Delay-Aware Dynamic Resource Management for High-Speed Railway Wireless Communications

Shengfeng Xu, Gang Zhu, Chao Shen, Shichao Li, Zhangdui Zhong
State Key Laboratory of Rail Traffic Control and Safety
Beijing Jiaotong University, Beijing, P. R. China
Email:{11120177, gzhu, shenchao, 14111022, zhdzhong}@bjtu.edu.cn

Abstract—In this paper, we investigate the delay-aware dynamic resource management problem for multi-service transmission in high-speed railway wireless communications, with a focus on resource allocation among the services and power control along the time. By taking account of average delay requirements and power constraints, the considered problem is formulated into a stochastic optimization problem, rather than pursuing the traditional convex optimization means. Inspired by Lyapunov optimization theory, the intractable stochastic optimization problem is transformed into a tractable deterministic optimization problem, which is a mixed-integer resource management problem. By exploiting the specific problem structure, the mixed-integer resource management problem is equivalently transformed into a single variable problem, which can be effectively solved by the golden section search method with guaranteed global optimality. Finally, we propose a dynamic resource management algorithm to solve the original stochastic optimization problem. Simulation results show the advantage of the proposed dynamic algorithm and reveal that there exists a fundamental tradeoff between delay requirements and power consumption.

I. INTRODUCTION

For the last decade, high-speed railway (HSR) has become the future trend of railway transportation worldwide, and attracted a lot of attentions as a fast, convenient and green public transportation system. With the continuous construction of HSR in recent years, the demand for mobile communication on high-speed trains is increasingly growing [1]. More and more services related with the railway controlling information need to be transmitted between the train and the ground in order to guarantee the train moving safety. Meanwhile, the passengers have an increasingly high demand on wireless Internet services when they are onboard. To fulfill the high demand for wireless data transmission, the study on efficiency of HSR communications is critical.

There have been some recent works to improve the transmission performance in HSR communication systems. From the network architecture perspective, a relay-assisted HSR network architecture has been proposed in [2] and [3], which can provide better performance than direct transmission in case of large penetration loss. To better utilize the network resources, [4] and [5] considered control/data signaling decoupled and cellular/infostation integrated HSR network architectures, respectively. From the transmission technology perspective, the radio-over-fiber (RoF) technology for HSR communications was proposed in [6], which can improve handover performance effectively. Multi-input Multi-output technology (MIMO) was

introduce into HSR scenarios in order to increase the network throughput [7]. However, when considering the multiple services transmission between the train and the ground, more investigations on resource management are necessary to further improve the transmission performance.

In HSR communications, many types of services need to be transmitted between the train and the ground [8]. In particular, these HSR services are classified into four categories [9], i.e., pure passenger internet, passenger comfort services, security-related services and cost saving applications. The effective transmission for these heterogenous services is a technical challenge. First, the channel condition cannot remain at the same level due to the fast-varying distance between the train and the ground, which causes that the power control along the time has a large influence on transmission performance. Second, there exist heterogeneous quality-of-service (QoS) requirements in HSR communications, especially the end-to-end delay requirements since the security-related services should be delivered in time. The resource allocation plays a key role in enhancing the QoS performance by making full use of the limited resources. Based on the above two aspects, the power control along the time and resource allocation among the delay-aware services in HSR wireless communications are still interesting and challenging problems.

To the best of our knowledge, resource allocation and power control in HSR communications are usually considered as separate problems. In this paper, we jointly optimize them for delay-aware multi-service transmission in HSR communications. Specifically, the main contributions are summarized as follows.

- A stochastic optimization framework for multi-service transmission in HSR communication systems is developed, which focuses on dynamic resource management under the heterogeneous delay requirements and power constraints. The proposed framework is based on a cross-layer design to improve the efficiency of resource management, which involves the interactions between physical (PHY) layer and media access control (MAC) layer.
- Inspired by the stochastic network optimization approach, the intractable stochastic optimization problem is transformed into a tractable deterministic optimization problem. A static resource management algorithm is proposed to solve it with guaranteed global optimality, by using

the golden section search method. Based on the static algorithm, we propose a dynamic resource management algorithm to solve the original stochastic optimization problem.

- The algorithm performance is evaluated by simulations under realistic assumptions for HSR communication systems. Simulation results show that compared with the traditional power control schemes, the proposed dynamic algorithm can effectively improve delay performance. In addition, we notice there exists a fundamental tradeoff between delay requirements and power consumption.

The remainder of the paper is structured as follows. In Section II, we review the related works. Section III describes the system model. The problem formulation and transformation are provided in Section IV. We propose a dynamic resource management algorithm in Section V. Numerical results and discussions are shown in Section VI. Finally, some conclusions are drawn in Section VII.

Notations: In this paper, $\mathbb{E}[\cdot]$ denotes expectation. $\lfloor x \rfloor = \max\{n \in \mathbb{Z} | n \leq x\}$. $\lceil x \rceil = \min\{n \in \mathbb{Z} | n \geq x\}$. $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{N}$ denote the sets of real numbers, all integers and all positive integers, respectively.

II. RELATED WORK

A. Resource Allocation

Resource allocation plays an important role in enhancing the data transmission efficiency and improving the QoS performance. In the literature, the resource allocation problem in HSR communications has attracted great research interest. For example, [10] and [11] investigated rate maximization resource allocation problem under the limited resource constraint. Different energy-efficiency resource allocation methods were developed in [12] and [13] to minimize the total transmit power while satisfy QoS requirements. However, a typical assumption in these works is the infinite backlog and the delay-insensitive services. As a result, these works focus only on optimizing the PHY layer performance metrics such as sum throughput and total transmit power, and the resultant resource allocation schemes are adaptive to the channel condition only.

In practical HSR communications, it is important to focus on cross-layer optimization design, which considers random bursty arrivals and delay performance metrics in addition to the PHY layer performance metrics. There is also plenty of literature on cross-layer resource optimization in HSR communications. [5] and [14] investigated the resource allocation problem for delivering multiple on-demand services while considering their deadline constraints. A cross-layer design approach was proposed in [15] to improve video transmission quality by jointly optimizing application-layer parameters and handoff decisions. In addition, [16] and [17] studied the downlink resource allocation problem with the delay constraint and packet delivery ratio requirement in relay-assisted HSR communications. All the above works treat the resource allocation problem with the assumption of a constant transmit power. When delivering multiple services between the ground and the train, the total resource allocated to the services is controlled by the instantaneous transmit power. Thus, it is necessary to jointly consider the resource allocation among the services and the power control along the time.

B. Power Control

There are three unique features in HSR communications [18], i.e., the deterministic moving direction, relatively steady moving speed and the accurate train location. The data transmission rate is highly determined by the transmit power and the distance between the ground and the train, thus these features make it necessary and feasible to implement power control along the time. Under the total power constraint, [19] presented four power allocation schemes to achieve different design objectives. As an extension, [20] investigated the utility-based resource allocation problem, which jointly considers the power allocation along the time and packet allocation among the services. The delay-aware power allocation policy has been proposed in [21] under the assumption of constant-rate data arrival. Moreover, [22] and [23] studied the energy-efficient data transmission problem in HSR communications, with the purpose of minimizing the total transmit power. However, these above works only take account of the time-varying channel state while do not consider the dynamic characteristics of the service or packet arrivals, which causes that the above power control schemes are not practical.

Dynamic power control is necessary to improve the performance of HSR communication systems, where the transmit power should be adaptive to the time-varying channel state and queue state. The work [24] investigated a joint admission control and resource allocation problem, which aims to maximize the system utility while stabilizing all transmission queues. Different from [24], our work in this paper focuses on the delay-aware dynamic resource allocation and power control problem under power constraints.

