Mobile relays for enhanced broadband connectivity in high speed train systems

Elias Yaacoub\textsuperscript{a,\ast}, Rachad Atat\textsuperscript{b}, Ahmad Alsharoa\textsuperscript{c}, Mohamed-Slim Alouini\textsuperscript{c}

\textsuperscript{a} Qatar Mobility Innovations Center (QMIC), Qatar Science and Technology Park, P.O. Box 210531, Doha, Qatar
\textsuperscript{b} Department of Electrical Engineering and Computer Science, University of Kansas, KS, United States
\textsuperscript{c} Electrical Engineering Program, King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 7 June 2013
Received in revised form 3 June 2014
Accepted 7 June 2014
Available online 13 June 2014

\textbf{Keywords:}
High speed train
Content distribution
Relay
LTE
Resource allocation

\textbf{ABSTRACT}

With the introduction of wireless modems and smart phones, the passenger transport industry is witnessing a high demand to ensure not only the safety of the trains, but also to provide users with Internet access all the time inside the train. When the Mobile Terminal (MT) communicates directly with the Base Station (BS), it will experience a severe degradation in the Quality of Service due to the path loss and shadowing effects as the wireless signal is traveling through the train. In this paper, we study the performance in the case of relays placed on top of each train car. In the proposed approach, these relays communicate with the cellular BS on one hand, and with the MTs inside the train cars on the other hand, using the Long Term Evolution (LTE) cellular technology. A low complexity heuristic LTE radio resource management approach is proposed and compared to the Hungarian algorithm, both in the presence and absence of the relays. The presence of the relays is shown to lead to significant enhancements in the effective data rates of the MTs. In addition, the proposed resource management approach is shown to reach a performance close to the optimal Hungarian algorithm.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

With the significant increase in the number of Internet users, wireless communications became integrated in people’s lives. Personal wireless devices such as laptops, Personal Digital Assistant (PDA), and smart phones are becoming widely used by users whether they are at homes, in their offices, inside their vehicles, on a train or a bus. In the past years, the passenger transport industry has witnessed a high demand for broadband services to ensure the safety of people and trains on one hand and to provide passengers with Internet access on the other hand [1]. Information between terrestrial control centers and trains, such as information on the location of the train, its schedule, its state, level crossings, permitted speed, etc., is usually exchanged through the Global System for Mobile communications-Railway (GSM-R) network to ensure that the railway service is under secure conditions [2]. In addition, since most of the journeys in high speed trains can last for several hours, passengers may want to have access to the Internet to be able to browse a website, read/send emails, perform real-time multimedia streaming, etc. [3]. All these services necessitate a wireless communication channel with large bandwidth and wide coverage [2]. The implementation of wireless communication systems has attracted interest from both the railway industry and the communication companies, which in turn are investigating new network architectures to provide the passengers with high speed mobile data services [4]. GSM-R can provide a maximum
data rate of 200 kbps and is only used to exchange train control information rather than passengers’ communication and thus, GSM-R cannot satisfy the needs of high speed data transmissions [5]. High speed mobile data services are provided by High Speed Downlink Packet Access (HS-DPA) that is able to achieve a data rate of 28 Mbps with a 5 MHz bandwidth. In addition, the Long Term Evolution Network (LTE) is being extensively deployed for cellular communications since it meets increasing bandwidth demand with high spectrum efficiency and latency. LTE can achieve a higher data rate of up to 300 Mbps with 20 MHz of channel bandwidth [5]. There exist various technologies for train-to-land connections, among these technologies, the satellite which is suitable for tracks without obstructions, WiFi which is suitable for train journeys with multiple stops, 2G/EDGE which is suitable for low bandwidth applications, and 3G which is suitable in urban areas. However, LTE is best suited for high speed broadband data [6]. An extension to LTE, called LTE-R (LTE-Railway), which is based on the standard of LTE and SAE (System Architecture Evolution), is being investigated as the next generation wireless communication system for high speed railways since it was shown to provide good performance with advanced channel estimation and disperse deployed antennas on the train [7]. However, many problems and challenges arise despite the wireless broadband technology in use. First, when the Mobile Terminal (MT) communicates directly with the Base Station (BS), it will experience a severe degradation in the Quality of Service (QoS) since the wireless signal has to travel through the train, and penetrate through the metalized windows, the fact that dramatically reduces and weakens the wireless link quality [5]. Multimedia applications such as video streaming require a high bandwidth in order to be delivered with high QoS to end users. Since the train can be perceived as a moving mobile network and since the users are moving fast, a reliable direct link with the outside cellular network can be hard to establish [3,8].

