Stochastic Delay Analysis for Train Control Services in Next-Generation High-Speed Railway Communications System

Lei Lei, *Member, IEEE*, Jiahua Lu, Yuming Jiang, *Senior Member, IEEE*, Xuemin (Sherman) Shen, *Fellow, IEEE*, Ying Li, Zhangdui Zhong, and Chuang Lin, *Senior Member, IEEE*

***Abstract*—The communication delay of train control services has a great impact on the track utilization and speed profile of high-speed trains. This paper undertakes stochastic delay analysis of train control services over a high-speed railway fading channel using stochastic network calculus. The mobility model of high-speed railway communications system is formulated as a semi-Markov process. Accordingly, the instantaneous data rate of the wireless channel is characterized by a semi-Markov modulated process, which takes into account the channel variations due to both large- and small-scale fading effects. The stochastic service curve of high-speed railway communications system is derived based on the semi-Markov modulated process. Based on the analytical approach of stochastic network calculus, the stochastic upper delay bounds of train control services are derived with both the moment generating function method and the complementary cumulative distribution function method. The analytical results of the two methods are compared and validated by simulation.**

***Index Terms*—LTE-R, stochastic network calculus, train control services.**

I. Introduction

RECENTLY, high-speed railway (HSR) has been developed rapidly all over the world, which puts forward requirements for a reliable and efficient wireless communication system between the moving train and the ground. According to International Union of Railways (UIC) E-Train Project [1], the train-ground wireless communication services for HSR system mainly include: (1) *train control services*, which are specific data and voice transmissions dedicated to the train crew with respect to the train control, train operator or other correspondents; (2) *train monitoring services*, which are data transmission in provenience from the train automatic monitoring and diagnosis systems; and (3) *passenger services* from/to Internet (all multimedia services accessible through Internet connection). The first and second categories are special services for train needs and provided by train operators mainly using Global System for Mobile Communications—Railway (GSM-R) [2], which is an international wireless communications standard for railway communication and applications. The third category is commercial services for passengers and currently provided by mobile network operators using cellular network standards, e.g., GSM/General Packet Radio Service (GPRS), Universal Mobile Telecommunications System (UMTS) and 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE).

Among the cellular network standards, LTE/LTE-Advanced represents the latest progress. It aims at providing a unified architecture to real-time and non real-time services and providing users with high data transfer rate, low latency and optimized packet wireless access technology. Although LTE is designed to support up to 350 km/h or even up to 500 km/h mobility speed, network performance is only optimized for $0 \sim 15$ km/h and high performance is possible only when the mobility speed is under 120 km/h [3]. This means that the quality of service (QoS) provided to the passengers on high-speed trains may be far from satisfactory. On the other hand, although GSM-R is specifically standardized for communication between train and railway regulation control centers, it is built on the GSM technology, which is a 2nd Generation (2G) cellular standard and much less efficient compared with the 4th Generation (4G) LTE standard. Therefore, it is important to design the next generation HSR communications system based on LTE technology while addressing the specific challenges of HSR environment, such as high mobility speeds (from 120 km/h for regional trains to 350 km/h for high-speed trains) and stringent QoS requirement of some railway-specific signalling, so that the above mentioned three types of communication services can be well supported by a unified network. Such a wireless communications system is commonly referred to as LTE-Railway (LTE-R).

Among the three categories of services for HSR communications system, the first category has higher priority over the other two categories, since the communication delay of the train control services between the train and track-side infrastructure is crucial for train movement control and safety [4]. Therefore, the LTE-R system needs to provide stringent QoS guarantee for

Manuscript received January 6, 2015; revised May 8, 2015; accepted June 24, 2015. This work was supported in part by the National Natural Science Foundation of China under Grants 61272168, U1334202, and 61472199; by the State Key Laboratory of Rail Traffic Control and Safety of Beijing Jiaotong University (RCS2014ZT10); and by the Key Grant Project of the Chinese Ministry of Education (313006). The Associate Editor for this paper was F. Qu.

L. Lei, J. Lu, Y. Li, and Z. Zhong are with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China (e-mail: leil@bjtu.edu.cn).

Y. Jiang is with the Department of Telematics, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway.

X. Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

C. Lin is with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2015.2450751

these mission-critical services. To this end, assigning dedicated radio resources to the first category of services is preferred to sharing radio resources with the other two categories of services, although higher resource efficiency can be achieved by the latter alternative due to statistical multiplexing gain. So, an interesting question is how many resources should be dedicated to these mission-critical services to guarantee their QoS performance or what is the expected QoS performance given a certain amount of dedicated resources for train control services transmission? In order to answer this question, we need to evaluate and quantify the QoS performance so that useful insights can be provided for LTE-R network dimensioning and design. Although the problem of cross-layer performance modeling and analysis of cellular networks and wireless ad-hoc networks has been addressed in literature [5], [6], the performance evaluation of LTE-R system is an open problem due to the following special features and requirements as compared to the LTE public communications system:

1) Traffic model: The characteristics of train control services are different from the user services in public communications system, which have to be studied and modeled for performance evaluation.
2) Wireless channel: The wireless channel characteristics for LTE-R system are unique due to the high mobility of the trains. The path loss varies rapidly as the train moves since it mainly depends on the distance between the train and the base station (referred to as evolved-NodeB (eNodeB) in LTE system). On the other hand, the time-correlation of the fading channel becomes very small with the increasing mobility speed. These effects together determine the instantaneous channel gains of the wireless channel.
3) Adaptive Modulation and Coding (AMC): Due to the high mobility of the trains and the induced rapid channel variations, it is very difficult to obtain accurate instantaneous Channel State Information (CSI) at the eNodeBs considering the channel measurements inaccuracy and CSI feedback delay. This will impact the performance of the AMC scheme in LTE system.

In this paper, we develop an analytical framework based on stochastic network calculus (snetal) taking into account the above unique characteristics to evaluate the performance of LTE-R system. The network calculus is a theory of queuing systems that has been developed as an initially deterministic framework for analysis of worst-case backlogs and delays, which are obtained by applying deterministic upper envelopes on traffic arrivals and lower envelopes on the offered service, the so-called arrival and service curves [7]. It is founded on the min-plus algebra and max-plus algebra to transform complex queuing systems into analytically tractable systems and mostly applied in the area of Internet QoS analysis. Compared with queuing theory which is largely constrained by the technical assumption of Poisson arrivals, network calculus can characterize a large variety of traffic arrival processes by their arrival curves. Although the worst-case performance bounds provided by deterministic network calculus (dnetal) were proven to be tight, the occurrence of such worst-case events is usually rare and statistical multiplexing gain can be captured when some violations of the deterministic bounds are tolerable. This has motivated considerable research for a stochastic network calculus which describes arrivals and service probabilistically while preserving the elegance and expressiveness of the original framework [8]–[12]. Generally speaking, existing work on snetal can be classified into two broad categories: the Moment Generating Function (MGF) approach [13] and the Complementary Cumulative Distribution Function (CCDF) approach [14]. Since it is easier to understand and simpler to implement, the MGF approach is more widely used in performance evaluation of wireless networks [15]–[18]. The research on CCDF approach for wireless channel focuses more on general principle and has mostly been applied for simple on-off impairment model [19], [20]. Notice that we use snetal instead of dnetal for the performance analysis of train control services mainly due to the following reasons: (1) although the delay performance of train control services is crucial to the safety of train operation, a small amount of violation probability can be tolerated according to the related standard [21]; and (2) much tighter bound can be derived by snetal compared with dnetal due to the stochastic nature of HSR fading channel and train control services. In addition, statistical multiplexing gain can be exploited for passenger services, which does not exist for the train control services studied in this paper.

This paper focuses on train control data traffic performance analysis of HSR fading channel. More specifically, we are interested in probabilistic delay and backlog guarantees of train control services in such a system. The impact of transmission delay to railway control system and the importance to provide delay guarantee to the train control services are discussed in [22], [23]. However, the above work assumes that the transmission delay is known as a fixed value and does not discuss on how to obtain the delay value. While increasing amount of literature on cross-layer modeling and optimization of HSR communications system has been proposed in recent years, most work only addresses the problem under the *infinite backlog traffic* model, assuming there will always be data to transmit from the queues [24]–[27]. Moreover, the models and optimization problems are *deterministic* considering a snapshot of the system instead of its dynamic behavior as a stochastic process over time. Different from the above work, [28], [29] consider dynamic optimization of radio resource management for HSR communication system. However, the transmission mechanisms considered are quite different from that of LTE-R. In order to analyze stochastic data traffic performance of HSR communications system, the HSR fading channel has to be modeled as a link between the physical layer and higher layers. Although HSR wireless channel modeling for physical layer has been a very active area, it is too complex to be incorporated into the cross-layer models for performance analysis and optimization. On the other hand, wireless channel can be modeled as a first-order Finite-State Markov Chain (FSMC) [30], which has been widely adopted in cross-layer performance analysis. However, most FSMC models in literature consider only low to medium mobility speed and assume that the average signal to noise ratio (SNR) remains constant [31], which is obviously not true for HSR

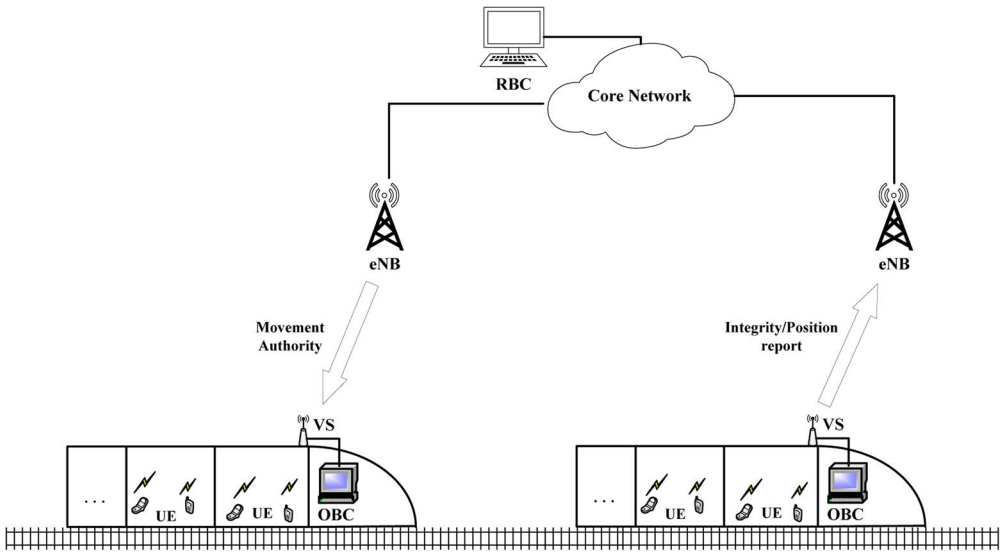

Fig. 1. LTE-R communications architecture for train control service.

fading channel. A FSMC model is developed for HSR fading channel in [32], which divides the coverage area of a base station along the railway line into multiple zones, assuming that the average SNR is constant within each zone and a FSMC similar to the traditional FSMC models are formulated for each zone. However, the FSMCs for different zones are considered separately, which cannot reflect the variation of average SNR over time as a train moves along the railway line. This "one FSMC per zone" modeling methodology is also used in other literature for HSR fading channel [33], [34], which are different from [32] in that real field measurement data is used to derive the SNR distribution.

The main contributions of this paper lie in the following aspects:

1) The mobility model of HSR communications system is formulated as a semi-Markov process. As such, the instantaneous data rate of wireless channel becomes a semi-Markov modulated process, which takes into account the channel variations due to both large-scale and small-scale fading effects. Moreover, the performance loss due to AMC selection with imperfect CSI is also considered. Finally, the stochastic service curve of HSR communications system is derived based on the semi-Markov modulated process.
2) Both CCDF snetal and MGF snetal approach are used to derive the delay and backlog bounds of train control services and the results are compared.
3) The analytical delay and backlog bounds are validated by simulation and can be used in the design and dimensioning of LTE-R system.

The remainder of this paper is organized as follows. Section II introduces system model including LTE-R architecture and the snetal basics. In Section III and Section IV, the stochastic arrival curve of train control services and the stochastic service curve for HSR fading channel are derived, respectively. In Section V, numerical and simulation results are presented, compared and discussed. Finally, Section VI concludes the paper.

