L_{i,t} - 1, B_{i,t}]. \end{cases}$$

When $L_{i,t} < 1$ or $B_{i,t} = 0$,

$$P(s_{i,t+1}|s_{i,t}, a_i) = \begin{cases} Q_i & s_{i,t+1} = [L_i, B_i], \\ \bar{Q}_i & s_{i,t+1} = [0, 0], \end{cases}$$

where we let $\bar{p}(c_m) \triangleq 1 - p(c_m)$ and $\bar{Q}_i \triangleq 1 - Q_i$.

Let $P(S_{t+1}\,|S_t, A_t\,)$ represent the transition probability of the P2P communication system from S_t to S_{t+1} when action A_t is taken, then we have

$$P(S_{t+1}|S_t, A_t) = \gamma_{mn} \prod_{i=1}^{N} P(s_{i,t+1}|s_{i,t}, a_{i,t}).$$

4) Reward function: Define $f_i(b)$ as the penalty function, where b is the remaining number of packets in the i-th of P2P pair at the deadline. When $b = 0$, let $f_i(0) = 0$, i.e., there is no penalty when the transmission task is completed. When $b > 0$, let $f_i(b)$ be an increasing concave function of b, i.e., the more the packets remain at the time deadline, the greater the penalty is. Then the reward of the i-th P2P pair can be defined as

$$R_i(s_{i,t}, a_i) = \begin{cases} (rp(c_m) - h(c_m))a_i & L_{i,t} > 1, B_{i,t} > 0, \\ (rp(c_m) - h(c_m))a_i - p(c_m)f_i(B_{i,t} - a_i) & \\ -\bar{p}(c_m)f_i(B_{i,t}) & L_{i,t} = 1, B_{i,t} > 0, \\ 0 & otherwise. \end{cases}$$

The reward received by the communication system at state S_t by taking action A_t is given as

$$R(S_t, A_t) = \sum_{i=1}^{N} R_i(s_{i,t}, a_i). \tag{4}$$

Now, we transform the P2P communication scheduling problem into an infinite horizon discounted MDP, and the total expected discount reward function $V(S_t)$ satisfies the following Bellman equation:

$$V(S_t) = \max_{A_t \in \mathcal{A}} \Big\{ R(S_t, A_t) + \beta \sum_{S_{t+1} \in S} P(S_{t+1}|S_t, A_t) V(S_{t+1}) \Big\}. \tag{5}$$

Based on the MDP theory [19], there exits an optimal stationary transmission scheduling strategy to maximize the total expected discount reward. The optimal strategy and the maximum

total expected reward can be solved by using algorithms such as the value iteration (VI), and Algorithm 1 gives the steps of VI algorithm.

In the P2P real-time communication system, the size of state space $\mathcal{S}$ is $Ml^N b^N$, where b and l are the cardinalities of $\mathcal{B}$ and $\mathcal{L}$. We can see that with the increase of number of users, the size of state space $\mathcal{S}$ increases exponentially. This indicates that the MDP model of P2P communication system suffers from the dimension disaster, and more computational power and time will be required.

Algorithm 1 VI algorithm for the optimal transmission scheduling strategy.

Step 1. Given $\delta > 0$, let $n = 0$, $\forall s \in \mathcal{S}$, initialization $V_0(s) = 0$;

Step 2. $\forall s \in \mathcal{S}$, compute:

$$V_{n+1}(s) = \max_{a \in \mathcal{A}} \{R(s;a) + \beta \sum_{s' \in \mathcal{S}} P_{ss'}(a) V_n(s')\};$$

Step 3. If $\max\left\{\delta, \left|V_n(s) - V_{n+1}(s)\right|\right\} \leq \delta$, go to step 4;
else let $n \leftarrow n+1$, return to step 2;

Step 4. $\forall s \in S$, calculate the optimal transmission strategy and the optimal value function:

$$a^*(s) = \arg\max_{a \in \mathcal{A}} \left\{R(s,a) + \beta \sum_{s' \in \mathcal{S}} P(s'|s,a) V_{n+1}(s')\right\},$$

$$V(s) = V_{n+1}(s).$$

IV. RMAB FRAMEWORK FOR SCHEDULING PROBLEM

To solve the "curse of dimensionality" problem described above, in this section, we introduce the Lagrange multiplication factor to resolve the transmission scheduling problem under the framework of the RMAB model.

A. Modelling Scheduling Problem by RMAB

We first define

$$R_i^w(s_{i,t}, a_i) = R_i(s_{i,t}, a_i) + wI_{\{a_i=0\}},$$

and R_i^w is called w-subsidy reward function, where w is the Lagrange multiplication factor, it can be considered as the additional subsidy received by the i-th P2P pair when the action

0 is taken. Then, the maximum total discounted expected reward function $V^w(s_{i,t})$ based on w-subsidy return can be expressed as

$$V_i^w(s_{i,t}) = \max_{a_i \in \{0,1\}} \{R^w(s_{i,t}, a_i) + \beta \sum_{s_{i,t+1} \in \mathcal{S}_i} P(s_{i,t+1}|s_{i,t}, a_i) V_i^w(s_{i,t+1})\}. \tag{6}$$

Here, $\mathcal{S}_i \triangleq \mathcal{L} \times \mathcal{B}$ denote the state space of the i-th P2P pair.