III. SYSTEM MODEL

As shown in Fig. 1, we consider a HSR communication network consisting of a linear cellular network and a backbone network. The linear cellular network deployed near the rail line can provide data transmission between the ground and the train. In the backbone network, the distributed content servers (CSs) are deployed in order to offload data traffic [25] and a central controller (CC) is responsible for resource management [5]. The base stations (BSs) in the cellular network are connected to the CSs via wireline links, thus BSs and CSs can communicate with a negligible delay. Considering the downlink transmission from the ground to the train, the data packets of the requested services are first delivered from the CSs to the vehicle station via the BSs, and then the vehicle station installed on the train transfers these data packets to the users on the train. Since the communication between the ground and the vehicle station suffers from the fast-varying wireless channel and may become the bottleneck in this network architecture. Therefore, this paper mainly considers

the multi-service downlink transmission from the BSs to the vehicle station.
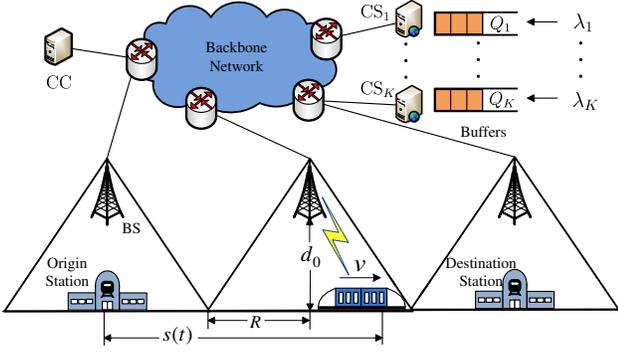
Fig. 1.   System model

A. Deterministic Train Trajectory

The deterministic train trajectory in HSR communication systems is a unique feature, which represents the train location at a specific time [5]. The train trajectory information can be obtained accurately due to the following two reasons. First, since the train moves on a predetermined rail line and the velocity is relatively steady. Second, many train positioning techniques are applied into railway communications, such as the global positioning system (GPS) and digital track maps [26].

In this paper, we develop a time-distance mapping function to describe the deterministic train trajectory. As shown in Fig. 1, we consider a train travels from the origin station to the destination station within the time duration $[0, T]$ at a constant speed v. The whole time duration is partitioned into slots of equal duration T_s. The distance which the train has traveled until slot t is $s(t) = vt$ and the train location between two adjacent BSs is $s(t) \bmod 2R$, where R is the cell radius. Define a time-distance mapping function $d(t) : [0, T] \rightarrow [d_0, d_{\max}]$, where $d_{\max} = \sqrt{R^2 + d_0^2}$ and d_0 is the distance between each BS and the rail line. With the help of the train trajectory information, the distance $d(t)$ at slot t can be obtained based on the geometry knowledge.

B. PHY Layer Model

The newly-built HSR routes are mainly composed of wide plain and viaduct, which yield a free space with few reflectors or scatterers. Most of the time, only the direct signal path between BS and vehicle station is available, which was confirmed by engineering measurements [27], [28]. Similar to [19], [21], [29], we assume that the channel condition variation results only from the time-varying distance between BS and train. Given the transmit power $P(t)$ and the distance $d(t)$, the received signal-to-noise ratio (SNR) at slot t can be denoted by

$$\text{SNR}(t) = \frac{P(t)d^{-\alpha}(t)}{BN_0} = \frac{P(t)}{N(t)}, \qquad (1)$$

where B is the system bandwidth, N_0 is the noise power spectral density, α is the pathloss exponent, and $N(t)$ is defined as $N(t) \triangleq BN_0 d^\alpha(t)$ for brevity.

Based on the SNR expression (1), the downlink transmission rate at slot t is expressed by

$$R(t) = B \log_2 \left(1 + \frac{P(t)}{N(t)} \right) \text{ bits/s.} \qquad (2)$$

Suppose that a packet is the transmission unit, hence the link capacity at slot t can be denoted as the maximum number of packets, which can be expressed by

$$C(t) = \left\lfloor \frac{T_s R(t)}{L} \right\rfloor = \left\lfloor \frac{1}{\eta} \log_2 \left(1 + \frac{P(t)}{N(t)} \right) \right\rfloor, \qquad (3)$$

where L is the packet size in bits, $\eta = \frac{L}{T_s B} > 0$, and T_s is a slot duration. Based on (3), we can see that the link capacity $C(t)$ is determined by transmit power $P(t)$. Although the equal packet size and equal time duration at each slot are considered, the results presented herein can be extended to the unequal scenarios.

We consider the erasure coding based service transmission in PHY layer, which has been adopted in [5], [17] to simplify the protocol design for HSR communications. The advantage is that no recovery scheme is required for the transmission error or loss of specific packets due to highly dynamic wireless channels.

C. MAC Layer Model

A set $\mathcal{K} \triangleq \{1, \ldots, K\}$ of delay-constrained services are supported over the trip. We assume that CS_k is equipped with a buffer and can provide service k. Thus, we can see K delay-constrained queues in MAC layer, as shown in Fig. 1. The maximum size of each buffer $Q_{\max}$ is assumed to be sufficiently large. Let $\mathbf{Q}(t) = (Q_1(t), \ldots, Q_K(t))$ represent the current queue backlogs vector, where $Q_k(t)$ denotes the number of packets at the beginning of slot t in the queue of CS_k. Let $\mathbf{A}(t) = (A_1(t), \ldots, A_K(t))$ represent the packet arrival vector, where $A_k(t)$ denotes the number of packets arriving into the buffer of CS_k at slot t. Suppose in general, $A_k(t)$ follows Poisson distribution with average packet arrival rate λ_k for service k.

The MAC layer is responsible for the resource allocation among the services. Let $\boldsymbol{\mu}(t) = (\mu_1(t), \ldots, \mu_K(t))$ be the resource allocation action vector at slot t, where $\mu_k(t)$ is the number of packets allocated to service k. Since the total number of allocated packets can not exceed the link capacity, the resource allocation at each slot t must satisfy the constraint $0 \leq \sum_k \mu_k(t) \leq C(t)$. In addition, the dynamics for all queues are given by

$$Q_k(t+1) = Q_k(t) - \mu_k(t) + A_k(t), \ \forall k \in \mathcal{K}. \qquad (4)$$

Since the arrival packets at slot t can only be transmitted after slot t, we have $0 \leq \mu_k(t) \leq Q_k(t), \ \forall k \in \mathcal{K}$.

IV. PROBLEM FORMULATION AND TRANSFORMATION

In this section, we first present a detailed description of the delay-aware multi-service transmission from the perspective of cross-layer design. The average delay constraint and average power constraint are formulated in terms of long-term time average. Then auxiliary variables are introduced to transform these long-term average constraints into the queue stability constraints. Based on the Lyapunov drift theory, we formulate the delay-aware resource allocation and power control problem as a stochastic optimization problem. Finally, the intractable stochastic optimization problem is transformed into a tractable deterministic optimization problem.

A. Problem Statement

This paper considers the delay-aware multi-service transmission in HSR communication systems, with a focus on dynamic resource allocation and power control problem. Based on the model in Section III, the problem can be stated as follows: During a trip, considering system dynamic characteristics, i.e., the random packet arrivals and time-varying wireless channels, how to dynamically optimize resource allocation and power control to satisfy the heterogenous delay requirements of multiple services under power constraints along the time.