II. System Model

A. LTE-R Architecture for Train Control System

Train control is an important part of the railway operating management system. Traditionally it connects the fixed signaling infrastructure with the trains. In modern train control systems, trains and control centers are connected by mobile communications links. Examples are European Train Control System (ETCS)/Chinese Train Control System (CTCS), which are used for main line railways in Europe/China; and Communications-Based Train Control (CBTC), which can mainly be used for urban railway lines. The current radio communication networks for ETCS/CTCS are based on GSM-R, which is envisioned to be upgraded to LTE-R in the future.

Fig. 1 depicts a simplified view of the LTE-R communication architecture. LTE-R eNodeBs are deployed along the railway line to provide a seamless coverage over the region. Although the LTE-R specifications have not been standardized yet, it is envisioned to be mostly based on the existing LTE specifications with some adaptations for the special characteristics of HSR communications, such as the high mobility and high priority of train control services. In this paper, the LTE-R eNodeBs can be considered as LTE eNodeBs, except that the proposed AMC scheme as described later is used in order to adapt to the HSR fading channel. The eNodeBs are connected to the core network via wireline links, while the core network provides connectivity to the train control centers. To overcome the penetration loss of train carriages, a vehicle station (VS) is fixed in the ceiling on top of the train. The data traffic dedicated to train control involves both downlink and uplink wireless transmissions between the VS and eNodeB. In modern train control systems, the train movement is controlled by exchanging messages with the control center, which

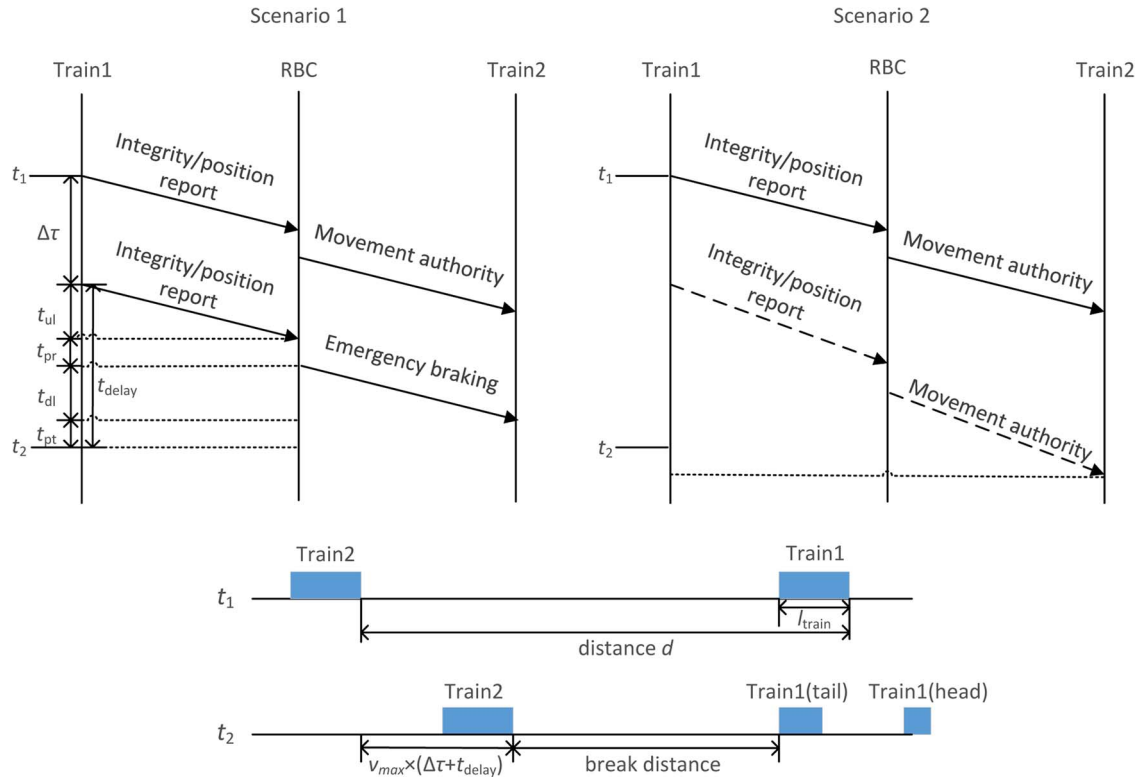

Fig. 2. The impact of communication delay to train distance and speed profile.

is referred to as radio block center (RBC) in the ETCS system. Each train features a train integrity control system and a computer (e.g., onboard controller (OBC) in ETCS) that can control train speed. It communicates via VS with eNodeBs, which are connected to the RBCs by the core network. Each train checks periodically its integrity and sends the integrity information together with the current position of the train head to the RBC, where such information is processed. The resulting information is sent to the following train, telling it either that everything is fine to go on driving (by sending a new movement authority message) or that an emergency braking is necessary immediately.

The communication delay between the VS and eNodeB of the train control services has great impact on the track utilization and speed profile of high-speed trains. The maximum track utilization will be achieved if trains are following each other with a minimum distance. Now we examine the minimum distance between trains operated under ETCS. We assume two trains (Train1 and Train2) directly follow each other with a maximum speed $v_{\max}$ and a distance d on a continuous track without stops, as shown in Fig. 2. At time t_1, Train1 completes its integrity check and sends a train integrity/position report to the RBC. Consider Scenario 1 where a part of Train1's carriages is lost immediately after t_1 from the main train and stop where they are. At time $t_1 + \Delta\tau$, an updated train integrity/position report is sent from Train1 to the RBC which informs the RBC that a part of its carriages is lost, where $\Delta\tau$ denotes the time between two successive integrity/position reports. After the RBC has processed this report, an emergency braking message is sent to the following Train2 which is processed there. As a result, Train2 starts to perform braking at a time no later than $t_2 = t_1 + \Delta\tau + t_{\text{delay}}$, where t_{delay} is the sum of the worst case values of the communication delay t_{ul} of the integrity/position report to the RBC, the processing time t_{pr} at the RBC, the communication delay t_{dl} of the emergency braking message to the Train2, and the processing time t_{pt} at Train2. The distance between the head of Train2 to the stopped part of Train1 is $d - l_{\text{train}} - v_{\max}(\Delta\tau + t_{\text{delay}})$ at time t_2, where l_{train} is the train length. Assume that the braking distance is l_{brake}. Then the minimum head-to-head distance d between the two trains should be $d = l_{\text{train}} + v_{\max}(\Delta\tau + t_{\text{delay}}) + l_{\text{brake}}$ to ensure train safety. Now consider Scenario 2 when Train1 is moving normally, but the communication delay of either the integrity/position report from Train1 to RBC or the movement authority message from RBC to Train2 exceeds the worst-case value, so consequently Train2 does not receive the second movement authority message in time. Without any information from RBC, Train2 needs to avoid accident if Scenario 1 of lost carriages as described above happens and brake at time t_2 even though it is actually safe to continue moving at the maximum speed. This means that if the communication delay exceeds the required worst-case value, the trains will perform unnecessary braking, which causes inefficiency in train operation and affects passenger comfort. Although increasing the required worst-case value of communication delay will solve this problem, the minimum distance d between trains will be increased and the track utilization decreased. Therefore, it is important to accurately evaluate the worst-case communication delay of the train control services to achieve the best tradeoff between track utilization and unnecessary braking.

B. Stochastic Network Calculus

This section provides a brief overview on the basic principle of stochastic network calculus and introduces the notation and basic assumptions in this paper. First, we introduce the following min-plus convolution and deconvolution operators, denoted by $\otimes$ and $\oslash$, respectively:

$$(f_1 \otimes f_2)(x) = \inf_{0<y\le x} \{f_1(y) + f_2(x-y)\}$$
$$(f_1 \oslash f_2)(x) = \sup_{y\ge 0} \{f_1(x+y) - f_2(y)\}.$$

We use $\mathcal{F}$ (resp. $\bar{\mathcal{F}}$) to denote the set of non-negative wide-sense increasing (resp. decreasing) functions as follows:

$$\mathcal{F} = \{f(\cdot) : \forall 0 \le x \le y, 0 \le f(x) \le f(y)\}$$
$$\bar{\mathcal{F}} = \{f(\cdot) : \forall 0 \le x \le y, 0 \le f(y) \le f(x)\}.$$

In this paper, the time model is discrete starting from zero. The time indices are denoted by the symbols n, k, t, and s. The stochastic processes are all considered as stationary. The cumulative arrivals and departures of a flow at/from a system up to time n are denoted by non-decreasing processes $A(n)$ and $A^*(n)$. The doubly-indexed extensions are $A(k,n) = A(n) - A(k)$ and $A^*(k,n) = A^*(n) - A^*(k)$. The delay of the flow at time n is

$$D(n) = \inf\{d \ge 0 : A(n) \le A^*(n+d)\} \tag{1}$$

and the backlog of the flow at time n is

$$B(n) = A(n) - A^*(n). \tag{2}$$

Let $S(n)$ denote the cumulative amount of workload that can be served by the system up to time n. The departure process $A^*(n)$ is determined by $A(n)$ and $S(n)$. Specifically, for a lossless queuing system, the following equality holds according to Lindley recursion:

$$B(n) = \sup_{0\le k\le n} \{A(k,n) - S(k,n)\}. \tag{3}$$

Combining (2) with (3), we have

$$A^*(n) = \inf_{0\le k\le n} \{A(k) + S(k,n)\} = A \otimes S(n). \tag{4}$$

Generally speaking, the snetal tackles the problem of performance analysis in two steps: (1) characterizing the stochastic arrival curve (SAC) for the flow arrival process $A(n)$ and stochastic service curve (SSC) for the system service process $S(n)$, respectively [35]; (2) deriving the stochastic delay and backlog bounds of the flow based on SAC and SSC. In order to achieve the above tasks, the MGF snetal and CCDF snetal take different approaches.

1) Step 1: Derivation of SAC and SSC: The SAC for $A(n)$ should be its upper envelop and the SSC for $S(n)$ should be its lower envelop, i.e., for all $0 \le k \le n$

$$A(k,n) - \hat{A}(n-k) \le 0 \tag{5}$$
$$A \otimes \hat{S}(n) - A^*(n) \le 0. \tag{6}$$

Since $A(n)$ and $S(n)$ are stochastic processes, it may be impossible to find deterministic processes $\hat{A}(n)$ and $\hat{S}(n)$ to satisfy the above inequalities as in dnetal, and will result in loose bounds even if such deterministic processes can be found. From this point on, the MGF snetal and CCDF snetal start to take different paths in dealing with this problem.

In MGF snetal, $\hat{A}(n)$ and $\hat{S}(n)$ are considered as stochastic processes referred to as stochastic envelop processes, which can be the arrival and service processes themselves. Then, the MGFs of stochastic envelop processes are derived, where the MGF for a stochastic process $X(t)$ is defined for any θ as

$$\mathrm{M}_X(\theta, n) = \mathrm{E}e^{\theta X(n)} \tag{7}$$

and E is the expectation of its argument. Note that another closely related concept is the effective bandwidth $\delta_X(\theta, n)$ of an arrival process $X(n)$, where

$$\delta_X(\theta, n) = \frac{1}{\theta n} \log \mathrm{E}e^{\theta X(n)} = \frac{1}{\theta n} \log \mathrm{M}_X(\theta, n). \tag{8}$$

In CCDF snetal, SAC $\hat{A}(n)$ and SSC $\hat{S}(n)$ are considered as deterministic processes. However, bounding functions $f(x)$ and $g(x)$ that bound the violation probabilities of (5) and (6) are defined in the CCDF form of $P(W > x) \le f(x)$ and $P(V > x) \le g(x)$ for all $x \ge 0$, where W and V can be the LHS term of (5) and (6), respectively. Alternatively, W and V can also be the maximum values of the LHS term of (5) and (6) over one or both of its free variable k and n. In [8], the three versions of SACs are thus defined as traffic-amount-centric (t.a.c.), virtual-backlog-centric (v.b.c.), and maximum (virtual)-backlog-centric (m.b.c.), while the two versions of SSC are defined as weak stochastic service curve and stochastic service curve. In this paper, we use the v.b.c. stochastic arrival curves and weak stochastic service curve, and their formal definitions are given below. For ease of understanding, we will use notations $\alpha(n)$ and $\beta(n)$ instead of $\hat{A}(n)$ and $\hat{S}(n)$ to represent SAC and SSC in CCDF snetal, respectively, where they are deterministic processes.