Give $s_{i,t}$, denote

$$J_0(s_{i,t}) \triangleq R_i^w(s_{i,t}, 0) + \beta \sum_{s_{i,t+1} \in \mathcal{S}_i} P(s_{i,t+1}|s_{i,t}, 0) V_i^w(s_{i,t+1}),$$

$$J_1(s_{i,t}) \triangleq R_i^w(s_{i,t}, 1) + \beta \sum_{s_{i,t+1} \in \mathcal{S}_i} P(s_{i,t+1}|s_{i,t}, 1) V_i^w(s_{i,t+1}).$$

Now, we give the following definitions.

Definition 1 (Whittle's index [20]) Given the state $s_{i,t}$, let

$$w_i^*(s_{i,t}) \triangleq \inf \Big\{ w | J_0(s_{i,t}) \geq J_1(s_{i,t}) \Big\},$$

and $w_i^*(s_{i,t})$ is referred to as the Whittle index of the state $s_{i,t}$ of the i-th P2P pair.

The monotonicity of $w_i^*(s_t)$ is characterized by the following lemma.

lemma 1 If the Whittle index exists given the states of the i-th P2P pair $s_{i,t} = [L_{i,t}, B_{i,t}]$, then

(1) $w_i^*(s_{i,t})$ is non-increase function of $L_{i,t}$;

(2) $w_i^*(s_{i,t})$ is non-decrease function of $B_{i,t}$.

Proof. The proof follows by a slight modification of the argument in [[11], Theorem 1 and Theorem 2], and we omit the proof in this paper. ■

Let $\Pi_i(w)$ be the set of states under which it is optimal to unschedule the i-th P2P pair in the w-subsidy problem, i.e.,

$$\Pi_i(w) \triangleq \{s_{i,t} | a_i^*(s_i) = 0\},$$

and $\Pi_i(w)$ is called the passive set.

Definition 2 (Indexability [20]) If $\Pi_i(w)$ increases with respect to w, i.e., for any $w_1, w_2 \in \mathbb{R}$, when $w_1 \leq w_2$, $\Pi_i(w_1) \subseteq \Pi_i(w_2)$, then the i-th P2P pair is said to be indexable.

B. Indexability and Whittle index

We now establish the indexability of the RMAB problem by considering the w-subsidy single P2P pair (arm) reward maximization problem. That is, our first main result is given as follows.

Theorem 1. The RMAB of the P2P wireless transmission is indexable.

Proof. The detailed proof is given in Appendix A. ■

Theorem 1 provides a theoretical basis for Whittle index-based transmission scheduling strategy, and our second main conclusion of this paper is the following theorem, in which the closed analytic formula of the Whittle index based transmission scheduling strategy is developed.

Theorem 2. For the P2P real-time communication system, the Whittle index based transmission scheduling strategy of the i-th P2P pair at state $s_{i,t} = [L_{i,t}, B_{i,t}]$ can be expressed as follows

$$w_i^*(s_{i,t}) = \begin{cases} -h(c_m) & B_{i,t} = 0 \\ rp(c_m) - h(c_m) + p(c_m)\delta(B_{i,t} - 1) & \\ & L_{i,t} = 1, B_{i,t} \neq 0 \\ p(c_m)(\beta\bar{\rho}(c_m))^{L_{i,t}-1}\delta(B_{i,t} - 1)/A(s_{i,t}) & \\ +A^r(s_{i,t})/A(s_{i,t}) - A^h(s_{i,t})/A(s_{i,t}) & \\ & L_{i,t} > 1, B_{i,t} \neq 0 \end{cases} \tag{7}$$

where

$$\begin{aligned} \delta(B_{i,t} - 1) &\triangleq f(B_{i,t}) - f(B_{i,t} - 1), \\ A(s_{i,t}) &\triangleq 1 - p(c_m) \sum_{i=1}^{L_{i,t}-1} \beta^i \bar{\rho}^{i-1}, \\ A^r(s_{i,t}) &\triangleq rp(c_m)(1 - \rho \sum_{i=1}^{L_{i,t}-1} \beta^i \bar{\rho}^{i-1}), \\ A^h(s_{i,t}) &\triangleq h(c_m) - p(c_m)\hbar \sum_{i=1}^{L_{i,t}-1} \beta^i \bar{\rho}^{i-1}, \\ \rho(c_m) &\triangleq \sum_{c_n \in \mathcal{C}} \gamma_{mn} p(c_n), \\ \hbar(c_m) &\triangleq \sum_{c_n \in \mathcal{C}} \gamma_{mn} h(c_n), \\ \bar{\rho}(c_m) &\triangleq= 1 - \rho(c_m). \end{aligned}$$

Then, for given subsidy w_0, the Whittle index of the i-th P2P pair is given as follows.

$$a_i^*(s_{i,t}) = \begin{cases} 1 & w_i^*(s_{i,t}) \geq w_0, \\ 0 & otherwise. \end{cases}$$

Proof. The detailed proof is given in Appendix B. ■

Based on Theorem 2, the Whittle index based transmission scheduling algorithm for P2P real-time communication is developed in Algorithm 2. In this algorithm it takes $O(N)$ time to calculate the Whittles index for all the P2P pairs in Step 1, and it may take $O((N)log2(N))$ time for the sorting process in Step 2. The time complexity of conducting Step 3 is linear in N. Thus, the time complexity of the Whittle index based scheduling algorithm is $O((N)log2(N))$.