To enhance the efficiency of resource utilization and improve delay performance of service transmission, it is necessary to dynamically control resource in a cross-layer way. Fig. 2 presents an illustration of cross-layer resource management, which involves the interactions between the PHY layer and the MAC layer. At the PHY layer, the channel state information (CSI) allows an observation of good transmission opportunity. At the MAC layer, the queue state information (QSI) provides the urgency of data packets. The control actions, including power control action P and resource allocation action vector $\boldsymbol{\mu}$, should be taken dynamically based on the PHY layer CSI and the MAC layer QSI. Specifically, the power control action decides the link capacity, i.e., the total allocated packets. The resource allocation action decides how many packets are allocated for each service.
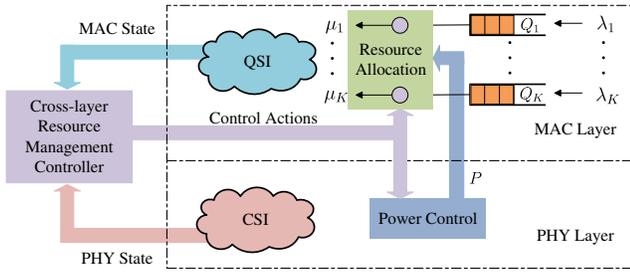
Fig. 2.   Cross-layer design for dynamic resource management

B. Constraint Formulation

We define the following notation for the long-term time average expectation of any quantity x,

$$\overline{x} \triangleq \lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}[x(\tau)], \tag{5}$$

Based on the definition (5), $\overline{Q}_k$ and $\overline{P}$ are denoted as average queue backlog for queue k and average power consumption, respectively.

For the erasure coding based service delivery, the average delay constraint is considered since the decoding delay will be closely related to all encoded packets. Mathematically, the average delay constraint for queue k can be expressed by

$$\overline{W}_k \le W_k^{\mathrm{av}}, \tag{6}$$

where $\overline{W}_k$ and W_k^{av} represent the average delay and maximum average delay for queue k, respectively. Based on Little's law, the average delay can be obtained by $\overline{W}_k = \frac{\overline{Q}_k}{\lambda_k}$. Thus, the constraint (6) is equal to

$$\overline{Q}_k \le W_k^{\mathrm{av}} \lambda_k. \tag{7}$$

The data transmission between the train and ground is subject to the transmit power constraints, including the maximum power constraint and average power constraint. Mathematically, the maximum power constraint at any slot t is given by $P(t) \le P_{\max}$ and the average power constraint can be expressed by

$$\overline{P} \le P_{\mathrm{av}}, \tag{8}$$

where P_{av} and $P_{\max}$ denote the maximum average power and the maximum instantaneous power, respectively.

The objective of this paper is to investigate the dynamic resource allocation and power control problem under the average delay constraints (6), the average power constraint (8) and the maximum power constraint. In order to better characterize the considered problem, we consider the constraint transformation in the following subsection.

C. Constraint Transformation

To facilitate satisfaction of the constraint (7), we define a virtual queue $X_k(t)$ for each k with the update equation

$$X_k(t+1) = \max[X_k(t) - W_k^{\mathrm{av}} \lambda_k, 0] + Q_k(t+1) \tag{9}$$

where $Q_k(t+1)$ is defined in (4) and the initial condition is assumed $X_k(0) = 0$ for all k. Intuitively, $Q_k(t+1)$ and $W_k^{\mathrm{av}} \lambda_k$ can be viewed as the "arrivals" and the "offered service" of queue $X_k(t)$, respectively.

Lemma 1: If the virtual queue $X_k(t)$ is rate stable, i.e., satisfies $\lim_{t \to \infty} \frac{X_k(t)}{t} = 0$, then $\overline{W}_k \le W_k^{\mathrm{av}}$ holds and the queue $Q_k(t)$ is stable.

Proof: From (9), we have $X_k(\tau+1) \ge X_k(\tau) - W_k^{\mathrm{av}} \lambda_k + Q_k(\tau+1)$, i.e., $X_k(\tau+1) - X_k(\tau) \ge Q_k(\tau+1) - W_k^{\mathrm{av}} \lambda_k$ for any τ. Summing the above over $\tau \in \{0, \ldots, t-1\}$ yields $X_k(t) - X_k(0) \ge \sum_{\tau=0}^{t-1}(Q_k(\tau+1) - W_k^{\mathrm{av}} \lambda_k)$. Dividing by t and taking limit as $t \to \infty$, we can get $\lim_{t \to \infty} X_k(t)/t \ge \overline{Q}_k - W_k^{\mathrm{av}} \lambda_k$. Thus, if $\lim_{t \to \infty} X_k(t)/t = 0$, then $\overline{Q}_k \le W_k^{\mathrm{av}} \lambda_k$ holds, which implies the queue $Q_k(t)$ is stable. In addition, based on Little's law, we have $\overline{Q}_k = \lambda_k \overline{W}_k$ and $\overline{W}_k \le W_k^{\mathrm{av}}$ for $k \in \mathcal{K}$. ∎

The intuition behind Lemma 1 is that if the excess backlog in the virtual queue is stabilized, it must be the case that the time average arrival rate $\overline{Q}_k$ is not larger than the service rate

$W_k^{\mathrm{av}} \lambda_k$. Based on the Lemma 1, the constraint (6) can be transformed into a single queue stability problem.

Similarly, for the constraint (8), we define the virtual queue $Y_k(t)$ for each k with the update equation

$$Y_k(t+1) = \max[Y_k(t) - P_{\mathrm{av}}, 0] + P(t). \tag{10}$$

Thus, stabilizing $Y_k(t)$ ensures $\overline{P} \leq P_{\mathrm{av}}$.

D. Problem Formulation

Define $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ as a vector of all virtual queues $X_k(t)$ and $Y_k(t)$, respectively. We denote $\mathbf{\Theta}(t)$ as the combined vector of all virtual queues, $\mathbf{\Theta}(t) \triangleq [\mathbf{X}(t), \mathbf{Y}(t)]$. Define the quadratic Lyapunov function [30]

$$L(\mathbf{\Theta}(t)) \triangleq \frac{1}{2} \left(\sum_{k \in \mathcal{K}} X_k(t)^2 + \omega \sum_{k \in \mathcal{K}} Y_k(t)^2 \right), \tag{11}$$

where $\omega \geq 0$ represents the weight on how much we emphasize the average power constraint.

Next, $\Delta(\mathbf{\Theta}(t))$ is defined as the one-slot conditional Lyapunov drift at slot t

$$\Delta(\mathbf{\Theta}(t)) = \mathbb{E}[L(\mathbf{\Theta}(t+1)) - L(\mathbf{\Theta}(t))|\mathbf{\Theta}(t)], \tag{12}$$

which can help to ensure that the virtual queues are stable and the desired constraints are met. At each slot t, observing the virtual queue vector $\mathbf{\Theta}(t)$ and real queue vector $\mathbf{Q}(t)$, the resource allocation action vector $\boldsymbol{\mu}(t)$ and power control action $P(t)$ should be jointly decided to minimize the drift (12). Thus, the resource management problem at slot t is formulated as

$$\min_{P(t), \boldsymbol{\mu}(t)} \quad \Delta(\mathbf{\Theta}(t)) \tag{13a}$$

$$\text{s.t.} \quad 0 \leq P(t) \leq P_{\max} \tag{13b}$$

$$0 \leq \mu_k(t) \leq Q_k(t), \ \mu_k(t) \in \mathbb{N}, \ \forall k \tag{13c}$$

$$\sum_{k \in \mathcal{K}} \mu_k(t) \leq C(t) \tag{13d}$$

The problem (13) is a stochastic optimization problem [30], but it cannot be solved efficiently since the difficulty from the form of the objective function (13a). In order to better characterize the problem (13) and develop an efficient algorithm to solve it, we consider the problem transformation in the following subsection.

E. Problem Transformation

To make the objective function (13a) easily handled, we have the following lemma.