Definition 1: A flow $A(n)$ is said to have a v.b.c. stochastic arrival curve (SAC) $\alpha \in \mathcal{F}$ with bounding function $f \in \bar{\mathcal{F}}$, denoted by $A \sim_{vb} \langle f, \alpha \rangle$, if, for all $x \ge 0$ and $n \ge 0$

$$P\left\{\sup_{0\le k\le n} \{A(k,n) - \alpha(n-k)\} > x\right\} \le f(x). \tag{9}$$

Definition 2: A system S is said to provide a weak stochastic service curve $\beta \in \mathcal{F}$ with bounding function $g \in \bar{\mathcal{F}}$, denoted by $S \sim_{ws} \langle g, \beta \rangle$, if, for all $x \ge 0$ and $n \ge 0$

$$P\{A \otimes \beta(n) - A^*(n) > x\} \le g(x). \tag{10}$$

2) Step 2: Derivation of Backlog and Delay Bounds: The objective is to find the error functions $\varepsilon_b(x)$ and $\varepsilon_d(x)$ for backlog and delay, respectively, such that $P\{B(n) > x\} \le \varepsilon_b(x)$ and $P\{D(n) > x\} \le \varepsilon_d(x)$.

Specifically, the backlog satisfies

$$\begin{aligned} P\{B(n) > x\} &= P\{A(n) - A^*(n) > x\} \\ &\le P\left\{\sup_{0\le k\le n} \left\{\hat{A}(k,n) - \hat{S}(k,n)\right\} > x\right\} \end{aligned} \tag{11}$$

where the first equality follows (2), while the second inequality can be derived by combining (5) and (6) with (2).

Moreover, the definition of delay in (1) implies that for any $x \geq 0$, if $D(n) > x$, $A(0,n) > A^*(0, n+x)$ is true [14]. Therefore, we have

$$\begin{aligned} P\{D(n) > x\} &\leq P\{A(n) > A^*(n+x)\} \\ &\leq P\left\{\sup_{0\leq k\leq n}\left\{\hat{A}(k,n) - \hat{S}(k,n+x)\right\} > 0\right\} \end{aligned} \tag{12}$$

where the second inequality follows taking (5) and (6) into its LHS term.

In MGF snetal, the second inequalities of (11) and (12) are used to derive the stochastic backlog and delay bounds, i.e., $\varepsilon_b(x)$ and $\varepsilon_d(x)$, using Boole's inequality and Chernoff bound. The results are given in Theorem 1 [13], [16]. Note that the second inequalities of both (11) and (12) become equalities if the stochastic envelop processes $\hat{A}(n)$ and $\hat{S}(n)$ are the arrival and service processes $A(n)$ and $S(n)$ themselves [8], [13], [16].

Theorem 1: Given the stochastic arrival envelop process $\hat{A}(n)$ with MGF $\mathrm{M}_A(\theta, n)$ and stochastic service envelop process $\hat{S}(n)$ with MGF $\overline{\mathrm{M}}_S(\theta, n) = \mathrm{M}_S(-\theta, n)$. If $\hat{A}(n)$ is independent of $\hat{S}(n)$, then an upper backlog bound and an upper delay bound, each with at most violation probability $\varepsilon \in (0, 1]$, are given by

$$x_b(\varepsilon) = \inf_{\theta>0}\left[\frac{1}{\theta}\left(\ln\sum_{k=0}^{\infty}\mathrm{M}_A(\theta,k)\overline{\mathrm{M}}_S(\theta,k) - \ln\varepsilon\right)\right] \tag{13}$$

$$\begin{aligned} x_d(\varepsilon) = \inf_{\theta>0}\Bigg\{\inf\Bigg[\tau : \frac{1}{\theta}\Bigg(\ln\sum_{k=\tau}^{\infty}\mathrm{M}_A(\theta,k-\tau) \\ \times \overline{\mathrm{M}}_S(\theta,k) - \ln\varepsilon\Bigg) \leq 0\Bigg]\Bigg\}. \end{aligned} \tag{14}$$

Note that $x_b(\varepsilon)$ is the inverse function of $\varepsilon_b(x)$, i.e., $x_b(\varepsilon) = x$ if and only if $\varepsilon_b(x) = \varepsilon$. The detailed proof of Theorem 1 is given in Appendix A.

In CCDF snetal, the first equality of (11) and first inequality of (12) are used to calculate $\varepsilon_b(x)$ and $\varepsilon_d(x)$, respectively. For example, by adding and subtracting $A \otimes \beta(n)$ to $A(n) - A^*(n)$, we derive

$$\begin{aligned} &A(n) - A^*(n) \\ &\quad= \sup\{A(k,n) - \alpha(n-k) + \alpha(n-k) - \beta(n-k)\} \\ &\qquad + [A \otimes \beta(n) - A^*(n)] \\ &\quad\leq \sup_{0\leq k\leq n}\{A(k,n) - \alpha(n-k)\} + \sup_{0\leq k\leq n}\{\alpha(k) - \beta(k)\} \\ &\qquad + [A \otimes \beta(n) - A^*(n)] \\ &\quad\leq \sup_{0\leq k\leq n}\{A(k,n) - \alpha(n-k)\} + [A \otimes \beta(n) - A^*(n)] \\ &\qquad + \sup_{n\geq 0}\{\alpha(n) - \beta(n)\}. \end{aligned} \tag{15}$$

Since the CCDFs of random variables (r.v.s) $X = \sup_{0\leq n\leq k}\{A(k,n) - \alpha(n-k)\}$ and $Y = A \otimes \beta(n) - A^*(n)$ are bounded by the bounding functions $f(x)$ and $g(x)$ by the definitions of v.b.c. stochastic arrival curve and weak stochastic service curve, we have $P(B(n) > x) = P(A(n) - A^*(n) > x)$ is bounded by $\varepsilon_b(x) = f \otimes g(x + \inf_{k\geq 0}[\beta(k) - \alpha(k)])$ according to probability theory given in Lemma 2 of Appendix B. Similarly, the stochastic delay bound $\varepsilon_d(x)$ can also be derived for $P(D(n) > x) \leq P\{A(n) - A^*(n+x) > 0\}$. The results are summarized in the following theorem.

Theorem 2: Consider a system S with input A. If the input has a v.b.c. stochastic arrival curve $\alpha \in \mathcal{F}$ with bounding function $f \in \bar{\mathcal{F}}$, (i.e., $A \sim_{vb} \langle f, \alpha\rangle$), the server provides to the input a weak stochastic service curve $\beta \in \mathcal{F}$ with bounding function $g \in \bar{\mathcal{F}}$, i.e., ($S \sim_{ws} \langle g, \beta\rangle$), then the backlog $B(n)$ and delay $D(n)$ are guaranteed such that, for all $x \geq 0$ and $n \geq 0$

$$P\{B(n) > x\} \leq f \otimes g\left(x + \inf_{k\geq 0}[\beta(k) - \alpha(k)]\right) \tag{16}$$

$$P\{D(n) > x\} \leq f \otimes g\left(\inf_{k\geq 0}[\beta(k) - \alpha(k-x)]\right). \tag{17}$$

According to Lemma 2 of Appendix B, if X and Y are independent r.v.s, the backlog and delay bounds can be further improved. However, since both X and Y defined above depend on the arrival process $A(n)$, they are not independent even if the arrival and service processes are independent. In order to further improve the performance bounds, the concept of a stochastic strict server is introduced in [14] which characterizes the service process $S(n)$ by a deterministic ideal service process with strict service curve $\hat{\beta}(n)$ and an impairment process $I(n)$ according to the following definition.

Definition 3: A system $S(n)$ is said to be a stochastic strict server providing strict service curve $\hat{\beta}(n) \in \mathcal{F}$ with impairment process $I(n)$ if, during any backlog period $(k, n]$, the actual service $S(k,n)$ provided by the system satisfies

$$S(k,n) \geq \hat{\beta}(n-k) - I(k,n). \tag{18}$$

If $I(n)$ has a v.b.c. stochastic arrival curve $\xi(n)$ with bounding function $g(x)$, it can be proved that the service process satisfies $A \otimes \beta(n) - A^*(n) \leq \sup_{0\leq k\leq n}\{I(k,n) - \xi(n-k)\}$, where $\beta(n) = \hat{\beta}(n) - \xi(n)$ (see Appendix B). Since $P\{\sup_{0\leq k\leq n}\{I(k,n) - \xi(n-k)\} > x\} \leq g(x)$ by the definition of v.b.c. stochastic arrival curve and also because $I(n)$ is independent from $A(n)$, the improved backlog and delay bounds can be derived according to Lemma 2 of Appendix B, which are given in the following theorem [8] and proved in Appendix B.

Theorem 3: Consider a system S with input A. If the input has a v.b.c. stochastic arrival curve $\alpha \in \mathcal{F}$ with bounding function $f \in \bar{\mathcal{F}}$, (i.e., $A \sim_{vb} \langle f, \alpha\rangle$). Also suppose the server is a stochastic strict server providing strict service curve $\hat{\beta}$ with impairment process $I \sim_{vb} \langle g, \xi\rangle$. If A and I are independent, the backlog $B(n)$ and delay $D(n)$ are guaranteed such that, for all $x \geq 0$

$$P\{B(n) > x\} \leq 1 - \bar{f} * \bar{g}\left(x + \inf_{k\geq 0}[\beta(k) - \alpha(k)]\right) \tag{19}$$

$$P\{D(n) > x\} \leq 1 - \bar{f} * \bar{g}\left(\inf_{k\geq 0}[\beta(k) - \alpha(k-x)]\right) \tag{20}$$

where $\beta(n) = \hat{\beta}(n) - \xi(n)$, $\bar{f}(x) = 1 - \min[f(x), 1]$, and $\bar{g}(x) = 1 - \min[g(x), 1]$.

III. Stochastic Arrival Curve for Train Control Services

As discussed in Section II, Position Report (PR) messages are transmitted in uplink direction from OBC (train) to RBC (ground) and Movement Authority (MA) messages are transmitted in downlink direction from RBC (ground) to OBC (train) periodically. Therefore, we use a periodic traffic source $A(n)$ to model each type of traffic with different parameters. The source generates σ units of workload at times $\{n = U\tau + c\tau, c = 0, 1, \ldots\}$ where τ is the period of the source and U is the initial start time which is uniformly distributed in the interval [0,1]. For all $n \geq 0$ and $\theta \geq 0$ it is known that the MGF of $A(n)$ is

$$\mathrm{M}_A(\theta, n) = e^{\theta\sigma\lfloor\frac{n}{\tau}\rfloor}\left[1 + \left(\frac{n}{\tau} - \left\lfloor\frac{n}{\tau}\right\rfloor\right)(e^{\theta\sigma} - 1)\right] \quad (21)$$

while the effective bandwidth of $A(n)$ is

$$\delta_A(\theta, n) = \frac{\sigma}{n}\left\lfloor\frac{n}{\tau}\right\rfloor + \frac{1}{\theta n}\log\left[1 + \left(\frac{n}{\tau} - \left\lfloor\frac{n}{\tau}\right\rfloor\right)(e^{\theta\sigma} - 1)\right]. \quad (22)$$

Now we derive the v.b.c. stochastic arrival curve of the periodic source $A(n)$ according to the following theorem, whose proof is given in Appendix C.

Theorem 4: A flow $A(n)$ with effective bandwidth $\delta_A(\theta, n)$ has stationary increments, then it has a v.b.c. stochastic arrival curve $A \sim_{vb} \langle\alpha, f\rangle$, where

$$\alpha(n) = [\delta(\theta, n) + \theta_1] \times n \quad (23)$$

$$f(x) = \frac{e^{-\theta\theta_1}}{1 - e^{-\theta\theta_1}}e^{-\theta x}. \quad (24)$$

for any $\theta_1 > 0$ and $\theta > 0$.

IV. Stochastic Service Curve for HSR Fading Channel

A. Mobility Model

We divide the communication region of a serving eNodeB along the railway line into multiple zones, $\mathbb{Z} = \{1, 2, \ldots, Z\}$, as shown in Fig. 3, where in each spatial zone z, $z \in \mathbb{Z}$, the average received Signal to Interference and Noise Ratio (SINR) over the wireless channel between the serving eNodeB and the VS on the train is approximately the same, denoted by $\bar{\gamma}_z^{\mathrm{DL}}$ for downlink and $\bar{\gamma}_z^{\mathrm{UL}}$ for uplink. Let d_z denote the length of zone z and c_z denote the average distance between the serving eNodeB and a train in zone z. The average received SINR is determined by

$$\bar{\gamma}_z^{\mathrm{DL}} = \frac{P_{\mathrm{eNB}} \times PL(c_z)}{N_0 W + I_z^{\mathrm{DL}}} \quad (25)$$

$$\bar{\gamma}_z^{\mathrm{UL}} = \frac{P_{\mathrm{VS}} \times PL(c_z)}{N_0 W + I_z^{\mathrm{UL}}}. \quad (26)$$

where P_{eNB} and P_{VS} are the transmit power of eNodeB and VS, respectively. $PL(c_z)$ is the path loss between the eNodeB and the VS given their distance c_z. N_0 is the noise spectral density and W is the system bandwidth. I_z^{DL} and I_z^{UL} denote the average received interference power in uplink and downlink for zone z, respectively.