Some comments on the Whittle index scheduling strategy are given in the following remark.

Remark Since the subsidy is zero in the original model, then the optimal action for the i-th P2P pair is unschedule (i.e., $a_i = 0$) when the Whittle index is less than 0, otherwise the optimal action is transmission (i.e., $a_i = 1$). For the P2P real-time communication system, at each decision time we just need to select at most Z P2P pairs with the largest non-negative Whittle indexes to transmit the data packet.

Algorithm 2 Whittle index based transmission scheduling algorithm for P2P real-time communication system.

Step 1. At each decision epoch, calculate the Whittle index of each P2P pair (arm) by Theorem 2;

Step 2. Sort the P2P pairs in descending order of the Whittle index;

Step 3. Select at most Z P2P pairs with the highest nonnegative Whittle index and transmit the packets in their data buffers.

V. Simulation Results

In this section, we give numerical results to investigate the Whittle index based transmission scheduling strategy of the P2P transmission communication system. The numerical setting is listed in Table II, and our numerical computations are run on a laptop with an Intel Core i5-9400 CPU, and the size of RAM is 16.0 GB.

For the P2P real-time communication system set up above, Algorithm 1 and Algorithm 2 are used to calculate the optimal transmission scheduling strategy and the maximum total expected

TABLE II

NUMERICAL SETTING OF P2P REAL-TIME COMMUNICATION SYSTEM

Channel state space	$\mathcal{C}$=$\{1, 2, 3\}$
Transition probability matrix	$\mathbf{\Gamma} = \begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.15 & 0.4 & 0.45 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$
Successful transmission probability	$p(c_m)$=[0.45 0.75 0.95]
Size of each transmission task	$B_i = 3$
Time delay of each transmission task	$L_i = 6$
Arrival probability of new data task	$Q_i = 0.5$
Reward function	$r = 3$
Penalty function	$f(b) = b^2$
Discounted factor	$\beta = 0.85$

reward under different N and Z. As shown in Fig.2 and Fig.3, Algorithm 1 and Algorithm 2 achieve almost the same total discounted reward and have almost the same strategy. In Fig.2, the total discounted rewards achieved by the Algorithm 1 are slightly larger than those of the Algorithm 2. The results demonstrate that the total discounted rewards achieved by the Whittle index based transmission scheduling algorithm are close to the optimal ones. In addition, it can be seen from Fig.3 that the strategies obtained by the two algorithms are highy consistent. Therefore, from the point of view of total discounted reward and scheduling strategy, Algorithm 2 can replace Algorithm 1 without losing accuracy.

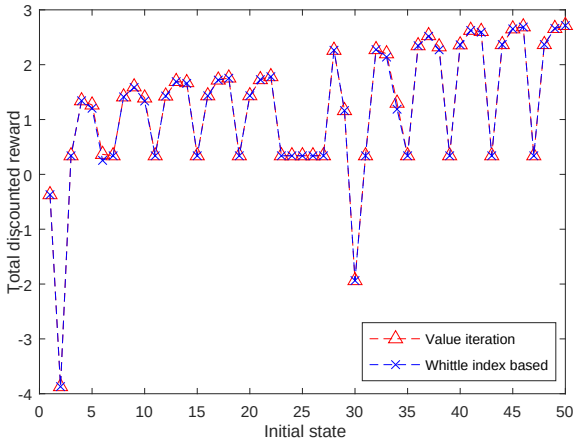

Fig. 2. When $N = 1$, the total discounted rewards under Algorithm 1 and Algorithm 2.

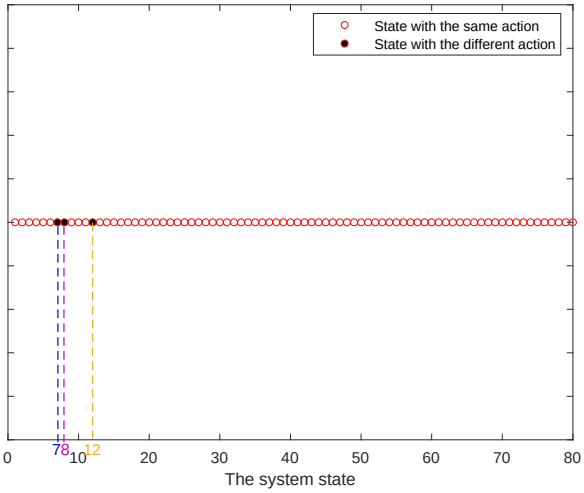

Fig. 3. When $N = 1$, the difference of the strategy under Algorithm 1 and Algorithm 2.

In order to show the effectiveness of the Whittle index based transmission scheduling strategy, we compare it with the polling strategy and the myopic strategy. The polling strategy is to make the actions in the action set be selected in turn, so that each P2P pair can be selected the same number of times as much as possible. The myopic strategy is to take the action of maximizing the current reward to select P2P pairs for transmission. In Fig.4, the total discounted rewards under the three strategies are obtained by the Monte Carlo (MC) method. From the figure, we can see that the discounted reward of Whittle index strategy is greater than that of polling and myopic strategies.