In line with standardization efforts [9], we suggest in this paper the use of efficient radio resource management with an LTE multihop relay-based network architecture, where a relay is installed on top of each car of the train (on the ceiling), and the closest wireless BS in the vicinity of the train communicates with the relays using LTE technology. Then, each relay communicates with the MTs that are located inside the train car, also using LTE. This network approach aims at enhancing the perceived data rates and QoS of the MTs compared to the case where the BS communicates directly with the MTs in the train. Thus, the BS does not need to communicate with the hundreds of passengers in the train which reduces radio resource management control significantly. This approach will help in avoiding the radio signal propagation losses and low QoS, and maintaining a stable high speed wireless link between the relay and the MTs inside the train car.

The paper is organized as follows. Related work is reviewed in Section 2. The system model is presented in Section 3. The adopted propagation models are described in Section 4. Data rate calculations with the proposed relay-based approach are presented in Section 5. The problem formulation and the proposed heuristic solution based on LTE radio resource management are presented in Section 6. Simulation results are studied and analyzed in Section 7. Section 8 outlines some limitations of the proposed approach and indicates directions for future research. Finally, conclusions are drawn in Section 9.

2. Related work

Relays have been widely investigated in the literature because they can minimize the total power consumption of the network nodes, maximize the network lifetime, extend the coverage and expand the capacity in wireless systems [10]. Relays for the purpose of reducing the energy consumption in sensor networks and cooperative ad-hoc networks are investigated in [11–16]. In [11], the authors propose a new scheme to be used in wireless sensor networks which are usually composed of a number of microsensors which have limited battery power. In the scheme of [11], one cluster head is elected, and it sends the data to the BS via a relay node, and thus, energy can be saved. An algorithm to select the best relays with minimum power consumption that would form cooperative links to establish a route from source to destination is proposed in [12]. For the cooperative link, total power consumption which is taken to be the summation of the transmit power of the source node and the transmit power of the relay node is minimized subject to a target bit error rate (BER). Then, the relay node that minimizes the total power consumption is selected and a one hop cooperative route from source to destination is established. An optimal power allocation in the case of multiple relays is suggested in [13]. The formulated optimization problem is based on maximizing the network lifetime at each transmission stage where channels are slowly varying over time by applying the optimal power allocation strategy. In [15], the authors minimize the average energy consumed in the network to transmit a message from source to destination via intermediate cooperative relays subject to a target outage probability. They propose an algorithm equivalent to shortest path computation with link cost being the energy consumed of transmitting a message from one node to another. In [16], suboptimal algorithms are proposed in order to increase the device lifetime by exploiting cooperative diversity and taking both location and energy advantages under the BER constraint.

The previous references investigated relaying in a relatively low mobility scenario. However, relaying in scenarios with higher mobility, such as vehicular networks and railroad networks, is more relevant to the scope of this paper. Cooperative wireless communication systems employed in vehicular networks have been investigated in [17–21]. Vehicles traversing a route have limited time to download a file or to access the Internet as they are moving fast and the Access Point (AP) cannot cover the whole path, which means vehicles would have intermittent connectivity. In [17,18], the authors present a cooperative Automatic Repeat-reQuest (ARQ) where, after leaving the AP coverage, vehicles communicate between each other to exchange the packets that were lost during the transmission from the AP to the vehicle nodes. In this way, retransmissions of packets can be avoided and thus packet losses will
be decreased improving the throughput and transfer delay. In [18], cars broadcast HELLO messages to know about the presence of other neighbor cars and to notify other nodes that they need to act as cooperators. A list of cooperators is contained in the HELLO message. After a vehicle finishes downloading data from the AP, it identifies the packets lost and requests them from other vehicles.