Since the eNodeBs are deployed along the railway line, we only consider interference from the two neighboring cells to the left and right of the considered cell. We consider the worst-case scenario where both neighboring cells are active and cause interference to the considered cell as shown in Fig. 3. Let cr_z and cl_z denote the average distances between a train in zone z and its right and left neighbor eNodeB, respectively. For the downlink, we have

$$I^{\mathrm{DL}} = P_{\mathrm{eNB}} \times PL(cr_z) + P_{\mathrm{eNB}} \times PL(cl_z). \quad (27)$$

For the uplink, we do not know the exact locations of the trains in the neighboring cells. However, we consider that the stationary probability π_z of there is a train in zone z, $\forall z \in \mathbb{Z}$ is the same for all the cells, which will be determined by our mobility model below. Therefore, we calculate the expected path loss between a train in a neighboring cell to the serving eNodeB given π_z, and the uplink interference power can be derived as

$$I^{\mathrm{UL}} = P_{\mathrm{VS}} \times \underbrace{\sum_{z=1}^{Z}(\pi_z \times PL(cl_z))}_{\substack{\text{expected path loss in} \\ \text{right neighboring cell}}} + P_{\mathrm{VS}} \times \underbrace{\sum_{z=1}^{Z}(\pi_z \times PL(cr_z))}_{\substack{\text{expected path loss in} \\ \text{left neighboring cell}}}. \quad (28)$$

Note that the first term is the uplink interference power from the right neighboring cell because the distance between a train in zone z of right neighboring cell and the serving eNodeB equals the distance cl_z between a train in zone z of the considered cell and the left neighbor eNodeB. For similar reason, the second term is the uplink interference power from the left neighboring cell. We will use $\bar{\gamma}_z$ to represent either uplink or downlink average SINR in the rest of the paper.

The movement of trains is modeled by a stochastic process $\{\mathbb{Z}_t, t = 0, 1, \ldots\}$ with discrete state space $\mathbb{Z}$, in which each state corresponds to one spatial zone. A discrete and integer time scale is adopted: t and $t + 1$ correspond to the beginning of two consecutive time slots, where the duration of a time slot $\Delta T = 1$ ms in LTE system. Within the duration of a time slot, a train either moves to the next zone, or remains in the current zone.[1] If a train leaves the current eNodeB and connects to a new eNodeB, it is regarded to move from state Z back to state 1 in the stochastic process, representing a new round of communication. Let the duration for which the trains stay in zone z be a random variable (r.v.) t_z, which is determined by the length of the partition zone d_z and the speed of trains v_t representing the distance the trains move during a time

[1]We reasonably assume that the length of a zone is large enough so that a train cannot move across multiple zones during a time slot. For example, the distance a train moves during one time slot equals 0.083 m when $v = 300$ km/h and $\Delta T = 1$ ms, while the length of a zone is at least several meters.

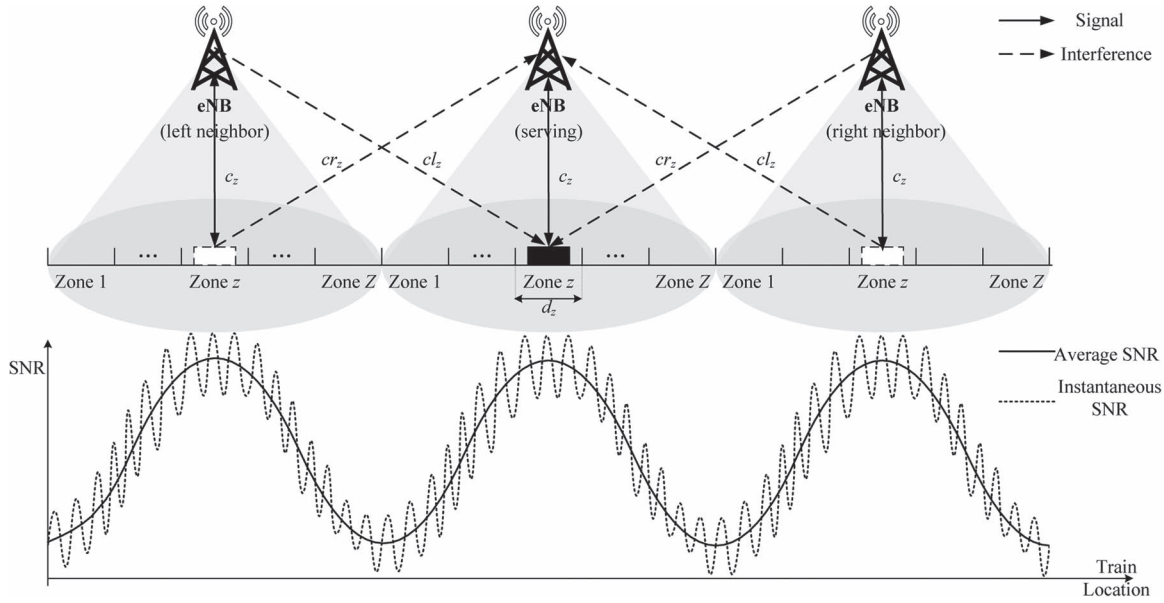

Fig. 3. HSR fading channel.

slot as $d_z = \sum_{t=1}^{t_z} v_t$, where $\{v_t, t = 0, 1, \ldots\}$ is considered as a discrete time stationary process after a train moves away from the station and accelerates to its maximum speed $v_{\max}$. Obviously, the distribution of t_z depends on that of the train speed v_t, which we only know is upper bounded by $v_{\max}$. In this paper, we consider the ideal scenario where $v_t = v_{\max}$ and thus $t_z = d_z/v_{\max}$. As the trains may have acceleration and decelerations due to unexpected random events in practical scenarios, e.g., communication delay of control signal etc., it is a challenging and interesting problem to determine the exact distribution of v_t and t_z.

The stochastic process $\{\mathbb{Z}_t, t = 0, 1, \ldots\}$ representing the movement of trains as described above is a semi-Markov process associated with a Markov renewal process $\{(X(n), T(n)), n = 0, 1, \ldots\}$ with some semi-Markov kernel $Q(z, y, t)$. Specifically, $X(n) \in \mathbb{Z}$ is the n-th state visited by the semi-Markov process and $T(n)$ is the time of this visit such that

$$\mathbb{Z}_t = X(n) \text{ whenever } T(n) \leq t < T(n+1). \quad (29)$$

Moreover, the semi-Markov kernel gives

$$Q(z, y, t) = P\left[X(n+1) = y, \; T(n+1) - T(n) \leq t | X(n) = z\right] \quad (30)$$

for all $z, y \in \mathbb{Z}$ and $t \geq 0$. The process $\{X(n), n = 0, 1 \ldots\}$ is a Markov chain with transition matrix $\mathbf{P}$, where each element $P(z, y)$ equals

$$P(z,y) = Q(z,y,+\infty) = \begin{cases} 1 & \text{if } z < Z,\ y = z+1 \text{ or} \\ & \quad z = Z,\ y = 1 \\ 0 & \text{Otherwise.} \end{cases} \quad (31)$$

Note that from (31) we can derive the stationary probabilities $\{\pi_z, z \in \mathbb{Z}\}$ if a train in zone z, which are used in (28) to derive the uplink interference from neighboring cells.

The above semi-Markov model of the train mobility process does not require that the speed of trains to be a deterministic value as assumed in this paper. Specifically, if $T(n)$ is geometrically distributed, the stochastic process $\{\mathbb{Z}_t, t = 0, 1, \ldots\}$ reduces to a Markov chain [36]. In the discrete time Markov chain $\{X(n), n = 0, 1 \ldots\}$, n and $n+1$ correspond to the beginning of two consecutive time units, where the duration $T(n)$ of a time unit n equals the duration for which the train stay in zone $z \in \mathbb{Z}$ if $X(n) = z$, i.e., t_z time slots or t_z ms. We will use s and t for the index of 1 ms time slots and k and n for the index of t_z ms time units in the rest of the paper.

B. Data Rate Process

Within any spatial zone z, the instantaneous received SINR $\gamma_{z,t}$ over the wireless channel between the eNodeB and the train is also affected by small-scale fading apart from large-scale fading, which makes $\gamma_{z,t}$ deviate from its average value $\bar{\gamma}_z$, as shown in Fig. 3. Due to the large fading rate $f_D \Delta T$ induced by the high mobility speed of the trains, $\gamma_{z,t}$ can be regarded as i.i.d. random variables over different time slots t [30]. Since the high speed trains typically run on the viaduct (such as in Chinese HSR), so the line of sight (LoS) path typically exists in the multipath environment. Thus, the multipath fast fading can be described using a Rician channel model [32], [37]. The instantaneous received SINR $\gamma_{z,t}$ in the downlink can be derived as

$$\gamma_{z,t}^{\mathrm{DL}} = \frac{P_{\mathrm{eNB}} \times PL(c_z) \times |h_t|^2}{N_0 W + P_{\mathrm{eNB}} \times PL(cr_z) \times |ir_t|^2 + P_{\mathrm{eNB}} \times PL(cl_z) \times |il_t|^2} \quad (32)$$

where $|h_t|$, $|ir(t)|$ and $|il(t)|$ are Rice distributed random variables whose square represent the small-scale fading gain

TABLE I
AMC PARAMETERS FOR LTE

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
Modulation order	2	2	4	4	6	6
Rate $r_z^{\text{ideal}}(bits/ms/180KHz)$	56	120	208	280	408	552
a_l	4.194	5.521	8.013	16.7	12.7	15.12
g_l	3.133	1.521	0.947	0.6359	0.2964	0.1211
γ_{pl}(dB)	-3.395	0.505	3.419	6.462	9.332	13.508
Γ_l(dB)	-0.37	3.09	5.63	8.31	11.23	15.31

of received signal power, received interference power from the right and left neighboring cells at time slot t, respectively. The instantaneous received SINR in the uplink can be derived similarly.

Since the average received signal power $P_{\text{eNB}} \times PL(c_z)$ or $P_{\text{VS}} \times PL(c_z)$ is usually much stronger than the average received interference power I_z^{DL} or I_z^{UL}, we ignore the effect of fast fading on the received interference power and approximate the denominator of (32) by $N_0 + I_z^{\text{DL}}$. Therefore, we have $\gamma_{z,t} = \bar{\gamma}_z |h_t|^2$. Using the more general and simpler Nakagami-m fading model to approximate the Rician fading model, the probability distribution function (PDF) of the SINR $\gamma_{z,t}$ can be presented by [38]

$$f(\gamma_{z,t}) = \left(\frac{m}{\bar{\gamma}_z}\right)^m \frac{\gamma_{z,t}^{m-1}}{\Gamma(m)} \exp\left(-\frac{m\gamma_{z,t}}{\bar{\gamma}_z}\right) \tag{33}$$

where $\Gamma(\cdot)$ is the Gamma function, $m = \frac{(K+1)^2}{2K+1}$ is the fading parameter, and K is the Rice factor.