When $N = 3$, the Algorithm 1 cannot be performed under the current computing power. In fact, when $N = 3$, the number of the state of the P2P real-time communication system is 65856, and in the case of $Z = 2$, then the required calculation force is $O(N) = 65856^2 \times 7$, and the corresponding required RAM is 32.3GB. In this numerical setting, it is difficult to obtain the optimal strategy and the maximum reward by Algorithm 1, while from the Table III and Table IV, we can see that Algorithm 2 can still works when the state space becomes larger. We list the Whittle index of the i-th P2P pair in Table III, and the transmission scheduling strategy when $N \geq 3$ in Table IV. In addition, compared with Algorithm 1, Algorithm 2 can not only save the calculation cost, but also reduce the computational time, as shown in Table V.

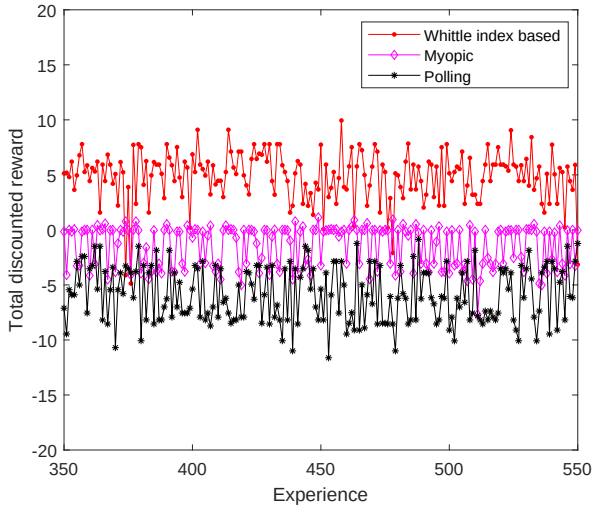

Fig. 4. When $N = 3$ and $Z = 2$, comparison of rewards of three strategies from the 350th simulation to the 550th simulation.

TABLE III

THE WHITTLE INDEX OF THE i-TH P2P PAIR, THE THREE COMPONENTS OF (a, b, c) REPRESENT THE WHITTLE INDEX OF CHANNEL STATE 1, 2, AND 3.

$L_{i,t}$	$B_{i,t} = 0$	$B_{i,t} = 1$	$B_{i,t} = 2$	$B_{i,t} = 3$
0	(-2.00,-1.00,-0.66)	(-2.00,-1.00,-0.66)	(-2.00,-1.00,-0.66)	(-2.00,-1.00,-0.67)
1	(-2.00,-1.00,-0.66)	(-0.20, 2.00, 3.13)	(0.70, 3.50, 5.03)	(1.60, 5.00, 6.93)
2	(-2.00,-1.00,-0.66)	(-1.06, 0.68, 1.83)	(-0.81, 1.38, 3.47)	(-0.57, 2.07, 5.10)
3	(-2.00,-1.00,-0.66)	(-1.25, 0.26, 1.27)	(-1.21, 0.43, 2.16)	(-1.16, 0.60, 3.05)
4	(-2.00,-1.00,-0.66)	(-1.29, 0.17, 0.91)	(-1.28, 0.20, 1.15)	(-1.27, 0.23, 1.38)
5	(-2.00,-1.00,-0.66)	(-1.30, 0.16, 0.81)	(-1.29, 0.16, 0.85)	(-1.29, 0.16, 0.89)
6	(-2.00,-1.00,-0.66)	(-1.31, 0.15, 0.79)	(-1.30, 0.15, 0.79)	(-1.30, 0.15, 0.80)

VI. CONCLUSIONS

In this paper, the P2P real-time communication system in fading channel is studied, which has strict constraints of time delay, random packet arrival and packet loss. We model the scheduling problem as a MDP, and analyze the difficulties of VI algorithm in solving the optimal strategy. The transmission scheduling problem is analysed under the framework of an RMAB by introducing the Lagrange multiplication factor. Then, the indexability of RMAB model is proved and the Whittle index based transmission scheduling strategy algorithm is developed. At last, the