Lemma 2: Under any $\boldsymbol{\mu}(t)$, $P(t)$ and $\mathbf{\Theta}(t)$, we have

$$\Delta(\mathbf{\Theta}(t)) \leq \frac{1}{2} D + \mathbb{E}[G(t)|\mathbf{\Theta}(t)], \tag{14}$$

where D is a finite constant defined as

$$D = \sum_{k \in \mathcal{K}} \left[Q_{\max}^2 + (\lambda_k W_k^{\mathrm{av}})^2 + \omega(P_{\max}^2 + P_{\mathrm{av}}^2) \right], \tag{15}$$

and $G(t)$ is defined as

$$G(t) \triangleq \sum_{k \in \mathcal{K}} \Big[X_k(t)(Q_k(t) - \mu_k(t) + A_k(t) - \lambda_k W_k^{\mathrm{av}})$$
$$+ \omega Y_k(t)(P(t) - P_{\mathrm{av}}) \Big]. \tag{16}$$

Proof: By squaring the equation (9), we have

$$X_k(t+1)^2 - X_k(t)^2$$
$$= (\max[X_k(t) - \lambda_k W_k^{\mathrm{av}}, 0] + Q_k(t+1))^2 - X_k(t)^2$$
$$\leq Q_k(t+1)^2 + (\lambda_k W_k^{\mathrm{av}})^2 + 2X_k(t)(Q_k(t+1) - \lambda_k W_k^{\mathrm{av}}) \tag{17}$$

where we use the fact that for any $x, y \geq 0$, $(\max[x, 0])^2 \leq x^2$ and $\max[x - y, 0] \leq x$ in the inequality.

Similarly, it can be shown for any $k \in \mathcal{K}$

$$Y_k(t+1)^2 - Y_k(t)^2 \leq P(t)^2 + P_{\mathrm{av}}^2 + 2Y_k(t)(P(t) - P_{\mathrm{av}}) \tag{18}$$

Based on (12), (17) and (18), we have

$$\Delta(\mathbf{\Theta}(t))$$
$$= \mathbb{E}\left[\frac{1}{2} \sum_{k \in \mathcal{K}} \left[X_k(t+1)^2 - X_k(t)^2 + \omega Y_k(t+1)^2 - \omega Y_k(t)^2 \right] \Big| \mathbf{\Theta}(t) \right]$$
$$\leq \mathbb{E}\left[\frac{1}{2} \sum_{k \in \mathcal{K}} \left[Q_k(t+1)^2 + (\lambda_k W_k^{\mathrm{av}})^2 + \omega(P(t)^2 + P_{\mathrm{av}}^2) \right.\right.$$
$$+ 2X_k(t)(Q_k(t) - \mu_k(t) + A_k(t) - \lambda_k W_k^{\mathrm{av}})$$
$$\left.\left. + 2\omega Y_k(t)(P(t) - P_{\mathrm{av}}) \right] \Big| \mathbf{\Theta}(t) \right]$$
$$\leq D + \mathbb{E}[G(t)|\mathbf{\Theta}(t)] \tag{19}$$

where $G(t)$ is defined by (16) and the last inequality can be obtained by

$$\mathbb{E}\left[\sum_{k \in \mathcal{K}} \left[Q_k(t+1)^2 + (\lambda_k W_k^{\mathrm{av}})^2 + \omega(P(t)^2 + P_{\mathrm{av}}^2) \right] \Big| \mathbf{\Theta}(t) \right]$$
$$\leq \mathbb{E}\left[\sum_{k \in \mathcal{K}} \left[Q_{\max}^2 + (\lambda_k W_k^{\mathrm{av}})^2 + \omega(P_{\max}^2 + P_{\mathrm{av}}^2) \right] \Big| \mathbf{\Theta}(t) \right]$$
$$= \sum_{k \in \mathcal{K}} \left[Q_{\max}^2 + (\lambda_k W_k^{\mathrm{av}})^2 + \omega(P_{\max}^2 + P_{\mathrm{av}}^2) \right] = D \tag{20}$$

where the inequality holds based on $Q_k(t+1) \leq Q_{\max}$ and $P(t) \leq P_{\max}$, and the equality holds since the constant in the square bracket is independent of $\mathbf{\Theta}(t)$. $\blacksquare$

Based on Lemma 2, the problem (13) can be simplified to minimize the drift upper bound, i.e., the right-hand-side of inequality (14). We notice that the control actions are independent of the first term and only affect the second term on the right-hand-side of the inequality (14). Thus, the objective turns to the minimization of the expression $\mathbb{E}[G(t)|\mathbf{\Theta}(t)]$. This conditional expectation is with respect to the virtual queue vector $\mathbf{\Theta}(t)$ and the possible control actions. Then, using the concept of opportunistically minimizing an expectation [30], the control actions are chosen to minimize $G(t)$ by observing

$\boldsymbol{\Theta}(t)$ and $\mathbf{Q}(t)$ at each slot t. Next, isolating $\boldsymbol{\mu}(t)$ and $P(t)$ in (16) leads to the following expression

$$\sum_{k \in \mathcal{K}} \left[\omega Y_k(t) P(t) - X_k(t) \mu_k(t) \right]. \qquad (21)$$

Therefore, the intractable stochastic optimization problem (13) can be transformed into a deterministic optimization problem at each slot, which is expressed by

$$\max_{P, \boldsymbol{\mu}} \quad \sum_{k \in \mathcal{K}} \left[X_k \mu_k - \omega Y_k P \right] \qquad (22a)$$

$$\text{s.t.} \quad 0 \le P \le P_{\max} \qquad (22b)$$

$$0 \le \mu_k \le Q_k, \ \mu_k \in \mathbb{N}, \ \forall k \qquad (22c)$$

$$\sum_{k \in \mathcal{K}} \mu_k \le C = \left\lfloor \frac{1}{\eta} \log_2 \left(1 + \frac{P}{N} \right) \right\rfloor \qquad (22d)$$

Note that the time index is omitted in problem (22) for brevity. Let $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_K^*)$ and P^* denote the optimal resource allocation action vector and the optimal power control action for problem (22), respectively.

V. Dynamic Resource Management Algorithm

The problem (22) is a mixed integer programming (MIP) problem. A common way of solving it is to relax the integer constraints and then the problem becomes a convex optimization problem, which can be solved by CVX [31]. In addition, the optimization solvers, such as CPLEX and LINDO, have been successfully applied to MIP problems. However, the above methods often have a high computational complexity. Thus, to overcome this disadvantage, we consider the problem transformation and then propose a static resource management algorithm to effectively solve problem (22). Finally, a dynamic resource management algorithm is proposed to solve the original problem (13).