The instantaneous data rate $r_{z,t}$ within spatial zone z can be determined from the instantaneous received SNR $\gamma_{z,t}$. The simplest method is using the Shannon formula, where the instantaneous data rate within spatial zone z is a random variable $r_{z,t} = C\log(1+\gamma_{z,t})$. However, this method can only provide an approximate data rate which is not accurate. In this paper, we consider the practical scenario where the instantaneous data rate within spatial zone z is determined by the adaptive modulation and coding (AMC) scheme. The SINR values are divided into L non-overlapping consecutive regions. Ideally, perfect channel state information (CSI) is available at the eNodeB, based on which the optimum modulation and coding scheme (MCS) can be selected. For any $l \in \{1, \ldots, L\}$, the l-th MCS is selected if the instantaneous SINR value γ falls within the l-th region $[\Gamma_l, \Gamma_{l+1})$. Obviously, $\Gamma_0 = 0$ and $\Gamma_{L+1} = \infty$. However, due to the rapidly varying channel condition induced by the high mobility of HST, the CSI at the eNodeB may be highly inaccurate. Therefore, the performance of the CSI-based AMC scheme may be seriously degraded. As an alternative, we propose to perform AMC based on the average received SINR instead, since the average received SINR is mainly impacted by the large-scale fading effect and varies on a much slower time scale than the instantaneous SINR. Based on the above assumption, a fixed MCS scheme is selected for each zone z according to its average SINR $\bar{\gamma}_z$, i.e., the l-th MCS is selected for zone z if $\bar{\gamma}_z$ falls within the l-th region $[\Gamma_{l-1}, \Gamma_l)$. The selected MCS determines the ideal transmission capability r_z^{ideal} of the wireless channel in zone z. However, as the channel condition is also impacted by the small-scale fading effect, transmission errors may be incurred when the channel is in deep fade. To simplify performance analysis, we will rely on the following approximate block error rate (BLER) expression over Additive White Gaussian Noise (AWGN) channel [39]:

$$\text{BLER}_l(\gamma) = \begin{cases} 1 & \text{if } 0 < \gamma < \gamma_{pl} \\ a_l \exp(-g_l \gamma) & \text{if } \gamma \geq \gamma_{pl} \end{cases} \tag{34}$$

Parameters a_l, g_l, and γ_{pl} are MCS-dependent, and are obtained by fitting and comparing curves by (34) to the simulated BLER according to the Monte-Carlo simulations with parameters given by 3G LTE specification [40]. We select $L = 6$ MCSs from the 32 MCSs in LTE and the parameters are given in Table I. In LTE system, a terminal can be allocated in the downlink or uplink with a minimum of 1 Resource Block (RB) during 1 subframe (1 ms), where an RB occupies 12 subcarriers (12×15 KHz $= 180$ KHz) in frequency domain. Therefore, the data rate r_z^{ideal} in Table I is the number of bits that can be transmitted on one RB within 1 ms time slot.[2] We consider an infinite-persistent ARQ protocol in the link layer, where an erroneous block is retransmitted until it is received correctly at the receiving end. Depending on the transmission outcome in each time slot, an acknowledgment (ACK) or a negative acknowledgement (NACK) is replied by the receiver to the transmitter for each transmitted packet. We assume that the ACK/NACK packets are available at the end of the transmission time slot, and the feedback channel carrying ACK/NACK packets is a reliable one. Based on the above assumptions, the instantaneous data rate of zone z given MCS index l is a random variable [41]

$$r_{z,t} = r_z^{\text{ideal}} \left(1 - \text{BLER}_l(\gamma_{z,t})\right). \tag{35}$$

The *data rate process* of a HSR wireless communication channel as described above can be modeled by a semi-Markov Modulated Process (SMMP). The modulation is done via a discrete-time homogeneous semi-Markov process (SMP) $\{\mathbb{Z}_t, t = 0, 1, \ldots\}$ on the states $\{1, 2, \ldots, Z\}$. Let $\{r_{z,t}, t = 0, 1, \ldots\}$, $z = 1, \ldots, Z$, be Z sequences of i.i.d. random variables, representing the instantaneous data rate at time slot t when the SMP $\mathbb{Z}_t$ is at state z. The *data rate process* $r_t = r_{\mathbb{Z}_t,t}$ is then an SMMP with the modulating process $\mathbb{Z}_t$.

[2]The term "time slot" in our paper is the same with the term "subframe" in LTE terminology.

C. Stochastic Service Curve

Define the *service process* $S_t \equiv \sum_{s=1}^{t} r_s$ as the cumulative amount of service provided by the wireless channel by time t. If the *data rate process* r_t as defined above is a Markov Modulated Process (MMP), the stochastic service curve of S_t can be derived. However, as r_t is an SMMP, we construct an *equivalent data rate process* which is an MMP. The modulation is done via a discrete-time homogeneous Markov process $\{X(n), n = 0, 1, \ldots\}$ on the states $\{1, 2, \ldots, Z\}$. Let $\{r_z(n) := \sum_{t=1}^{t_z} r_{z,t}, n = 0, 1 \ldots, z = 1, \ldots, Z\}$, be Z sequences of i.i.d. random variables, representing the total achievable data rate during the n-th state visited by the SMP $\{\mathbb{Z}_t, t = 0, 1, \ldots\}$, when the n-th state is state z. The *equivalent data rate process* $r(n) = r_{X(n)}(n)$ is then an MMP with the modulating process $X(n)$. We define the *equivalent service process* $S(n)$ as

$$S(n) := \sum_{k=1}^{n} r(k) = S\left(\sum_{k=1}^{n} t_{X(k)}\right). \tag{36}$$

1) MGF Snetal: Define $\phi_{\mathrm{S},z}(-\theta) := \mathrm{E}[e^{-\theta r_z(n)}] = (\mathrm{E}[e^{-\theta r_{z,1}}])^{t_z}$ as the MGF of $r_z(n)$ and let $\boldsymbol{\phi}_{\mathrm{S}}(\theta)$ be the diagonal matrix $\mathrm{diag}\{\phi_{\mathrm{S},1}(-\theta), \ldots, \phi_{\mathrm{S},Z}(-\theta)\}$. For all $n \geq 0$ and all $\theta > 0$, the MGF of the *equivalent service process* $S(n)$ can be derived as [42]

$$\overline{\mathrm{M}}_S(\theta, n) = \boldsymbol{\pi} \left(\boldsymbol{\phi}_{\mathrm{S}}(-\theta)\mathbf{P}\right)^{n-1} \boldsymbol{\phi}_{\mathrm{S}}(-\theta)\mathbf{1}. \tag{37}$$

where $\boldsymbol{\pi}$ is a row vector of the stationary state distribution of the modulating process $X(n)$, $\mathbf{P}$ is the transition matrix of $X(n)$ given in (31), and $\mathbf{1}$ is a column vector of ones.

2) CCDF Snetal: Now we use two stochastic processes to characterize the equivalent service process $S(n)$, i.e., an ideal deterministic service process $\hat{S}(n) = \hat{r} \times n$ and an impairment process $I(n)$, where $S(n) = \hat{S}(n) - I(n)$ with $\hat{S}(0) = I(0) = 0$ by convention. Therefore, $I(n) = \sum_{k=1}^{n}(\hat{r} - r(k))$ according to (35). We can see that the impairment process is also the cumulative process of an MMP $i(n) := i_{X(n)}(n)$ with modulating process $X(n)$, where $i_z(n) := \hat{r} - r_z(n)$ $(n = 0, 1, \ldots, z = 1, \ldots, Z)$ are Z sequences of i.i.d. random variables, representing the amount of impaired services during the n-th state visited by the SMP $\{\mathbb{Z}_t, t = 0, 1, \ldots\}$, when the n-th state is state z. Now, the *equivalent service process* $S(n)$ is a stochastic strict server with strict service curve $\hat{\beta}(n) = \hat{S}(n)$ and impairment process $I(n)$ by Definition 3.

Define $\phi_{\mathrm{I},z}(\theta) := \mathrm{E}[e^{\theta i_z(n)}] = e^{\theta\hat{r}}(\mathrm{E}[e^{-\theta r_{z,1}}])^{t_z}$ as the MGF of $i_z(n)$ and let $\boldsymbol{\phi}_{\mathrm{I}}(\theta)$ be the diagonal matrix $\mathrm{diag}\{\phi_{\mathrm{I},1}(\theta), \ldots, \phi_{\mathrm{I},Z}(\theta)\}$. Let $sp(\boldsymbol{\phi}_{\mathrm{I}}(\theta)\mathbf{P})$ be the spectral radius[3] of the matrix $\boldsymbol{\phi}(\theta)\mathbf{P}$, where the transition matrix $\mathbf{P}$ is given in (31). For all $n \geq 0$ and all $\theta > 0$, the effective bandwidth of the impairment process $I(n)$ can be derived as [42]

$$\delta_I(\theta, n) = \frac{1}{\theta n} \log\left(\boldsymbol{\pi} \left(\boldsymbol{\phi}_{\mathrm{I}}(\theta)\mathbf{P}\right)^{n-1} \boldsymbol{\phi}_{\mathrm{I}}(\theta)\mathbf{1}\right). \tag{38}$$

[3]The spectral radius of a matrix is the maxima of the absolute values of the eigenvalues of that matrix.

The impairment process $I(n)$ can be characterized by v.b.c. stochastic arrival curve according to the following Lemma, whose proof is similar to that of Theorem 4 and omitted here.

Lemma 1: If impairment process $I(n)$ with effective bandwidth $\delta_I(\theta, n)$ has stationary increments, then it has a v.b.c. stochastic arrival curve $A \sim_{vb} \langle \xi, g \rangle$, where

$$\xi(n) = [\delta_I(\theta, n) + \theta_1] \times n \tag{39}$$

$$g(x) = \frac{e^{-\theta\theta_1}}{1 - e^{-\theta\theta_1}} e^{-\theta x} \tag{40}$$

for any $\theta_1 > 0$ and $\theta > 0$.

Given the stochastic strict server and v.b.c. stochastic arrival curve of the impairment process, Theorem 3 can be applied to derive the stochastic backlog and delay bounds using independence case analysis.

Alternatively, we can first characterize the *equivalent service process* $S(n)$ using weak stochastic service curve according to the following theorem.

Theorem 5: The *equivalent service process* $S(n)$ provides a weak stochastic service curve, i.e., $S \sim \langle g, \beta \rangle$, where

$$\beta(n) = [\hat{r} - \delta_I(\theta, n) - \theta_1]^+ n \tag{41}$$

$$g(x) = \frac{e^{-\theta\theta_1}}{1 - e^{-\theta\theta_1}} e^{-\theta x} \tag{42}$$

for $\forall\theta > 0$ and $\theta_1 > 0$.

Proof: The proof follows directly from Lemma 1 and Lemma 3 in Appendix B. ■

Given the weak stochastic service curve of the *equivalent service process* $S(n)$, Theorem 2 can be applied to derive the stochastic backlog and delay bounds.

V. Performance Evaluation

A. Derivation of Delay Bound

Given the SAC of train control services in Section III and the SSC provided by the HSR fading channel in Section IV, the stochastic delay bound of the flow can be determined using the following three methods.

1) MGF method: Theorem 1 is used to derive the delay bound where the MGFs of the arrival process and service process are derived from (21) and (37), respectively;
2) CCDF method: For ease of notation, we denote $m = \inf_{k \geq 0}[\beta(k) - \alpha(k - x)]$.

   a) method 1: With the v.b.c. stochastic arrival curve of arrival process given in Theorem 4 and the weak stochastic service curve of service process given in Theorem 5, the delay bound can be derived by Theorem 2. Taking $f(x) = g(x) = \frac{e^{-\theta\theta_1}}{1-e^{-\theta\theta_1}} e^{-\theta x}$ from (24) and (42) into (17), we have

$$P\{D(n) > x\} \leq \frac{2e^{-\theta\theta_1}}{1 - e^{-\theta\theta_1}} e^{\frac{-\theta m}{2}} \tag{43}$$

TABLE II
SYSTEM PARAMETERS

Parameter	Value
Transmit power of eNodeB P_{eNB}	43dBm
Transmit power of VS P_{VS}	23dBm
Bandwidth W	3MHz
Noise spectral density N_0	-174dBm/Hz
Carrier frequency f_c	1.9GHz
Train velocity $v_{\max}/\Delta T$	100m/s
Inter-site Distance	3km
length of a zone d_z	5m

b) method 2: With the v.b.c. stochastic arrival curve of arrival process given in Theorem 4, and the stochastic strict server of service process with v.b.c. stochastic arrival curve of impairment process given in Lemma 1, the delay bound can be derived by Theorem 3. When taking $f(x) = g(x) = \frac{e^{-\theta\theta_1}}{1-e^{-\theta\theta_1}} e^{-\theta x}$ into (20), we have

$$P\{D(n) > x\} \leq 1 - \frac{e^{-\theta\theta_1}}{1-e^{-\theta\theta_1}}(1-e^{-\theta m}) + \left(\frac{e^{-\theta\theta_1}}{1-e^{-\theta\theta_1}}\right)^2 \theta m e^{-\theta m}. \quad (44)$$

Note that $m = \inf_{k\geq 0}[(\hat{r} - \delta_I(\theta,k) - \delta_A(\theta,k-x) - 2\theta_1)k + (\delta_A(\theta,k-x)+\theta_1)x]$ according to (41) and (23). θ and θ_1 are free parameters to optimize the performance of the delay bound so that $P\{D(n) > x\}$ can be as small as possible.