TABLE IV

WHITTLE INDEX BASED COMMUNICATION SCHEDULING STRATEGY OF REAL-TIME P2P REAL-TIME COMMUNICATION SYSTEM

N=3 and $Z=2$		N=4 and $Z=3$		N=5 and $Z=3$		...
system state	scheduling strategy	system state	scheduling strategy	system state	scheduling strategy	...
(1,0,0,0,0,0,0)	(0,0,0)	(1,0,0,0,0,0,0,0,0)	(0,0,0,0)	(1,0,0,0,0,0,0,0,0,0,0)	(0,0,0,0,0)	...
...	...	...	...	...	...	...
(2,5,2,3,3,0,3)	(1,1,0)	(1,1,3,1,2,6,2,1,2)	(1,1,0,1)	(3,4,2,2,3,5,1,2,1,1,3)	(0,1,0,1,1)	...
(2,5,2,3,3,1,0)	(1,1,0)	(1,1,3,1,2,6,2,1,3)	(1,1,0,1)	(3,4,2,2,3,5,1,2,1,2,0)	(1,1,0,1,0)	...
(2,5,2,3,3,1,1)	(0,1,1)	(1,1,3,1,2,6,2,2,0)	(1,1,0,0)	(3,4,2,2,3,5,1,2,1,2,1)	(0,1,0,1,1)	...
(2,5,2,3,3,1,2)	(0,1,1)	(1,1,3,1,2,6,2,2,1)	(1,1,0,0)	(3,4,2,2,3,5,1,2,1,2,2)	(0,1,0,1,1)	...
(2,5,2,3,3,1,3)	(0,1,1)	(1,1,3,1,2,6,2,2,2)	(1,1,0,0)	(3,4,2,2,3,5,1,2,1,2,3)	(0,1,0,1,1)	...
(2,5,2,3,3,2,0)	(1,1,0)	(1,1,3,1,2,6,2,2,3)	(1,1,0,0)	(3,4,2,2,3,5,1,2,1,3,0)	(1,1,0,1,0)	...
...	...	...	...	...	...	...

TABLE V

COMPUTATIONAL TIME OF THE ALGORITHM 1 AND ALGORITHM 2 IN DIFFERENT SCALES OF P2P COMMUNICATION SYSTEM.

Scale of P2P communication system	Number of states	Computational time	
		Algorithm 1	Algorithm 2
$N=1$, $Z=1$	84	0.4431	0.4194
$N=2$, $Z=1$	2352	3.3027	1.5268
$N=3$, $Z=2$	65856	\	1.8467
$N=4$, $Z=2$	1843968	\	54.0544
...	...	...	...

numerical results show that the transmission scheduling algorithm is with low time complexity and greatly reduces the computational cost in comparison with the VI algorithm.

However, sometimes the BS may only know part of the link states, so it will be one of our future work to consider the transmission scheduling of partially observable CSI. In addition,

with the development of wireless technology, it is a trend to combine P2P communication with MIMO (Multi-input Multi-output) or MU-MIMO (Multi-user Multi-input Multi-output) and NOMA (Non-Orthogonal Multiple Access) technologies, which makes the scheduling of P2P real-time communication more challenging.

APPENDIX A
PROOF OF THEOREM 1

For any $s_{i,t} = [L_{i,t}, B_{i,t}]$, let

$$J(s_{i,t}) \triangleq J_0(s_{i,t}) - J_1(s_{i,t}). \tag{8}$$

It is easy to see that, when $w = -\infty$, $J(s_{i,t}) = -\infty$, and when $w = +\infty$, $J(s_{i,t}) = +\infty$. Then the indexability can be proved if the monotonicity of $J(s_{i,t})$ holds. Thus, we only need to show that $\frac{\partial J(s_{i,t})}{\partial w} \geq 0$.

Before proving theorem 1, we first give two lemmas to describe the existence and the characteristics of the Whittle index in some special states.

Lemma A.1 Given the state $s_{i,t} = [L_{i,t}, B_{i,t}]$ of the i-th P2P pair, for any channel state $c_m \in \mathcal{C}$, the following statements hold.

(a) When $L_{i,t} = 0$, the Whittle index of state $s_{i,t}$ exists;

(b) When $L_{i,t} \neq 0, B_{i,t} = 0$, the Whittle index of state $s_{i,t}$ exists;

(c) When $L_{i,t} = 1$, the Whittle index of state $s_{i,t}$ exists.

Proof. (a) When $L_{i,t} = 0$, it must be $B_{i,t} = 0$, and then

$$V^w(0,0) = \max\{w + \beta U^w, -h(c_m) + \beta U^w\}, \tag{9}$$

where $U^w = \mathbb{E}[Q_i V^w(L_i, B_i) + \bar{Q}_i V^w(0,0)]$.

From (9), we obtain $J(0,0) = w + h(c_m)$, and then

$$w_i^*(0,0) = -h(c_m). \tag{10}$$

The statement (a) is proved.

(b) When $L_{i,t} \neq 0, B_{i,t} = 0$, let $L_{i,t} = k$, similar to (a), we get

$$\begin{aligned} V^w(k,0) = \max\{&w + \beta \mathbb{E}_{c_m}[V^w(k-1,0)], \\ &- h(c_m) + \beta \mathbb{E}_{c_m}[V^w(k-1,0)]\} \end{aligned}$$

then

$$w_i^*(k,0) = -h(c_m). \tag{11}$$

The statement (b) is proved.