A. Static Resource Management Algorithm

In this subsection, the problem (22) is equivalently transformed into a single variable problem, which will be discussed below. First, we focus on analyzing the constraints in problem (22). Specifically, we notice that the optimal solution will always achieve the equality in constraint (22d), which can be given by

$$\sum_{k \in \mathcal{K}} \mu_k = C = \frac{1}{\eta} \log_2 \left(1 + \frac{P}{N} \right). \qquad (23)$$

Otherwise we can reduce the value of P so as to increase the objective value without any violation of the constraints (22b) and (22c). Based on the first equality in (23) and the constraint (22c), the link capacity C should satisfy $0 \le C \le \sum_k Q_k$. In addition, since the link capacity C is the sum of integers, it is also an integer, i.e., $C \in \mathbb{N}$. From the second equality in (23), there exists a one-to-one relationship between P and C. Thus, the power consumption P can be expressed by

$$P = N(2^{\eta C} - 1). \qquad (24)$$

Based on (24), the constraint (22b) is equivalent to $0 \le C \le C_{\max}$, where $C_{\max} \triangleq \frac{1}{\eta} \log_2 \left(1 + \frac{P_{\max}}{N} \right)$. From the above analysis, the link capacity C should satisfy

$$0 \le C \le \min \left(\sum_{k \in \mathcal{K}} Q_k, C_{\max} \right). \qquad (25)$$

Then, the problem (22) can be transformed into a single variable problem as shown below

$$\max_{C \in \mathbb{N}} \quad M(C) \triangleq M_1(C) - M_2(C) \qquad (26a)$$

$$\text{s.t. (25),} \qquad (26b)$$

where $M_1(C)$ is given by

$$M_1(C) \triangleq \max_{\boldsymbol{\mu}} \quad \sum_{k \in \mathcal{K}} X_k \mu_k \qquad (27a)$$

$$\text{s.t.} \quad 0 \le \mu_k \le Q_k, \ \mu_k \in \mathbb{N}, \ \forall k \in \mathcal{K}, \qquad (27b)$$

$$\sum_{k \in \mathcal{K}} \mu_k = C, \qquad (27c)$$

and $M_2(C)$ is given by

$$M_2(C) \triangleq \sum_{k \in \mathcal{K}} \omega Y_k P = \beta \left(2^{\eta C} - 1 \right), \qquad (28)$$

with $\beta \triangleq \omega N \sum_{k \in \mathcal{K}} Y_k$.

Lemma 3: Problems (22) and (26) are equivalent.

Proof: We prove the equivalence from both the objective function and the constraints. On one hand, we can observe that the objective functions of problem (22) and problem (26) are same, although the objective function of problem (26) is divided into two parts, i.e., (27a) and (28). On the other hand, as for the constraints on the variable μ_k, the constraint (22c) is the same as the constraint (27b) and the constraint (22d) is equivalent to (27c) based on the necessary condition of optimality (23). As for the constraints on the variable P, based on the expression (24), the constraint $0 \le P \le P_{\max}$ in (22b) is equivalent to $0 \le C \le C_{\max}$ in (25). From the above analysis, we can conclude that problem (22) is equivalent to problem (26). $\blacksquare$

Thus, we can solve the problem (26) instead of (22). Let C^* denote the optimal solution of the problem (26). Before solving the problem (26), we first focus on the subproblem (27) with any given C. It is worth noting that the maximum value of $M_1(C)$ can always be achieved by allocating link capacity C to the services in the descending order of X_k. For convenience, all the services are sorted in descending order of X_k with the set $\{k_1, k_2, \ldots, k_K\}$. Mathematically, the optimal solution to the subproblem (27) is given by

$$\mu_{k_n} = \min \left\{ \max \left\{ C - \sum_{m=0}^{n-1} Q_{k_m}, 0 \right\}, Q_{k_n} \right\}, \ \forall n, \quad (29)$$

where $Q_{k_0} = 0$.

After solving the subproblem (27), we focus on how to solve the problem (26). We relax $C \in \mathbb{N}$ to $C \in \mathbb{R}$ in problem (26), and then the property of the objective function $M(C)$ will be exploited in the following lemma.

Lemma 4: $M(C)$ is concave over $[0, \sum_{m=0}^{K} Q_{k_m}]$.

Proof: On one hand, for a sufficiently small $\delta > 0$, since $\Delta M_1(C) = M_1(C + \delta) - M_1(C) = \delta X_{k_n}$ for $\sum_{m=0}^{n-1} Q_{k_m} \leq C < \sum_{m=0}^{n} Q_{k_m}, \forall n \in [1, K]$, $\Delta M_1(C)$ is a non-increasing function of C. On the other hand, since $\Delta M_2(C) = M_2(C + \delta) - M_2(C) = \beta 2^{\eta C}(2^{\eta \delta} - 1)$, $\Delta M_2(C)$ is a monotonically increasing function of C. Therefore, $\Delta M(C) = \Delta M_1(C) - \Delta M_2(C)$ is a monotonically decreasing function of C, which implies that $M(C)$ is concave over $[0, \sum_{m=0}^{K} Q_{k_m}]$. ∎

Based on [32], if $M(C)$ is concave, then $M(C)$ is unimodal. Since $M(C)$ is an unimodal function of C over $[0, \sum_{m=0}^{K} Q_{k_m}]$, the golden section search method [33] is very suitable for searching without derivative for the maximum of objective function $M(C)$ with unimodal. Then the static resource management algorithm is proposed based on the golden section search method, as described in Algorithm 1.

Algorithm 1 Static Resource Management Algorithm

Input: X_k, Q_k, β, the golden ratio $\varphi = \frac{\sqrt{5}-1}{2}$;
1: Initialize two endpoints, i.e., $\hat{C} = 0$ and $\check{C} = \min\{C_{\max}, \sum_{m=0}^{K} Q_{k_m}\}$;
2: Determine two intermediate points C_1 and C_2 such that $C_1 = \hat{C} + \varphi(\check{C} - \hat{C})$ and $C_2 = \check{C} - \varphi(\check{C} - \hat{C})$;
3: **while** $\check{C} - \hat{C} > \varepsilon$ **do**
4:    Obtain $M_1(C_1)$ and $M_1(C_2)$ by solving (27), respectively;
5:    Compute $M_2(C_1)$ and $M_2(C_2)$ based on (28);
6:    $M(C_1) = M_1(C_1) - M_2(C_1)$, $M(C_2) = M_1(C_2) - M_2(C_2)$;
7:    **if** $M(C_1) \geq M(C_2)$ **then**
8:       $\hat{C} := C_2$, $C_2 := C_1$, $C_1 := \hat{C} + \varphi(\check{C} - \hat{C})$;
9:    **else**
10:      $\check{C} := C_1$, $C_1 := C_2$, $C_2 := \check{C} + \varphi(\check{C} - \hat{C})$;
11:   **end if**
12: **end while**
13: $\tilde{C} := \frac{1}{2}(\check{C} + \hat{C})$;
14: Obtain C^* by solving (30);
15: Calculate P^* by (24) when $C = C^*$;
16: Obtain μ_k^* by (29) when $C = C^*$;
Output: P^*, $\boldsymbol{\mu}^*$

From step 1 to step 13, the golden section search method is used to get the optimal solution while relaxing C as a positive real number in problem (26). Specifically, two endpoints and two intermediate points in the search region are determined in step 1 and step 2, respectively. The iterative calculation of the golden section search is implemented from step 3 to step 12 until $\check{C} - \hat{C} \leq \varepsilon$, where ε is the iterative accuracy. In each iteration, $M(C_1)$ and $M(C_2)$ are calculated from step 4 to step 6 and then are evaluated from step 7 to step 11. The endpoints and intermediate points are updated during the evaluation. When the iteration converges, the optimal solution $\tilde{C}$ after relaxation is obtained in step 13. In step 14, the optimal solution of problem (26) is obtained by considering the integer nature of optimal variable C. Due to the concavity of $M(C)$

based on Lemma 4, $M(C)$ has a non-negative slope at $C = \lfloor \tilde{C} \rfloor$. Thus, $M(\lfloor \tilde{C} \rfloor) \geq M(\lfloor \tilde{C} \rfloor - \epsilon)$ for $0 \leq \epsilon \leq \tilde{C}$. Similarly, we have $M(\lceil \tilde{C} \rceil) \geq M(\lceil \tilde{C} \rceil + \epsilon)$ for $0 \leq \epsilon \leq \tilde{C}$ due to the non-positive slope of $M(C)$ at $C = \lceil \tilde{C} \rceil$. Thus, the optimal integer value of C is either $\lfloor \tilde{C} \rfloor$ or $\lceil \tilde{C} \rceil$. Mathematically, it can be obtained by

$$C^* = \underset{C \in \{\lfloor \tilde{C} \rfloor, \lceil \tilde{C} \rceil\}}{\operatorname{argmax}} M(C). \qquad (30)$$

Finally, the optimal solutions P^* and $\boldsymbol{\mu}^*$ can be obtained in step 15 and step 16, respectively.