Railroad networks correspond to very high speeds where additional challenges arise compared to vehicular networks. Consequently, corresponding propagation characteristics need to be investigated. Therefore, in [1], the authors study the path loss characteristics for intra- and inter-car communications in train systems. For intra-car communications, they use a 2 GHz band with access points (APs) installed either at the center of the car (on the ceiling) or on the sidewalk. For inter-car communications, they use a frequency band of 5 GHz to study the path loss characteristics between adjacent APs. Different techniques were proposed in the literature to ensure broadband connectivity for train passengers. In [2], to exchange information between terrestrial control centers and the train, WiFi networks are placed in locations where the train stops for long periods of time, and that is to allow the discharge of certain amount of information. A broadband communications manager is developed to assign communication requests for terrestrial applications to communicate with the train through the WiFi networks. In [3], a hierarchical architecture is developed to provide passengers with communications and entertainment services through an on-board Local Area Network (LAN), that connects the end user to the outside world. Wireless Asynchronous Transfer Mode (WATM) is a high capacity mobile network used to provide communications between the high speed train and the land station, as it can reach a transmission rate of 155 Mbps or higher. In [4], Ethernet is used to deliver broadband services to fast moving users. The moving train communicates with antennas on the railroad via WiFi or WiMAX. To guarantee a stable connection all the time, moving tunnels between train and Service Gateway (SGW), that acts as uplink towards the Internet, provide guaranteed overall connection when the train moves from one access network to the next. In order to provide end users with good video quality, Video Adapting Software (VAS) packages are placed along the railway to transform video files into a different format and shape that meets the needs of users and the device capabilities. In [5,8], a two-hop network is suggested to reduce signal propagation losses and to provide a stable high speed wireless link between MTs and in-train APs. The MT communicates with an AP installed in the train, then the data received by the AP is aggregated and sent to the BS via an antenna installed on top of the train (on the ceiling). Wireless LAN (WLAN) is used for in-train coverage, while Radio-over-Fiber (RoF) is used to provide broadband wireless access for high speed trains, where antennas are placed along the railway and connected to one control center via a fiber ring. In [6], LTE is viewed as the best candidate for train-to-land communications since it meets the increasing bandwidth demand with high spectrum efficiency and low latency for high speed broadband services. In [7], LTE-R is viewed as a promising solution for the next generation broadband mobile communication systems of high speed train systems.

The novelty in this paper consists of proposing an efficient two-hop LTE radio resource management approach using LTE relays placed on top of each train car, in line with the architecture discussed in [9]. Adequate propagation models for high speed rail scenarios are used in the simulations. The contributions of this paper compared to the existing literature can be summarized as follows:

- APs are placed inside the train car, not outside on the terrestrial control center as in [2]. Furthermore, they ensure user connectivity as the train moves at high speed, not when it stops. Each AP, placed on the ceiling of a train car, has an antenna outside the train, communicating with the LTE BSs, and another antenna inside the train, communicating with the MTs.
- Compared to other papers suggesting two-hop communications, the proposed approach does not require the installation of specific antennas on the railroads as in [4,5,8]. Furthermore, in this paper, we investigate the use of the state-of-the-art wireless communications technology, LTE, on the connections between the relays and the BS. No additional antennas need to be installed outside the train along the railroad track, and the LTE BSs can be co-located with GSM-R BSs, already deployed along most of the railroads.
- As for papers that discuss LTE and LTE-R [6,7], they do not suggest the use of relays. Instead, they mainly focus on LTE-R as a replacement of GSM-R. In this paper, detailed performance evaluation with LTE resource allocation in the presence and absence of relays is presented. Furthermore, LTE is not considered as a GSM-R replacement in terms of exchanging control information, although it can easily assume this role. However, it is investigated here in order to ensure broadband connectivity for train passengers, and LTE BSs can be co-located with GSM-R BSs, where GSM-R can be used for railroad control and LTE can be used to ensure passenger connectivity. The novelty in the proposed approach consists of using LTE relays on top of each train car, along with efficient radio resource management, in order to provide enhanced passenger connectivity.

3. System model

A high speed moving train is considered. Cellular coverage inside the train is ensured by LTE BSs deployed parallel to the train path. These BSs could be co-located with GSM-R BSs. Typical GSM-R deployments consist of having a separation distance $d_{BS}$ between BSs on the order of 7–15 km along the railroad. Although in traditional urban LTE deployments BSs are deployed with considerably shorter inter-site distances, this separation is suitable for railroad networks: the BSs are covering only the railway system, they are deployed along the railroad path, and they are equipped with directive antennas concentrating the radiated power in the direction of the rail track. In addition, each BS site can accommodate two sectors, so the actual cell radius would be half of the inter-site distance, as shown in Fig. 1. However, distinction should be made between GSM-R used for railway control, which is out of the scope of this paper, and LTE, which is used to ensure cellular connectivity to passengers inside the train.
The system model is depicted in Fig. 2. The train consists of a number $K$ of train cars. On top of each car, a mobile relay (MR) is fixed in the ceiling. The relay performs heterogeneous communications using two antennas (or two sets of antennas in case multiple-input multiple-output (MIMO) communication is used, which could be an interesting future extension of this work): one antenna located outside the train car, used for communication with the BS on the long range (LR) LTE links, and another antenna inside the train car, used to communicate with the MTs inside the train using short range (SR) LTE communications. In a given car, relay $k$ serves a number $M_k$ of MTs belonging to the train passengers. Omnidirectional antennas are considered at the MRs and MTs.