(c) When $L_{i,t} = 1$, two cases should be considered:

(c.1) When $B_{i,t} = 0$, we have

$$V^w(1,0) = \max\left\{w + \beta U^w, -h(c_m) + \beta U^w\right\}.$$

Similar to (a), we can obtain

$$w_i^*(1,0) = -h(c_m). \tag{12}$$

(c.2) When $B_{i,t} \neq 0$, we have

$$\begin{aligned} J_0(1,B_{i,t}) =& w - f(B_{i,t}) + \beta U^w, \\ J_1(1,B_{i,t}) =& rp(c_m) - h(c_m) + \beta U^w - \bar{p}(c_m) f(B_{i,t}) \\ & - p(c_m) f(B_{i,t}-1), \end{aligned}$$

and then

$$\begin{aligned} J(1,B_{i,t}) =& w - rp(c_m) + h(c_m) - p(c_m) f(B_{i,t}) \\ & + p(c_m) f(B_{i,t}-1). \end{aligned}$$

Denote $\delta(B_{i,t}-1) \triangleq f(B_{i,t}) - f(B_{i,t}-1)$, and noting $J(1,B_{i,t}) = 0$, we then get

$$w_i^*(1,B_{i,t}) = rp(c_m) - h(c_m) - p(c_m)\delta(B_{i,t}-1). \tag{13}$$

The statement (c) is proved. ■

Lemma A.2 Given the state $s_{i,t} = [L_{i,t}, B_{i,t}]$ of the i-th P2P pair, for any channel state $c_m \in \mathcal{C}$, when $L_{i,t} \geq 2$, the Whittle index $w_i^*(s_{i,t})$ of state $s_{i,t}$ is existence. Define

$$\upsilon(s_{i,t}) = V^w(L_{i,t}, B_{i,t}+1) - V^w(L_{i,t}, B_{i,t}),$$

then the following inequality holds

$$\frac{\partial \mathbb{E}[\upsilon(L_{i,t}-1, B_{i,t}-1)]}{\partial w} \geq -\frac{1}{\beta p(c_m)}. \tag{14}$$

Proof. Two cases should be considered:

(I) The first case: $B_{i,t}=0$.

In this case, the lemma is obvious by Lemma A.1.

(II) The second case: $B_{i,t}\geq 1$.

In this case, we have

$$\begin{aligned}J_0(L_{i,t},B_{i,t}) =& w+\beta\mathbb{E}[V^w(L_{i,t}-1,B_{i,t})],\\ J_1(L_{i,t},B_{i,t}) =& rp(c_m)-h(c_m)\\ &+\beta p(c_m)\mathbb{E}[V^w(L_{i,t}-1,B_{i,t}-1)]\\ &+\beta\bar{p}(c_m)\mathbb{E}[V^w(L_{i,t}-1,B_{i,t})].\end{aligned}$$

Thus,

$$\begin{aligned}J(L_{i,t},B_{i,t}) =&\beta p(c_m)\mathbb{E}[\upsilon(T_{i,t}-1,B_{i,t}-1)]\\ &-rp(c_m)+h(c_m)+w.\end{aligned}$$

Taking the derivative of $J(L_{i,t},B_{i,t})$ with respect to w, we have

$$\frac{\partial J(L_{i,t},B_{i,t})}{\partial w}=\beta p(c_m)\frac{\partial\mathbb{E}[\upsilon(L_{i,t}-1,B_{i,t}-1)]}{\partial w}+1.$$

By the existence of the Whittle index, we have

$$1+\beta p(c_m)\frac{\partial\mathbb{E}[\upsilon(L_{i,t}-1,B_{i,t}-1)]}{\partial w}\geq 0.$$

Based on (I)and (II), when $L_{i,t}\geq 2$, we get

$$\frac{\partial\mathbb{E}[\upsilon(L_{i,t}-1,B_{i,t}-1)]}{\partial w}\geq -\frac{1}{\beta p(c_m)}.$$

The lemma is proved. ■

Proof of Theorem 1. The existence of Whittle index is proved by mathematical induction.

When $L_{i,t}=0$ and $L_{i,t}=1$, the conclusion holds by Lemma A.1.

We suppose that the conclusion holds when $L_{i,t}=k$, then from Lemma A.2, it holds

$$\frac{\partial\mathbb{E}[\upsilon(k,B_{i,t}-1)]}{\partial w}\geq -\frac{1}{\beta p(c_m)}. \tag{15}$$

Now, we prove the existence for $L_{i,t}=k+1$, and two cases are investigated:

Case 1: $B_{i,t}=0$. From $J(k+1,0)=w+h(c_m)$, we can get

$$w_i^*(k+1,0)=-h(c_m).$$

Case 2: $B_{i,t} \geq 1$. We have

$$\begin{aligned} J(k+1, B_{i,t}) =& p(c_m)\mathbb{E}[\upsilon(k, B_{i,t}-1)] - rp(c_m) \\ &+ h(c_m) + w. \end{aligned}$$

From (15), we can get

$$\frac{\partial J(k+1, B_{i,t})}{\partial w} = 1 + \beta p(c_m)\frac{\partial \mathbb{E}[\upsilon(k, B_{i,t}-1)]}{\partial w} \geq 0.$$

Thus, when $L_t = k+1$, the index $w^*(s_{i,t})$ exists. The theorem is proved. ■

APPENDIX B

PROOF OF THEOREM 2

Proof of Theorem 2. The first two expressions of equation (7) can be obtained from (10)-(13).

We now prove the third expression of equation (7). To do that, we first give the function $g(L_{i,t}, B_{i,t})$ under different cases.