B. System Parameter

The system parameters are given in Table II. We use the Winner Phase II model D2a sub-scenario to calculate the path loss $PL(d)$ in (25)–(28), which is a measurement based physical layer channel model for links between the trackside base station and the roof-top antenna of a train

$$PL(d) = \begin{cases} 44.2 + 21.5\log(d) + L & d < d_{bp} \\ 44.2 + 40\log(d/d_{bp}) + L_{bp} + L & d \geq d_{bp} \end{cases} \quad (45)$$

where d is the distance from the roof-top antenna of the train to the eNodeB, which could be either c_z, cr_z or cl_z in (14)–(17). The distance from the eNodeB to the track is set to 50 m in order to calculate the above distances. L and L_{bp} are constants in dB. $L = 20\log(f_c/(5\times 10^9))$ is the carrier frequency loss, where f_c is the carrier frequency in Hz. $L_{bp} = 21.5\log(d_{bp})$, where d_{bp} is the break point of the path-loss curve. d_{bp} equals to $4h_{\mathrm{eNB}}h_{\mathrm{VS}}f_c/c$, where $h_{\mathrm{eNB}} = 45$ m and $h_{\mathrm{VS}} = 5$ m are the eNodeB and VS antenna heights in meter compared to the ground, respectively, and c is velocity of light in vacuum.

The system bandwidth is 3 MHz containing 15 RBs. However, we assume that only one RB is dedicated to the train control services in the following numerical experiments. We set the velocity of trains to be 100 m/s if not specified otherwise in the following numerical experiments, so that $v_{\max} =$ 0.1 m/time slot. Moreover, the inter-site distance between two neighboring eNodeBs is set to be 3 km. Let the length of a zone z be $d_z = 5$ m, and we have the duration for which a train stays in zone z is $t_z = d_z/v_{\max} = 50$ time slots or ms for any $z \in \mathbb{Z}$. Moreover, the number of zones $Z = 600$. Note that smaller zone size d_z will result in smaller time unit duration t_z. As our delay bound is in terms of time unit, shorter length of a time unit shall result in more precise measurement of the delay bound. Therefore, we should set d_z to be as small as possible. However, smaller d_z will lead to larger number of zones Z within a cell and thus larger state space of the semi-Markov process, which results in larger computational complexity in analysis. Moreover, if d_z is too small, a train may move from zone z to $z+i$ with $i > 1$ during a time slot, which further complicates the analysis. Therefore, the zone size d_z should be set to a proper value according to the train speed and coverage region of a cell considering the above tradeoff.

C. Numerical and Simulation Results

In this section, the delay performance of train control services over HSR fading channel is evaluated through both simulation and numerical results. Our simulation program is built on the MATLAB platform. The BS and the VS each has a buffer, where the arrived packets wait for transmission. At the start of each 50 ms time unit when the train is in zone z, the instantaneous SNR values $\gamma_{z,t}$ of $t_z = 50$ i.i.d. Rician fading channels with mean SNR $\bar{\gamma}_z$ are generated, each of which represents the instantaneous SNR of the HSR fading channel during 1 ms time slot. Then, the instantaneous data rate of each Rician fading channel is derived using both the AMC method by (35) and the Shannon method by the Shannon formula, respectively. The sum of the $t_z = 50$ instantaneous data rates represents the total amount of data that can be transmitted by the HSR fading channel during the period when the train is in zone z. The sojourn time of each data unit in buffer is recorded when it is transmitted. With this, the delay performance is obtained. The results are collected over Z simulations, each of which runs for 10^6 time units, where in the z-th simulation ($z \in \{1,\ldots,Z\}$) the train is assumed to be in zone z when the simulation starts. In the following experiments, we focus on the downlink transmission, since the analytical principles for deriving the stochastic delay bounds of uplink and downlink transmissions are the same.

Figs. 4 and 5 compare the analytical bounds by MGF and CCDF snetals and the simulation results under different violation probabilities, where the burst size and period of the periodical source are set to $\sigma = 4000$ bits and $\tau = 120$ time units (6 s), respectively. Fig. 4 uses the proposed AMC method to obtain instantaneous data rate, while Fig. 5 uses the Shannon method. As expected, the estimated delay bound with Shannon method is smaller than that with the AMC method, which means that the Shannon method will provide results that are more optimistic than that can be actually achieved. It can be observed that in both figures, the analytical bounds provided by the MGF method are the tightest while those provided by the CCDF method 1 are the loosest. This is because the MGF method only uses the Boole's inequality and Chernoff bound when deriving the analytical bound by (46), while both CCDF methods use the above inequalities twice (when obtaining the

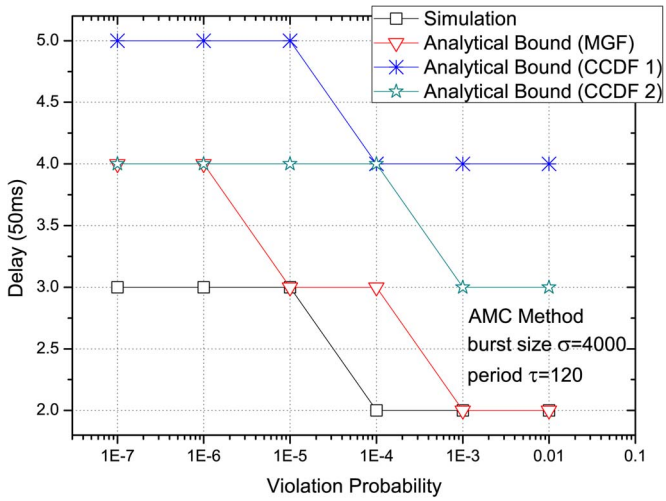

Fig. 4. Comparison of simulation results and analytical bounds under different violation probabilities with AMC method (burst size $\sigma = 4000$ bits, period $\tau = 120$ time units (6 s)).

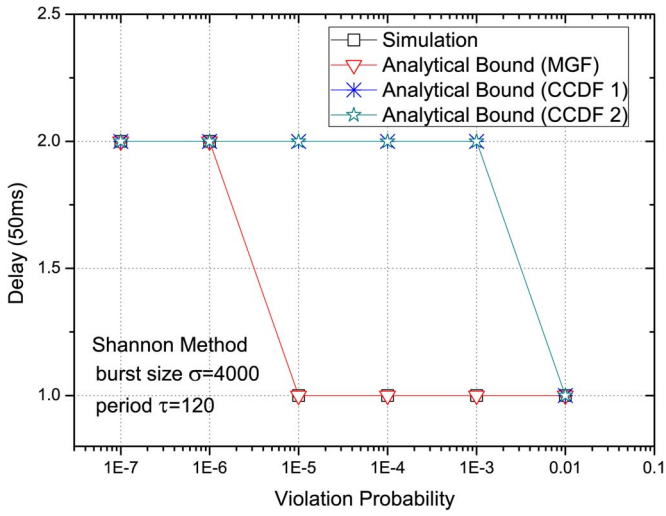

Fig. 5. Comparison of simulation results and analytical bounds under different violation probabilities with Shannon method (burst size $\sigma = 4000$ bits, period $\tau = 120$ time units (6 s)).

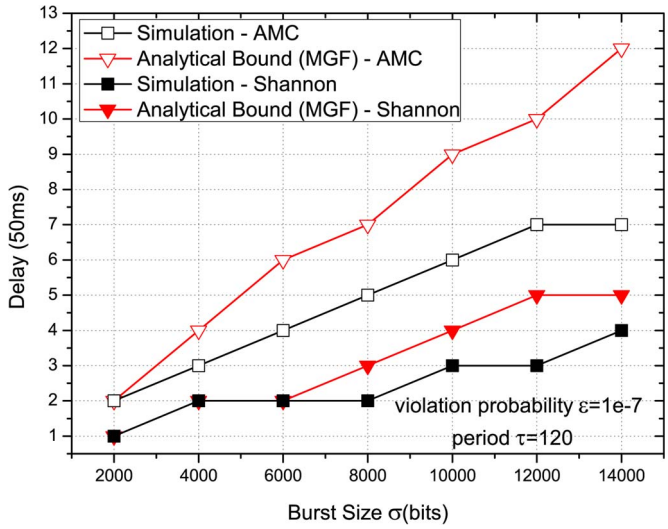

Fig. 6. Comparison of simulation results and analytical bounds for periodical source under different burst sizes with AMC method and Shannon method (violation probability $\varepsilon = 1\mathrm{e}-7$ bits, period $\tau = 120$ time units (6 s)).

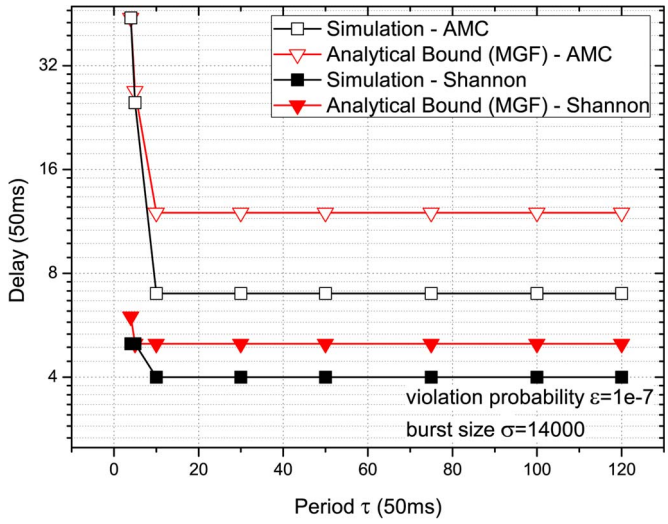

Fig. 7. Comparison of simulation results and analytical bounds under different periods with AMC method and Shannon method (violation probability $\varepsilon = 1\mathrm{e}-7$ bits, burst size $\sigma = 14000$ bits.

stochastic arrival curves for periodical source and impairment process by (51)). Moreover, CCDF method 2 provides tighter bound than CCDF method 1 in Fig. 4 and the same bound with CCDF method 1 in Fig. 5, because the second bound in Lemma 2 is generally better than the first bound. Note that the MGF snetal is easier to implement than the CCDF snetal, because that in the MGF snetal, only one free parameter θ needs to be optimized in order to derive the performance bound as in (14). In the CCDF snetal, on the other hand, two free parameters θ and θ_1 need to be optimized as in (43) and (44). Therefore, in Figs. 6 and 7, we only use MGF snetal to derive the analytical bounds.

Fig. 6 compares the analytical bounds by MGF snetal and the simulation results under different burst sizes with AMC method and Shannon method, where the period of the periodical source is set to $\tau = 120$ time units (6 s) and the violation probability is set to $1\mathrm{e}-7$, respectively. Fig. 6 shows that the delay bound increases with the increasing burst size, and the increasing rate of Shannon method is slower than that of the AMC method. This is because that the period of burst arrival is $\tau = 120$ time units and the largest delay experienced by the data in the buffer (when the burst size is 14000 bits with AMC method) is larger than 7 time units with probability no larger than $1\mathrm{e}-7$. Therefore, we can safely conclude that a burst is fully transmitted before the next one arrives, which means that the largest backlog equals the burst size and thus the delay bound depends on the burst size and the instantaneous data rate at every zone. From Fig. 6 we can also observe that the MGF method can provide a relatively tight bound with both AMC method and Shannon method.

Fig. 7 compares the analytical bounds by MGF snetal and the simulation results under different periods with AMC method and Shannon method, where the burst size is set to

$\tau = 14000$ bits and the violation probability is set to $1\mathrm{e}-7$, respectively. It can be observed that the delay bound remains to be the same when the period reduces from 120 time units to 10 time units for both the AMC method and Shannon method. However, when the period reduces from 10 time units to 4 time units, the delay bound grows quickly for the AMC method (from 7 to 44 time units in the simulation results) and grows a little for the Shannon method (from 4 to 5 time units in the simulation results). This observation is because when the period is larger than 10 time units, a burst is almost always fully transmitted before the next one arrives, as explained in Fig. 6. However, when the period becomes smaller than 10 time units, a burst may not be fully transmitted before the next one arrives, so the remaining data in the previous burst will be backlogged to the next period for transmission. For the AMC method, since the delay bound corresponding to violation probability $1\mathrm{e}-7$ is 7 time units (in the simulation result) when the period is 10 time units, the delay bound increase quickly when the period reduces to 4 and 5 time units. When the period further reduces to be smaller than 4 time units, the MGF snetal fails to derive the delay bound since the term $\mathrm{M}_A(\theta, k-\tau)\overline{\mathrm{M}}_S(\theta, k)$ in (14) increases with increasing k and the sum of this term over $k = \{0, 1, \ldots\}$ becomes infinity. This is because the traffic intensity becomes too large so the queuing system is not stable anymore and the backlog accumulates to infinity over time. On the other hand, for the Shannon method, since the delay bound corresponding to violation probability $1\mathrm{e}-7$ is 4 time units (in the simulation result) when the period is 10 time units, the bursts are fully transmitted before the next one arrives even when the period is reduced to 4 time units. Therefore, the delay bound only increases a little in this case.