LTE is based on Orthogonal Frequency Division Multiple Access (OFDMA). The available spectrum is divided into resource blocks (RB) consisting of 12 adjacent subcarriers, allocated in a 0.5 ms time slot. The shortest assignment unit consists of two consecutive slots, i.e., for a duration of 1 ms, which is the duration of one transmission time interval (TTI) [22,23]. Each subcarrier has a bandwidth of $W_{\text{sub}} = 15$ kHz, such that the bandwidth of an RB is $W_{\text{RB}} = 180$ kHz [22]. In addition, we consider that a 20 MHz LTE bandwidth is subdivided into three orthogonal blocks: a block of $W_{\text{LR}} = 10$ MHz LTE bandwidth on the LR (at the BS), and two blocks of $W_{\text{SR}} = 5$ MHz on the SR, with these two blocks used at consecutive MRs inside the train to avoid interference. They are reused at MRs further away. We assume that interference can be neglected in this case, due to the relatively low transmission power of MRs and due to the separation, consisting of multiple metallic walls, between the MRs using the same bandwidth, which leads to significant attenuation. In LTE, a bandwidth of 5 MHz is subdivided into $N_{\text{RBs}} = 25$ RBs, whereas a bandwidth of 10 MHz is subdivided into $N_{\text{RBs}} = 50$ RBs, with each RB consisting of 12 consecutive subcarriers [22,23]. With LTE-Advanced (LTE-A), additional bandwidth can be used with carrier aggregation (up to 100 MHz), in order to increase both the LR and SR bandwidths. The approach presented in this paper is applicable both to the case of a single carrier where the bandwidth is subdivided into multiple blocks, and to the case of carrier aggregation.
In this paper, the presented approach is applicable to both the downlink (DL) and uplink (UL) directions. In the DL, data is sent from the BS to the MRs or MTs. In the UL, data is sent from the MTs or MRs to the BS. Although the results of Section 7 focus on the DL as an example, the proposed approach can be implemented similarly on the UL. Whenever the formulations are applicable to multiple scenarios, we use indices r and t to avoid repetition. We use r to indicate a receiver, which could be an MR or an MT in the DL, or an MR or the BS in the UL. We use t to indicate a transmitter, which would be a BS or an MR in the DL, or an MT or an MR in the UL.

3.1. Data rates

Letting $I_{\text{sub},r}$ be the set of subcarriers allocated to receiver r, $I^{(t)}_{\text{RB}}$, the set of RBs allocated to receiver r, $N^{(t)}_{\text{RB}}$, the total number of RBs at transmitter t, $P^{(t)}_{\text{tx}}$, the total power transmitted by transmitter t over subcarrier x, and $P^{(t)}_{\text{max}}$ the maximum transmission power of transmitter t, then the transmission of transmitter t leads to an achievable rate at receiver r $R_{t,r}$ that is given by

$$R_{t,r} = \frac{P^{(t)}_{\text{tx}}}{N^{(t)}_{\text{RB}}},$$

(1)

where $\Gamma_{t,r,x}$ is the signal to interference plus noise ratio (SINR) on the link between transmitter t and receiver r over subcarrier x, and $B^{(t)}_{\text{sub}}$ is the subcarrier bandwidth. It is expressed as

$$B^{(t)}_{\text{sub}} = \frac{B^t}{N^{(t)}_{\text{sub}}},$$

(2)

with $B^t$ the total usable bandwidth at transmitter t, and $N^{(t)}_{\text{sub}}$ the total number of subcarriers at transmitter t.