When $B_{i,t} = 0$, from Lemma 1, we have $w_i^*(L_{i,t}, 1) \geq w_i^*(L_{i,t}, 0) = -h(c_m)$, and then

$$\upsilon(L_{i,t}, 0) = \begin{cases} \beta\bar{p}(c_m)\mathbb{E}[\upsilon(L_{i,t}-1, 0)] + rp(c_m) & \\ & w < -h(c_m), \\ \beta\bar{p}(c_m)\mathbb{E}[\upsilon(L_{i,t}-1, 0)] + rp(c_m) - h(c_m) & \\ -w & -h(c_m) \leq w < w_i^*(L_{i,t}, 1), \\ \beta\mathbb{E}[\upsilon(L_{i,t}-1, 0)] & w \geq w_i^*(L_{i,t}, 1). \end{cases} \tag{16}$$

Similarly, when $B_{i,t} \neq 0$, we have $w_i^*(L_{i,t}, B_{i,t}+1) \geq w_i^*(L_{i,t}, B_{i,t})$, and then

$$\begin{aligned} &\upsilon(L_{i,t}, B_{i,t}) \\ &= \begin{cases} \beta p(c_m)\mathbb{E}[\upsilon(L_{i,t}-1, B_{i,t}-1)] & \\ +\beta\bar{p}(c_m)\mathbb{E}[\upsilon(L_{i,t}-1, B_{i,t})] & w < w_i^*(L_{i,t}, B_{i,t}), \\ \beta\bar{p}(c_m)\mathbb{E}[\upsilon(L_{i,t}-1, B_{i,t})] + rp(c_m) - h(c_m) - w & \\ & w_i^*(L_{i,t}, B_{i,t}) \leq w < w_i^*(L_{i,t}, B_{i,t}+1), \\ \beta\mathbb{E}[\upsilon(L_{i,t}-1, B_{i,t})] & w \geq w_i^*(L_{i,t}, B_{i,t}+1). \end{cases} \end{aligned} \tag{17}$$

Thus, it can be obtained

$$\upsilon(1,0)=\begin{cases} rp(c_m)-\bar{p}(c_m)\delta(0) & w<-h(c_m), \\ rp(c_m)-h(c_m)-\bar{p}(c_m)\delta(0)-w & \\ \qquad -h(c_m)\leq w<w_i^*(1,1), & \\ -\delta(0) & w\geq w(1,1). \end{cases} \tag{18}$$

$$\upsilon(1,B_{i,t})=\begin{cases} -\bar{p}(c_m)\delta(B_{i,t})-p(c_m)\delta(B_{i,t}-1) & \\ \qquad w<w_i^*(1,B_{i,t}), & \\ rp(c_m)-h(c_m)-\bar{p}(c_m)\delta(B_{i,t})-w & \\ \qquad w_i^*(1,B_{i,t})\leq w<w_i^*(1,B_{i,t}+1), & \\ -\delta(B_{i,t}) & w\geq w_i^*(1,B_{i,t}+1). \end{cases} \tag{19}$$

We prove the third expression of equation (7) by the following two cases.

(I) When $L_{i,t}>1, B_{i,t}=1$, it holds

$$\begin{aligned} J(L_{i,t},1)=&\beta p(c_m)\mathbb{E}[\upsilon(L_{i,t}-1,0)]-rp(c_m) \\ &+h(c_m)+w. \end{aligned} \tag{20}$$

For $\forall c_n\in\mathcal{C}$, taking the second item of equation (16) and substituting it into equation (20), we can get

$$\begin{aligned} J(L_{i,t},1)=&w-rp(c_m)+h(c_m)+\beta p(c_m)\mathbb{E}[rp(c_n) \\ &-h(c_n)+\beta\bar{p}(c_n))\mathbb{E}[\upsilon(L_{i,t}-2,0)]-w]. \end{aligned}$$

Similarly, for $\forall c_n, c_{n'}, \cdots, c_{n^{L_{i,t}-2}} \in \mathcal{C}$, we have

$$\begin{aligned}
J(L_{i,t},1) =& w - rp(c_m) + h(c_m) + \beta p(c_m)\mathbb{E}[rp(c_n) \\
& - h(c_n) + \beta\bar{p}(c_n)\mathbb{E}[\upsilon(L_{i,t}-2,0)] - w] \\
=& w - rp(c_m) + h(c_m) + \beta p(c_m)\mathbb{E}[rp(c_n) \\
& - h(c_n) + \beta\bar{p}(c_n)\mathbb{E}[rp(c_n) - h(c_{n'}) \\
& + \beta\bar{p}(c_{n'})\mathbb{E}[\upsilon(L_{i,t}-3,0)] - w] - w] \\
=& \cdots \\
=& w - rp(c_m) + h(c_m) + \beta p(c_m)\mathbb{E}[rp(c_n) \\
& - h(c_n) + \beta\bar{p}(c_n)\mathbb{E}[rp(c_{n'}) - h(c_{n'}) + \cdots \\
& + \beta\bar{p}(c_{n^{L_{i,t}-3}})\mathbb{E}[\upsilon(1,0)] - w]\cdots - w] - w].
\end{aligned}$$

Substituting the second term of equation (18) and taking $J(L_{i,t}, B_{i,t}) = 0$, and Theorem 2 is proved.