From both Figs. 6 and 7, it can be seen that the analytical bounds for Shannon method is much tighter than those for AMC method. The reason for this difference is due to the mathematical principle of snetal. In snetal, some general purpose methods are commonly used in order to derive the performance bounds, such as the Chernoff's bound and Boole's inequality. The derived bounds may be tight or loose depending on the specific distribution of the arrival and service processes. Therefore, the usage of Shannon method or AMC method leads to different distributions of the service process, which results in the different tightness of the derived bounds.

In the above numerical experiments, we set the velocity of trains to be 100 m/s. Note that the train speed may affect the delay bound, since the channel variation due to changing path loss shall be faster with higher train speed. In order to examine its impact on the delay bound, we vary the train speed from 50 m/s to 200 m/s in a step of 50 m/s. Moreover, in order to facilitate comparison, we set the zone size d_z correspondingly so that the duration of a time unit t_z remains to be 50 ms irrespective of the train speed. Fig. 8 shows the analytical bound and simulation results under different train speeds with AMC method when the burst size $\sigma = 14000$ bits, period $\tau = 4$ time units and violation probability $\varepsilon = 1\mathrm{e}-7$. It can be seen that the delay bound is improved with increasing train speed. Although not shown in Fig. 8, the impact of train speed reduces to almost zero when we reduce the burst size or increase the period. This is because the delay bound improves when the train speed is higher because the train will travel longer distance and thus the HSR fading channel will experience larger channel state variation during the transmission of a message. When the burst size is reduced or the period is increased, it can be seen from Figs. 6 and 7 that the message transmission delay is significantly reduced compared to that corresponds to the parameter setting in Fig. 8, which results in reduced impact of train speed on the delay bound.

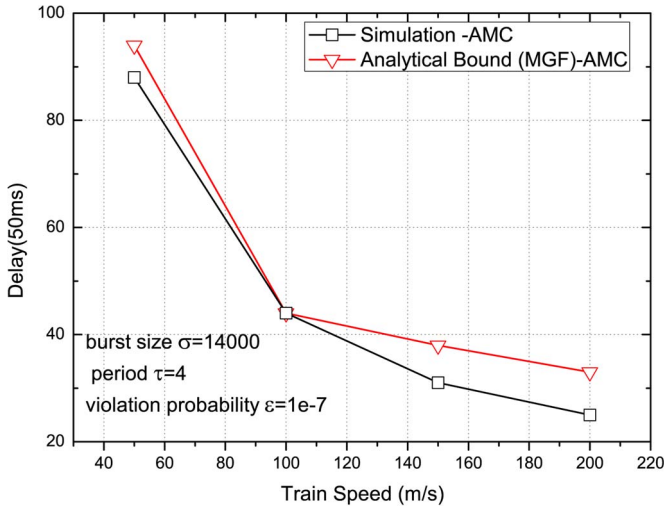

Fig. 8. Impact of train speed on delay bound (violation probability $\varepsilon = 1\mathrm{e}-7$ bits, burst size $\sigma = 14000$ bits, period $\tau = 4$ time units, AMC method).

We would like to remark that the length of the MA message in the current ETCS/CTCS system is typically around 1600 bits and the length of the PR message is 192 bits, and the arrival periods of both messages are typically around 6 s (120 time units) based on our measurement data in practical system. Moreover, it is required in the ETCS/CTCS system that the maximum end-to-end transfer delay should be $\leq$0.5 s (10 time units) under 99% probability [21]. Therefore, we can conclude that the LTE system can provide satisfactory performance guarantee for the train control services using one RB based on our analytical and simulation results above.

Generally speaking, the analytical principles and delay bounds derived for uplink and downlink transmissions are the same for both FDD-LTE and TD-LTE. However, for the total end-to-end transfer delay from the time when an integrity/position report is transmitted by Train1 until a movement authority message is received by Train2 as illustrated in Fig. 2, TD-LTE may cause several milliseconds of more delay than FDD-LTE. This is because in TD-LTE, a 10 ms frame is divided into 10 1ms subframes, which are reserved for downlink or uplink transmissions according to different configurations. Therefore, a message buffered at the train or base station needs to wait for the proper type of subframe for transmission, which may cause some further delay.

VI. Conclusion

In this paper, stochastic delay bounds of the train control services over HSR fading channel have been derived based on the analytical principle behind stochastic network calculus. In specific, the service process of the HSR fading channel

was modeled as a semi-Markov modulated process, where the channel variations due to both the large-scale fading and small-scale fading effects were taken into account. Moreover, the performance loss due to AMC selection with imperfect CSI was also considered. The stochastic service curve of the semi-Markov modulated service process was derived using both the MGF and CCDF methods. The train control service was modeled as a periodical source, where the stochastic arrival curve can be derived using both the MGF and CCDF methods. Based on the stochastic arrival and service curves, the stochastic delay bounds were derived using both the MGF and CCDF methods. It has been shown that for our specific arrival and service processes, the MGF method can provide tighter bound than the CCDF method and is also simpler to use than the CCDF method, since it only needs to optimize one free parameter. Moreover, we have also shown that the delay bound derived when considering AMC method is indeed more conservative than that derived when using the Shannon formula to derive the instantaneous data rate. The numerical and simulation results demonstrate that the LTE system can provide very good delay bounds for train control services using only one resource block. Our focus in this paper is on the performance of train control services which are transmitted over dedicated radio resources. The extension of our analysis method to support all three types of services including train monitoring services and passenger services with a priority queuing system is currently under investigation.

APPENDIX

A. Proof of Theorem 1

Due to space limitation, we will only prove (14) for the delay bound, while the backlog bound follows similarly [16]:

$$\begin{aligned} P\{D(n)>x\} &\le P\left\{\sup_{0\le k\le n}\left\{\hat{A}(k,n)-\hat{S}(k,n+x)\right\}>0\right\} \\ &\le \mathrm{E}\left[e^{\theta\sup_{0\le k\le n}\left\{\hat{A}(k,n)-\hat{S}(k,n+x)\right\}}\right] \\ &\le \sum_{k=0}^{n}\mathrm{E}\left[e^{\theta\hat{A}(k,n)-\theta\hat{S}(k,n+x)}\right] \\ &\le \sum_{k=x}^{\infty}\mathrm{M}_A(\theta,k-x)\overline{\mathrm{M}}_S(\theta,k) \end{aligned} \tag{46}$$

for $\forall\theta>0$. The first inequality stems from the Chernoff's bound of random variable X that for any x and all $\theta\ge 0$ it is known that

$$P\{X\ge x\}\le e^{-\theta x}\mathrm{E}e^{\theta X}$$

while the second inequality is based on Boole's inequality.

Suppose the allowable delay violation probability is $\varepsilon\in(0,1]$. Then, by letting the right hand side of (46) equal to ε and with some mathematical manipulation, a delay bound can be found as (14). Similarly, a backlog bound can be found as (13).

B. Proof of Theorem 3

The proof is based on the following lemma of probability theory.

Lemma 2: For any random variables X and Y, and $\forall x\ge 0$, if $P\{X>x\}\le f(x)$ and $P\{Y>x\}\le g(x)$, where $f,g\in\bar{\mathcal{F}}$, then

$$P\{X+Y>x\}\le(f\otimes g)(x).$$

If X and Y are nonnegative and independent, then

$$P\{X+Y>x\}\le 1-(\bar{f}*\bar{g})(x)$$

where $\bar{f}(x)=1-\min[f(x),1]$ and $\bar{g}(x)=1-\min[g(x),1]$. Note that the second bound for the CCDF of $X+Y$ may be significantly better than the first bound, which motivates the independent case analysis.

In the following, we first introduce and prove that the following lemma on weak stochastic service curve.

Lemma 3: Consider a stochastic strict server S providing strict service curve $\hat{\beta}$ with impairment process I. If the impairment process I provides a v.b.c. stochastic arrival curve, or $I\sim_{vb}\langle g,\xi\rangle$ then the server provides a weak stochastic service curve $S\sim_{ws}\langle g,\beta\rangle$ with $\beta(n)=[\hat{\beta}(n)-\xi(n)]^+$ if $\beta\in\mathcal{F}$.

Proof: For any time $n\ge 0$, there are two cases. In case 1, n is not within any backlogged period. In this case, there is no backlog in the system at time n, which implies that all traffic that arrived up to time n has left the server. Hence, $A^*(t)=A(t)$ and consequently $A\otimes\beta(n)-A^*(n)\le A(n)+\beta(0)-A^*(n)\le 0$.

In case 2, n is within a backlogged period. Without loss of generality, assume the backlogged period starts from n_0. Then, $A(n_0)=A^*(n_0)$ and

$$\begin{aligned} A\otimes\beta(n)-A^*(n) &\le A(n_0)+\beta(n-n_0)-A^*(n) \\ &= \beta(n-n_0)+A^*(n_0)-A^*(n) \\ &= \beta(n-n_0)-S(n,n_0). \end{aligned} \tag{47}$$

In addition, since the server is a stochastic strict server providing strict service curve $\hat{\beta}$ with impairment process I, we have for this backlogged period $(n_0,n]$, by definition

$$\begin{aligned} \beta(n-n_0)-S(n_0,n) &= \hat{\beta}(n-n_0)-S(n_0,n)-\xi(n-n_0) \\ &\le I(n_0,n)-\xi(n-n_0) \\ &\le \sup_{0\le k\le n}\{I(k,n)-\xi(n-k)\}. \end{aligned} \tag{48}$$

Combining both cases, we conclude that, for any $k\ge 0$

$$A\otimes\beta(n)-A^*(n)\le\left(\sup_{0\le k\le n}\{I(k,n)-\xi(n-k)\}\right)^+. \tag{49}$$

Since the impairment process $I(n)$ provides a v.b.c. stochastic arrival curve, or $I\sim_{vb}\langle g,\xi\rangle$. By definition, there holds

$$P\left\{\sup_{0\le k\le n}\{I(k,n)-\xi(n-k)\}>x\right\}\le g(x).$$

Hence we get

$$\begin{aligned} &P\{A\otimes\beta(n)-A^*(n)>x\} \\ &\qquad\le P\left\{\sup_{0\le k\le n}\{I(k,n)-\xi(n-k)\}>x\right\}\le g(x). \end{aligned}$$

which completes the proof by the definition of weak stochastic service curve. ■

Applying (49) to (15), we get

$$A(n) - A^*(n) \leq \sup_{0\leq k\leq n}\{A(k,n) - \alpha(n-k)\} + \left(\sup_{0\leq k\leq n}\{I(k,n) - \xi(n-k)\}\right)^+ + \sup_{n\geq 0}\{\alpha(n) - \beta(n)\}. \tag{50}$$

If A and I are independent random processes, since α, β, and ξ are non-random functions, the first two terms on the right-hand side of (50) are also independent. Then, together with the fact that A provides v.b.c. stochastic arrival curve, we have Theorem 3 by applying Lemma 2.

C. Proof of Theorem 4

$$\begin{aligned} P&\left\{\sup_{0\leq k\leq n}\{A(k,n) - [\delta(\theta,n)+\theta_1]\times(n-k)\} > x\right\} \\ &\leq \sum_{k=0}^{n-1} P\{A(k,n) - [\delta(\theta,n)+\theta_1]\times(n-k)\} > x\} \\ &= \sum_{u=1}^{t} P\{A(0,u) - [\delta(\theta,n)+\theta_1]\times u\} > x\} \\ &\leq \sum_{u=1}^{t}[e^{-\theta x}e^{-\theta\theta_1 u}] \leq \frac{e^{-\theta\theta_1}}{1-e^{-\theta\theta_1}}e^{-\theta x}. \end{aligned} \tag{51}$$

where the first inequality is based on Boole's inequality, while the second inequality stems from the Chernoff's bound.

REFERENCES

[1] G. Barbu, "E-Train—Broadband communication with moving trains technical report—Technology state of the Art," Int. Union Railways, Paris, France, Tech. Rep., Jun. 2010.