B. Dynamic Resource Management Algorithm

In this subsection, we propose a dynamic resource management algorithm to solve the original problem (13) based on the static resource management algorithm. Specifically, by observing the queue states at each slot, the dynamic algorithm is designed to choose control actions via solving problem (26). The detailed steps are described in Algorithm 2. All system parameters should be initialized before the transmission process begins. At the beginning of each slot, the problem (26) is solved by calling Algorithm 1. At the end of each slot, the queues $Q_k(t+1)$, $X_k(t+1)$, and $Y_k(t+1)$ are updated according to (4), (9) and (10), respectively. The algorithm will be repeated until all service transmissions are finished.

Algorithm 2 Dynamic Resource Management Algorithm

1: Initialize T_s, B, N_0, ω, η, $Q_k(0) = X_k(0) = Y_k(0) = 0$ for all k;
2: Obtain the trajectory information $d(t)$;
3: **for** $t = 0$ to T **do**
4:    Calculate $\beta(t)$, $N(t)$, and $C_{\max}(t)$;
5:    Obtain $P(t)$ and $\boldsymbol{\mu}(t)$ by calling Algorithm 1;
6:    Update $Q_k(t+1)$, $X_k(t+1)$, and $Y_k(t+1)$ according to (4), (9), and (10), respectively;
7: **end for**

VI. SIMULATION RESULTS AND DISCUSSIONS

A. Simulation Setup

We consider a real train schedule based on the Huhang high-speed railway [5], [14]. The simulations in this paper are built on the train G7302 and the train trajectory is generated according to the mobility model proposed in [34]. The other parameters are summarized in Table I.

For the purpose of comparison, we evaluate two related static power allocation schemes as reference benchmarks, i.e., constant power allocation (CPA) scheme and water filling power allocation (WFPA) scheme [19]. In the static CPA scheme, BS maintains a constant transmit power at all times, i.e., $P(t) = P_{\mathrm{av}}$. In the static WFPA scheme, the water filling method is used to maximize the total throughput along the time. Since the power allocation in these two static schemes is determined in advance, we modify them to the corresponding dynamic schemes in order to adapt to the variations of data
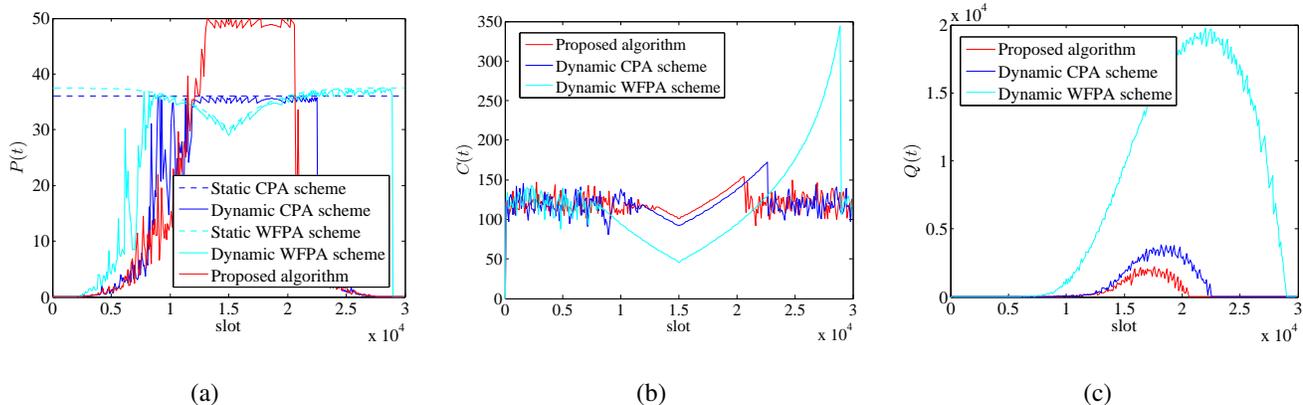
Fig. 3. The dynamic performance along the time, where $\lambda_k = 20$ packets/slot, $W_k^{\text{av}} = 15$ slots, $P_{\text{max}} = 50$ W and $\omega = 0.8$.

Parameter	Description	Value
P_{av}	maximum average power	36 W
B	system bandwidth	5 MHz
L	packet size	240 bits
T_s	slot duration	1 ms
α	pathloss exponent	4
N_0	noise power spectral density	-174 dBm/Hz
v	constant moving speed	360 km/h
R	cell radius	1.5 km
d_0	distance between BS and rail	50 m
K	number of services	6

traffic and channel state. Specifically, the dynamic schemes can be obtained by replacing the maximum power P_{max} in proposed dynamic algorithm with the static power allocation results. The resultant dynamic schemes are denoted as dynamic CPA scheme and dynamic WFPA scheme, respectively.

B. Performance Comparison

Fig. 3 shows the dynamic performance along the time for different power allocation schemes. For the sake of performance comparison, we only plot the simulation results during a time period when the train moves from the center of one cell to that of adjacent cell.

Fig. 3(a) shows the power allocation along the travel time for different schemes. We can see that the predetermined power allocation results in the static CPA scheme and static WFPA scheme are independent of random packet arrivals. Considering the packet arrival process, the transmit power changes dynamically in the three dynamic schemes. Specifically, when the train moves towards the cell edge, the power consumption increases in all dynamic schemes since the wireless link quality degrades. When the train moves at the cell edge (from 1×10^4th slot to 2×10^4th slot), nearly maximum transmit power is consumed in the proposed algorithm, while the power consumptions in the dynamic CPA scheme and dynamic WFPA scheme are limited by the predetermined power allocation.

Fig. 3(b) and 3(c) show the instantaneous link capacity and average queue backlog of services for the three dynamic schemes, respectively. It can be observed in Fig. 3(b) that when the train moves at the cell center (from 0 to 10^4th slot), the link capacity is just around the total packet arrivals 120 packets/slot for all schemes, which results in a small queue backlog shown in Fig. 3(c). This result can be explained as follows: Since the channel condition is good at the cell center, little power will be consumed shown in Fig. 3(a). Thus, the packets can be transmitted immediately once they arrive at the buffer, which implies that the queue backlog is small. As the train moves far from the cell center, the channel condition turns bad and much power will be consumed for transmitting one packet. The link capacity decreases for all schemes due to the power constraint, which causes the increasing queue backlog. As shown in Fig. 3(b), when the train locates near the cell edge, the link capacity in the proposed algorithm is more than that in the other two dynamic schemes due to the different power allocations in Fig. 3(a). The difference in the link capacity results in different queue backlogs among these dynamic schemes. Furthermore, we can see from Fig. 3(b) that the three curves suddenly drop after steady increasing, which implies the queue backlogs have been emptied. Compared with the other two schemes, less time will be spent on emptying queue backlogs and less buffer size is needed in the proposed algorithm.