(II) When $L_{i,t} > 1, B_{i,t} > 1$, it holds

$$\begin{aligned}
J(L_{i,t},B_{i,t}) = & \beta p(c_m)\mathbb{E}[g(L_{i,t}-1,B_{i,t}-1)] \\
& + w - rp(c_m) + h(c_m). \qquad (21)
\end{aligned}$$

Similar to the proof of (I), substituting the second term of equation (17) into equation (21) for $L_{i,t} - 1$ times, we obtain

$$\begin{aligned}
J(L_{i,t},B_{i,t}) =& w - rp(c_m) + h(c_m) + \beta p(c_m)\mathbb{E}[rp(c_n) \\
& - h(c_n) + \beta\bar{p}(c_n)\mathbb{E}[\upsilon(L_{i,t}-2,B_{i,t}-1)] - w] \\
=& w - rp(c_m) + h(c_m) + \beta p(c_m)\mathbb{E}[rp(c_n) \\
& - h(c_n) + \beta\bar{p}(c_n)\mathbb{E}[rp(c_{n'}) - h(c_{n'}) + \cdots \\
& + \beta\bar{p}(c_{n^{L_{i,t}-3}}))\mathbb{E}[\upsilon(1,B_{i.t}-1)] - w]\cdots - w].
\end{aligned}$$

Replacing $\upsilon(1, B_{i.t} - 1)$ with equation (19) and taking $J(L_{i,t}, B_{i,t}) = 0$. Thus, the proof is completed. ■

REFERENCES

[1] A. M. Odlyzko, "Internet traffic growth: Sources and implications," in *Optical transmission systems and equipment for WDM networking II*, vol. 5247, pp. 1–15, SPIE, 2003.

[2] J. Verdyck and M. Moonen, "Dynamic spectrum management in digital subscriber line networks with unequal error protection requirements," *IEEE Access*, vol. 5, pp. 18107–18120, 2017.

[3] J. Simarata and S. Suherman, "Downlink ratio impact on downstream traffic performances on wimax," in *IOP Conference Series: Materials Science and Engineering*, vol. 725, p. 012057, IOP Publishing, 2020.

[4] W. Coomans, H. Chow, and J. Maes, "Introducing full duplex in hybrid fiber coaxial networks," *IEEE Communications Standards Magazine*, vol. 2, no. 1, pp. 74–79, 2018.

[5] B. Di, L. Song, Y. Li, and H. V. Poor, "Ultra-dense leo: Integration of satellite access networks into 5g and beyond," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 62–69, 2019.

[6] R. Liu, Y. Li, M. Zhang, Z. Ding, S. Yang, and S. Zhu, "The wireless iot device identification based on channel state information fingerprinting," in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, vol. 9, pp. 534–541, IEEE, 2020.

[7] Y. Yildirim, "Optical solitons in dwdm technology with four-wave mixing by trial equation integration architecture," *Optik*, vol. 182, pp. 625–632, 2019.

[8] A. Tamayo-Dominguez, J.-M. Fernandez-Gonzalez, and M. S. Castaner, "Low-cost millimeter-wave antenna with simultaneous sum and difference patterns for 5g point-to-point communications," *IEEE Communications Magazine*, vol. 56, no. 7, pp. 28–34, 2018.

[9] J. Xu and C. Guo, "Scheduling stochastic real-time d2d communications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 6022–6036, 2019.

[10] N. Salodkar and A. Karnik, "Point-to-point scheduling over a wireless channel with costly channel state information," in *2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011)*, pp. 1–6, IEEE, 2011.

[11] M. H. Ngo and V. Krishnamurthy, "Optimality of threshold policies for transmission scheduling in correlated fading channels," *IEEE Transactions on Communications*, vol. 57, no. 8, pp. 2474–2483, 2009.

[12] G. Federico, D. Caratelli, G. Theis, and A. Smolders, "A review of antenna array technologies for point-to-point and point-to-multipoint wireless communications at millimeter-wave frequencies," *International Journal of Antennas and Propagation*, vol. 2021, 2021.

[13] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE communications magazine*, vol. 49, no. 6, pp. 101–107, 2011.

[14] F. Yilmaz and M.-S. Alouini, "A new simple model for composite fading channels: Second order statistics and channel capacity," in *2010 7th international symposium on wireless communication systems*, pp. 676–680, IEEE, 2010.

[15] K. Mahender, T. A. Kumar, and K. Ramesh, "Analysis of multipath channel fading techniques in wireless communication systems," in *AIP Conference Proceedings*, vol. 1952, p. 020050, AIP Publishing LLC, 2018.

[16] G. Sadeque, S. C. Mohonta, and F. Ali, "Modeling and characterization of different types of fading channel," *International Journal of Science, Engineering and Technology Research*, vol. 4, no. 5, pp. 1410–1415, 2015.

[17] H. S. Wang and N. Moayeri, "Finite-state markov channel-a useful model for radio communication channels," *IEEE transactions on vehicular technology*, vol. 44, no. 1, pp. 163–171, 1995.

[18] Q. Zhang and S. A. Kassam, "Finite-state markov model for rayleigh fading channels," *IEEE Transactions on communications*, vol. 47, no. 11, pp. 1688–1692, 1999.

[19] Y. Guan and L. Turner, "Generalised fsmc model for radio channels with correlated fading," *IEE Proceedings-Communications*, vol. 146, no. 2, pp. 133–137, 1999.

[20] J. C. Gittins, "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 41, no. 2, pp. 148–164, 1979.