[2] A. Coraiola and M. Antscher, "GSM-R network for the high speed line Rome-Naples," *Signal Draht*, vol. 92, no. 5, pp. 42–45, 2000.

[3] *"Requirements for Further Advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced) (Release 11)*," 3GPP TR Specification 36.913, Technical Specification Group Radio Access Network, Sep. 2012.

[4] B. A. *et al.*, "Challenges toward wireless communications for high-speed railway," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2143–2158, Oct. 2014.

[5] K. Abboud and W. Zhuang, "Stochastic analysis of a single-hop communication link in vehicular ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2297–2307, Oct. 2014.

[6] N. Lu *et al.*, "Vehicles meet infrastructure: Toward capacity-cost tradeoffs for vehicular access networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1266–1277, Sep. 2013.

[7] R. Cruz, "A calculus for network delay—I: Network elements in isolation," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 114–131, Jan. 1991.

[8] Y. Jiang and Y. Liu, *Stochastic Network Calculus*. London, U.K.: Springer-Verlag, 2008.

[9] A. Burchard, J. Liebeherr, and S. Patek, "A min-plus calculus for end-to-end statistical service guarantees," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4105–4114, Sep. 2006.

[10] F. Ciucu, A. Burchard, and J. Liebeherr, "A network service curve approach for the stochastic analysis of networks," in *Proc. ACM SIGMETRICS*, 2005, pp. 279–290.

[11] C. Li, A. Burchard, and J. Liebeherr, "A network calculus with effective bandwidth," *IEEE/ACM Trans. Netw.*, vol. 15, no. 6, pp. 1442–1453, Dec. 2007.

[12] M. Fidler, "Survey of deterministic and stochastic service curve models in the network calculus," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 1, pp. 59–86, 1st Quart. 2010.

[13] M. Fidler, "An end-to-end probabilistic network calculus with moment generating functions," in *Proc. 14th IEEE IWQoS*, 2006, pp. 261–270.

[14] Y. Jiang, "A basic stochastic network calculus," in *Proc. ACM SIGCOMM*, 2006, pp. 123–134.

[15] M. Fidler, "A network calculus approach to probabilistic quality of service analysis of fading channels," in *Proc. IEEE GLOBECOM*, 2006, pp. 1–6.

[16] K. Zheng, F. Liu, L. Lei, C. Lin, and Y. Jiang, "Stochastic performance analysis of a wireless finite-state Markov channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 782–793, Feb. 2013.

[17] H. Al-Zubaidy, J. Liebeherr, and A. Burchard, "A (min, $\times$) network calculus for multi-hop fading channels," in *Proc. IEEE INFOCOM*, 2013, pp. 1833–1841.

[18] D. Wu, *"Providing quality-of-service guarantees in wireless networks*," Ph.D. dissertation, Dept. Elect. Comput. Eng., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2003.

[19] Y. Jiang and P. J. Emstad, "Analysis of stochastic service guarantees in communication networks: A server model," in *Proc. IWQoS*, 2005, pp. 233–245.

[20] Y. Gao and Y. Jiang, "Analysis on the capacity of a cognitive radio network under delay constraints," *IEICE Trans.*, vol. 95-B, no. 4, pp. 1180–1189, 2012.

[21] EEIG ERTMS User Group, "ETCS/GSM-R quality of service—Operational analysis [04E117]," Int. Union Railways, Paris, France, Tech. Rep., 2005.

[22] A. Zimmermann and G. Hommel, "Towards modeling and evaluation of ETCS real-time communication and operation," *J. Syst. Softw.*, vol. 77, no. 1, pp. 47–54, Jul. 2005.

[23] J. Xun, B. Ning, K. Li, and W. Zhang, "The impact of end-to-end communication delay on railway traffic flow using cellular automata model," *Transp. Res. C, Emerging Technol.*, vol. 35, pp. 127–140, Oct. 2013.

[24] Y. Dong, P. Fan, and K. B. Letaief, "High-speed railway wireless communications: Efficiency versus fairness," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 925–930, Feb. 2014.

[25] O. 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, May 2012.

[26] 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. 2343–2347.

[27] Q. Xu, X. Li, H. Ji, and L. Yao, "A cross-layer admission control scheme for high-speed railway communication system," in *Proc. IEEE ICC*, 2013, pp. 6343–6347.

[28] L. Zhu, F. Yu, B. Ning, and T. Tang, "Handoff performance improvements in MIMO-enabled communication-based train control systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 582–593, Jun. 2012.

[29] H. Liang and W. 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.

[30] P. Sadeghi, R. Kennedy, P. Rapajic, and R. Shams, "Finite-state Markov modeling of fading channels—A survey of principles and applications," *IEEE Signal Process. Mag.*, vol. 25, no. 5, pp. 57–80, Sep. 2008.

[31] R. Zhang and L. Cai, "Joint AMC and packet fragmentation for error control over fading channels," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3070–3080, Jul. 2010.

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

[33] H. Wang, F. Yu, L. Zhu, T. Tang, and B. Ning, "Finite-state Markov modeling for wireless channels in tunnel communication-based train control systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 3, pp. 1083–1090, Jun. 2014.

[34] X. Li, C. Shen, A. Bo, and G. Zhu, "Finite-state Markov modeling of fading channels: A field measurement in high-speed railways," in *Proc. IEEE/CIC ICCC*, 2013, pp. 577–582.

[35] S. Mao and S. S. Panwar, "A survey of envelope processes and their applications in quality of service provisioning," *IEEE Commun. Surveys Tuts.*, vol. 8, no. 3, pp. 2–20, 3rd Quart. 2006.

[36] T. Luan, X. Ling, and X. Shen, "MAC in motion: Impact of mobility on the MAC of drive-thru internet," *IEEE Trans. Mobile Comput.*, vol. 11, no. 2, pp. 305–319, Feb. 2012.

[37] R. He *et al.*, "Measurements and analysis of propagation channels in high-speed railway viaducts," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 794–805, Feb. 2013.

[38] G. L. Stber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA, USA: Kluwer, 2001.

[39] Q. Liu, S. Zhou, and G. Giannakis, "Queuing with adaptive modulation and coding over wireless links: Cross-layer analysis and design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1142–1153, May 2005.
[40] *Evolved universal terrestrial radio access (E-UTRA); Physical channels and modulation*, 3GPP TR Specification 36.211, Technical Specification Group Radio Access Network, Jun. 2008.
[41] H. Zheng and H. Viswanathan, "Optimizing the ARQ performance in downlink packet data systems with scheduling," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 495–506, Mar. 2005.
[42] C.-S. Chang, *Performance Guarantees in Communication Networks*. London, U.K.: Springer-Verlag, 2000.

Lei Lei (M'13) received the B.S. and Ph.D. degrees in telecommunications engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2001 and 2006, respectively. From July 2006 to March 2008, she was a Postdoctoral Fellow with the Department of Computer Science, Tsinghua University, Beijing. From April 2008 to August 2011, she was with the Wireless Communications Department, China Mobile Research Institute. Since September 2011, she has been an Associate Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing. Her current research interests include performance evaluation, quality of service, and radio resource management in wireless communication networks.

Jiahua Lu received the B.E. degree from Beijing Jiaotong University (BJTU), Beijing, China, in 2014. He is currently working toward the Master's degree with BJTU. His research interests lie in wireless communications for railways.

Yuming Jiang (SM'14) received the B.Sc. degree from Peking University, Beijing, China, in 1988, the M.Eng. degree from Beijing Institute of Technology, Beijing, in 1991, and the Ph.D. degree from the National University of Singapore, Singapore, in 2001. From 1996 to 1997, he was with Motorola. From 2001 to 2003, he was a Member of Technical Staff and a Research Scientist with the Institute for Infocomm Research, Singapore. From 2003 to 2004, he was an Adjunct Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore. From 2004 to 2005, he was with the Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, supported in part by the Fellowship Programme of the European Research Consortium for Informatics and Mathematics. Since 2005, he has been a Professor with the Department of Telematics, NTNU. From 2009 to 2010, he was with Northwestern University, Evanston, IL, USA. He is the first author of *Stochastic Network Calculus*. His research interests are the provision, analysis, and management of quality-of-service guarantees in communication networks. In the area of network calculus, his focus has been on developing models and investigating their basic properties for stochastic network calculus (snetcal) and, recently, on applying snetcal to performance analysis of wireless networks.

Prof. Jiang was a Cochair for IEEE GLOBECOM2005—General Conference Symposium, a General/TPC Cochair for International Symposium on Wireless Communication Systems 2007–2010, a TPC Cochair for IEEE Vehicular Technology Conference 2008, and a General Chair for the IFIP Networking 2014 Conference.

Xuemin (Sherman) Shen (M'97–SM'02–F'09) received the B.Sc. degree from Dalian Maritime University, Dalian, China, in 1982 and the M.Sc. and Ph.D. degrees from Rutgers University, Newark, NJ, USA, in 1987 and 1990, respectively, all in electrical engineering. He is currently a Professor and a University Research Chair with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, where he was the Associate Chair for Graduate Studies from 2004 to 2008. He is a coauthor/editor of six books and has authored or coauthored over 600 papers and book chapters in wireless communications and networks and control and filtering. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. Dr. Shen is an Elected Member of the IEEE ComSoc Board of Governor and the Chair of the Distinguished Lecturers Selection Committee. He is a Fellow of the IEEE, The Engineering Institute of Canada, and The Canadian Academy of Engineering and a Distinguished Lecturer of the IEEE Vehicular Technology and IEEE Communications Societies. He served as the Technical Program Committee Chair/Cochair for IEEE Infocom'14 and IEEE VTC'10 Fall, the Symposia Chair for IEEE ICC'10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, the Technical Program Committee Chair for IEEE GLOBECOM'07, the General Cochair for Chinacom'07 and QShine'06, and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications and P2P Communications and Networking. He also serves/served as the Editor in Chief of IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; a Founding Area Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, Computer Networks, and ACM/Wireless Networks; and the Guest Editor of IEEE Journal on Selected Areas in Communications, *IEEE Wireless Communications*, *IEEE Communications Magazine*, and *ACM Mobile Networks and Applications*. He was a recipient of the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo; the Premier's Research Excellence Award in 2003 from the Province of Ontario; and the Excellent Graduate Supervision Award in 2006 and the Outstanding Performance Award in 2004, 2007, and 2010 from the University of Waterloo. He is also a Registered Professional Engineer in Ontario, Canada.

Ying Li received the B.E. degree from Beijing Jiaotong University (BJTU), Beijing, China, in 2012. She is currently working toward the Master's degree with BJTU. Her research interests lie in wireless communications for railways.

Zhangdui Zhong received the B.Eng. and M.Sc. degrees from Northern Jiaotong University (currently Beijing Jiaotong University), Beijing, China, in 1983 and 1988, respectively. He is currently a Professor and an Advisor of Ph.D. candidates with Beijing Jiaotong University, where he is also the Dean of the School of Computer and Information Technology and a Chief Scientist in the State Key Laboratory of Rail Traffic Control and Safety. He is also a Director of the Innovative Research Team of the Ministry of Education and a Chief Scientist of the Ministry of Railways of China. He is an Executive Council Member of the Radio Association of China and a Deputy Director of the Radio Association of Beijing. His interests are wireless communications for railways, control theory and techniques for railways, and GSM-R system. His research has been widely used in railway engineering, such as the Qinghai–Xizang railway, the Datong–Qinhuangdao Heavy Haul railway, and many high-speed railway lines of China. He has authored/coauthored seven books, five invention patents, and over 200 scientific research papers in his research areas. Prof. Zhong was a recipient of the Mao YiSheng Scientific Award of China, the Zhan TianYou Railway Honorary Award of China, and the Top 10 Science/Technology Achievements Award of Chinese Universities.

Chuang Lin (SM'04) received the Ph.D. degree in computer science from Tsinghua University, Beijing, China, in 1994. He is currently a Professor with the Department of Computer Science and Technology, Tsinghua University. He is also an Honorary Visiting Professor with the University of Bradford, Bradford, U.K. His current research interests include computer networks, performance evaluation, network security analysis, and Petri net theory and its applications. He has authored or coauthored over 500 papers in research journals and IEEE conference proceedings in these areas and has published five books. Prof. Lin is a Senior Member of the IEEE. He served as the Technical Program Vice Chair for the 10th IEEE Workshop on Future Trends of Distributed Computing Systems and the General Chair for ACM SIGCOMM Asia Workshop 2005 and the 2010 IEEE International Workshop on Quality of Service. He has been an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the Area Editor of *Journal of Computer Networks*.