Fig. 4 shows the average power consumption and average delay performance with different packet arrival rates, respectively. As expected, the average power consumption in all the schemes increases with the average packet arrival rates. Transmitting more packets with the same delay requirement will lead to more power consumption. However, the increment in the proposed algorithm is large while that in the other two schemes is small. This is because that the power consumptions in the dynamic CPA scheme and dynamic WFPA scheme are limited by the predetermined power allocation. From Fig. 4(b), we can see that the average delay in all the schemes also increases with the average packet arrival rates. When the packet arrival rate increases, the queue backlog gets larger,

which further results in longer queue delay. Furthermore, we observe that as for the same packet arrival rate, the average delay in the proposed algorithm is much lower than the other two schemes. Specifically, when the packet arrival rate is 25 packets/slot, the average delay in the proposed algorithm can respectively be 6.3% and 22.2% of that in the other two schemes, which demonstrates that the proposed algorithm outperforms the other two schemes in term of delay performance.
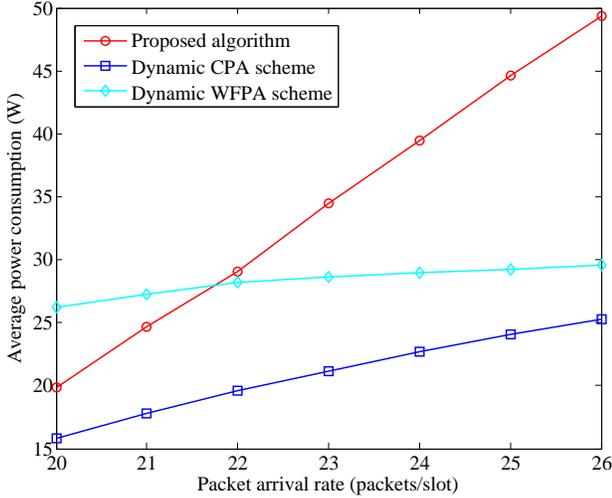
(a)
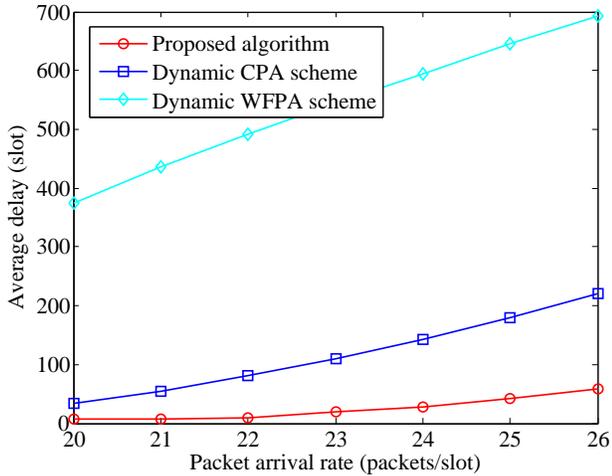
(b)

Fig. 4. Average power consumption and average delay under different packet arrival rates, where $\lambda_k = 20$ packets/slot, $W_k^{\mathrm{av}} = 15$ slots, $P_{\max} = 100\ W$ and $\omega = 0.8$.

C. Effects of the Parameter ω

We show the effects of the weight ω on the power consumption and delay performance in the proposed algorithm. Fig. 5 plots the average power consumption and average delay versus the weight ω. It can be seen that the average power consumption decreases with increasing ω while the average delay increases with increasing ω. The reason is that increasing
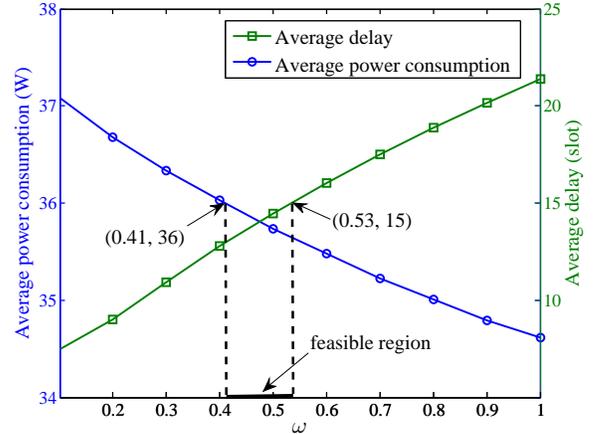
Fig. 5. Average power consumption and average delay under different weights, where $\lambda_k = 23$ packets/slot, $W_k^{\mathrm{av}} = 15$ slots, and $P_{\max} = 100\ W$.

ω leads to more weight putting on average power consumption constraint and hence less power is utilized for transmitting buffered packets, which results in longer queue delay. Thus, we can see that ω plays a key role in balancing the average delay and average power consumption. Furthermore, to satisfy the average delay constraint and average power constraint simultaneously, it is necessary to find the reasonable range of ω, which is called as "feasible region". As shown in Fig. 5, we can find the feasible region of ω is $[0.41, 0.53]$ such that both the average delay constraint and average power constraint can be satisfied simultaneously.

D. Effects of the Maximum Transmit Power

We evaluate how the maximum transmit power $P_{\max}$ effects the delay performance and power consumption in the proposed algorithm. Fig. 6(a) shows average power consumption and average delay with different maximum transmit powers, respectively. As shown in Fig. 6(a), the average power consumption increases with the maximum transmit power while the average delay decreases with the maximum transmit power. This can be explained by the observation in Fig. 6(b). As the maximum transmit power gets larger, more power is consumed when the train moves at the cell edge, resulting that the buffered packets can be transmitted as soon as possible.

VII. Conclusions

In this paper, we investigate the delay-aware dynamic resource allocation and power control problem in HSR wireless communications. The problem is formulated into a stochastic optimization problem, rather than pursuing the traditional convex optimization means. A dynamic resource management algorithm is proposed to solve the intractable stochastic optimization problem. The novelty of the proposed dynamic algorithm lies in the applications of stochastic network optimization approach and the ideas such as the virtual queue-based constraint transformation and opportunistically minimizing an expectation. Simulation results are presented to show that the proposed dynamic algorithm can reasonably
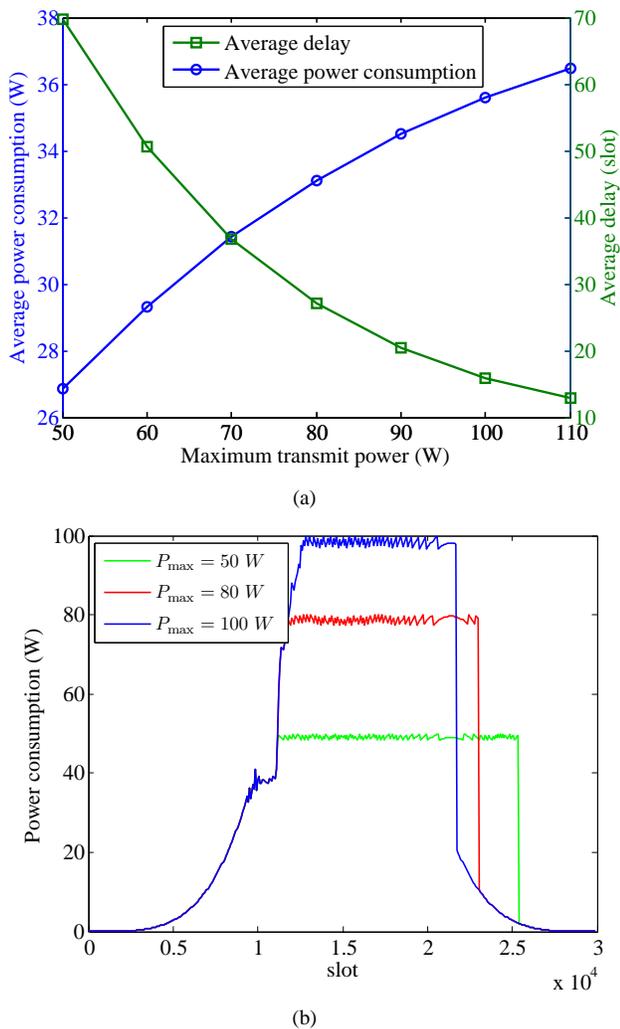
Fig. 6. Power consumption and delay performance under different maximum transmit powers, where $\lambda_k = 23$ packets/slot, $W_k^{\mathrm{av}} = 15$ slots and $\omega = 0.6$.

use the limited resource and significantly improve the delay performance under power constraints in HSR communications.

In the future, we plan to broaden and deepen this work in several directions. First, we attempt to investigate delay-aware cross-layer design for multi-service transmission with more practical assumptions in HSR communication systems, e.g., multi-service resource management problem under more practical HSR channel model and more realistic packet arrival model. Second, since the weight parameter plays a key role in balancing the queue delay and power consumption, we plan to analyze theoretically how to obtain the feasible region of the weight parameter. Third, to address the issue that the delay requirement and power constraint can not be satisfied simultaneously, we would like to design a dynamic admission control scheme for supporting multi-service transmission. Finally, as for the safety-related services transmission in HSR communications, we also plan to propose a new multi-service transmission scheme, which can take full consideration of different priorities and strict delay requirements.

REFERENCES

[1] B. Ai, X. Cheng, T. Kurner, Z. Zhong, K. Guan, R. He, L. Xiong, D. Matolak, D. Michelson, and C. Briso-Rodriguez, "Challenges toward wireless communications for high-speed railway," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2143–2158, 2014.

[2] J. Wang, H. Zhu, and N. J. Gomes, "Distributed antenna systems for mobile communications in high speed trains," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 675–683, 2012.

[3] L. Tian, J. Li, Y. Huang, J. Shi, and J. Zhou, "Seamless dual-link handover scheme in broadband wireless communication systems for high-speed rail," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 708–718, 2012.

[4] L. Yan, X. Fang, and Y. Fang, "Control and data signaling decoupled architecture for railway wireless networks," *IEEE Wireless Communications*, vol. 22, no. 1, pp. 103–111, February 2015.

[5] H. Liang and W. H. Zhuang, "Efficient on-demand data service delivery to high-speed trains in cellular/infostation integrated networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 780–791, May 2012.

[6] B. Lannoo, D. Colle, M. Pickavet, and P. Demeester, "Radio-over-fiber-based solution to provide broadband Internet access to train passengers," *IEEE Commun. Mag.*, vol. 45, no. 2, pp. 56–62, 2007.

[7] W. Luo, X. Fang, M. Cheng, and Y. Zhao, "Efficient multiple-group multiple-antenna (MGMA) scheme for high-speed railway viaducts," *IEEE Trans. Veh. Technol.*, vol. 62, no. 6, pp. 2558–2569, 2013.

[8] D. Pareit, E. V. de Velde, D. Naudts, and J. Bergs, "A novel network architecture for train-to-wayside communication with quality of service over heterogeneous wireless networks," *EURASIP J. Wirel. Comm.*, vol. 2012, no. 1, pp. 1–30, 2012.

[9] W. Brussel, "Bringing ICT services to trains technical and economical challenges," in *Proc. CTTE*, 2010, pp. 1–7.

[10] Y. Zhao, X. Li, Y. Li, and H. Ji, "Resource allocation for high-speed railway downlink MIMO-OFDM system using quantum-behaved particle swarm optimization," in *Proc. IEEE ICC*, 2013, pp. 936–940.

[11] O. B. Karimi, J. Liu, and C. Wang, "Seamless wireless connectivity for multimedia services in high speed trains," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 729–739, 2012.

[12] H. L. Zhu, "Radio resource allocation for OFDMA systems in high speed environments," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 748–759, 2012.

[13] Y. Zhao, X. Li, X. Zhang, and H. Ji, "Multidimensional resource allocation strategy for high-speed railway MIMO-OFDM system," in *Proc. IEEE GLOBECOM*, 2012, pp. 1653–1657.

[14] T. Y. Chen, H. G. Shan, and X. Wang, "Optimal scheduling for wireless on-demand data packet delivery to high-speed trains," *IEEE Trans. Veh. Technol.*, 2014.

[15] L. Zhu, F. R. Yu, B. Ning, and T. Tang, "Cross-layer design for video transmissions in metro passenger information systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1171–1181, 2011.

[16] S. F. Xu, G. Zhu, C. Shen, Y. Lei, and Z. Zhong, "Delay-aware online service scheduling in high-speed railway communication systems," *Math. Probl. Eng.*, vol. 2014, Article ID 248174, 10 pages, 2014.

[17] S. F. Xu, G. Zhu, C. Shen, and B. Ai, "A QoS-aware scheduling algorithm for high-speed railway communication system," in *Proc. IEEE ICC*, 2014, pp. 2855–2860.

[18] S. Lin, Z. Zhong, L. Cai, and Y. Luo, "Finite state markov modelling for high speed railway wireless communication channel," in *Proc. IEEE GLOBECOM*, 2012, pp. 5421–5426.

[19] Y. Q. Dong, P. Y. Fan, and K. B. Letaief, "High speed railway wireless communications: efficiency v.s. fairness," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 925–930, 2014.

[20] S. F. Xu, G. Zhu, C. Shen, Y. Lei, and Z. Zhong, "Utility-based resource allocation in high-speed railway wireless networks," *EURASIP J. Wirel. Comm.*, vol. 2014, no. 1, pp. 1–14, 2014.

[21] C. Zhang, P. Y. Fan, K. Xiong, and P. Z. Fan, "Optimal power allocation with delay constraint for signal transmission from a moving train to base stations in high-speed railway scenarios," *IEEE Trans. Veh. Technol.*, 2015.

[22] C. D. Thai, M. Berbineau, and F. Sun, "Exploiting the forecast channels for power allocation in high-speed railway communications," in *Proc. ITST*, 2013, pp. 402–406.

[23] X. Ma, J. Liu, and H. Jiang, "Energy-efficient mobile data uploading from high-speed trains," *Mobile Netw. Appl.*, vol. 17, no. 1, pp. 143–151, 2012.

[24] S. F. Xu, G. Zhu, C. Shen, Y. Lei, and Z. Zhong, "Analysis and optimization of resource control in high-speed railway wireless networks," *Math. Probl. Eng.*, vol. 2014, Article ID 781654, 13 pages, 2014.

[25] S. Spagna, M. Liebsch, and R. Baldessari, "Design principles of an operator-owned highly distributed content delivery network," *IEEE Commu. Mag.*, vol. 51, no. 4, pp. 132–140, 2013.

[26] R. D. Pascoe and T. N. Eichorn, "What is communication-based train control?" *IEEE Veh. Technol. Mag.*, vol. 4, no. 4, pp. 16–21, 2009.

[27] L. Liu, C. Tao, J. Qiu, H. Chen, Y. Li, W. Dong, and Y. Yuan, "Position-based modeling for wireless channel on high-speed railway under a viaduct at 2.35 GHz," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 834–845, 2012.

[28] R. He, Z. D. Zhong, B. Ai, and J. Ding, "An empirical path loss model and fading analysis for high-speed railway viaduct scenarios," *IEEE Antenn. Wirel. Pr.*, vol. 10, pp. 808–812, 2011.

[29] C. Zhang, P. Y. Fan, Y. Q. Dong, and K. Xiong, "Service-based high-speed railway base station arrangement," *Wireless Communications and Mobile Computing*, vol. 15, no. 13, pp. 1681–1694, 2015.

[30] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan & Claypool, 2010.

[31] M. Grant, S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, 2014.

[32] M. J. Bertin and R. Theodorescu, "Some characterizations of discrete unimodality," *Statistics & Probability Letters*, vol. 2, pp. 23–30, 1984.

[33] S. L. Jacoby, J. S. Kowalik, and J. T. Pizzo, *Iterative Methods for Nonlinear Optimization Problems*. Prentice-Hall, 1972.

[34] A. Ahmad, "Simulation-based local train mobility model," in *Proc. ACM SpringSim*, Apr. 2010, pp. 1–8.