The SINR $\Gamma_{t,r,x}$ is given by

$$\Gamma_{t,r,x} = \frac{P^{(t)}_{\text{tx}}G_tG_r\xi^x_{t,r}}{I_{t,r} + \sigma^2_{t,r}},$$

(3)

where $P^{(t)}_{\text{tx}}$ is the transmission power of transmitter t over subcarrier x, $H^x_{t,r}$ is the channel gain between transmitter t and receiver r over subcarrier x, $\sigma^2_{t,r}$ is the noise power over subcarrier x in receiver r, and $I_{t,r}$ is the interference on subcarrier x measured at receiver r. The antenna gains of the transmitter and receiver are denoted by $G_t$ and $G_r$, respectively.

4. Channel model

In this paper, three types of communication links can be noted: BS–MR links, MR–MT links, and BS–MT links. The first two types correspond to the scenario with relays, whereas the last type corresponds to the traditional approach. Each of these link types should be described by a dedicated propagation model, since the propagation characteristics differ between each of them.

4.1. Path Loss analysis in the scenario with relays

The Path Loss (PL) used in the backhaul link between the BS and MR (BS–MR link) follows the D1 Line of Sight (LoS) propagation model described in the WINNER II specifications [24]. Thus, PL between the BS and MR k is denoted by $\text{PL}_{\text{BS-MR}_k}$ and expressed as

$$\text{PL}_{\text{BS-MR}_k} = 44.2 + 21.5 \log_{10}(d_{\text{BS-MR}_k}) + 20 \log_{10}(f[gHz]/5),$$

$$10 \text{ m} < d_{\text{BS-MR}_k} < d_{BP} \cdot 10.5$$

$$+ 40.0 \log_{10}(d_{\text{BS-MR}_k}) - 18.5 \log_{10}(f[GHz]) / 5,$$

$$d_{BP} < d_{\text{BS-MR}_k} < 10 \text{ km},$$

(4)

where $d_{\text{BS-MR}_k}$ is the distance from the BS to the kth MR in [m], f is the center frequency, $h_{\text{BS}}$ and $h_r$ are the height of the BS and the train in [m], respectively, and $d_{BP} = 4h_{\text{BS}}h_{r}f(s)^{1/2}$ [Hz/c] is the breaking point distance value, where $c = 3 \times 10^8$ m/s is the speed of light in vacuum.

The PL used in the access link between MR k and MT m (MR–MT link) inside the corresponding kth wagon follows the A1 indoor LoS propagation model described in the WINNER II specifications [24]:

$$\text{PL}^{(k)}_{\text{MR}k-\text{MT}m}(d_{\text{MR}k-\text{MT}m}) = 46.4 + 18.7 \log_{10}(d_{\text{MR}k-\text{MT}m})$$

$$+ 20 \log_{10}(f[GHz]/5),$$

(5)

where $d_{\text{MR}k-\text{MT}m}$ is the distance from MR k to MT m in [m].

4.2. Path Loss analysis in the scenario without relays

In this section, the PL of the direct link between BS and MT (BS–MT link) without a relay under outdoor–indoor propagation scenario is given by [24]

$$\text{PL}_{\text{BS-MT}}(d_{\text{BS-MT}}) = \text{PL}_{\text{BS-MR}}(d_{\text{BS-MR}}) + \text{PL}_{\text{MR-MT}}(d_{\text{MR-MT}}) + \text{PL}_{\text{outer}},$$

(6)

where $d_{\text{BS-MT}}$ are the distance between the BS and closest point of the train wall to the MT and the distance from train wall to the MT in [m], respectively. $\text{PL}_{\text{outer}} = W_{\text{e}} + W_{\text{c}}(1 - \cos(\theta))^2$ is the path loss through the outer wall, where $W_{\text{e}}$ is the loss through the train wall for the perpendicular penetration while $W_{\text{c}}$ is the loss through the train wall for the parallel penetration, and $\theta$ is the angle between the normal of the train wall and the outgoing (incoming) ray.

The total PL at a given distance for both relay and direct links is given as follows:

$$\text{PL}^{(m)} = \begin{cases} \text{PL}_{\text{relay}}^{(m)} = \text{PL}_{\text{BS-MR}}(d_{\text{BS-MR}}) + \text{PL}_{\text{MR-MT}}(d_{\text{MR-MT}}), & \text{scenario with MRs,} \\ \text{PL}_{\text{direct}}^{(m)} = \text{PL}_{\text{BS-MT}}(d_{\text{BS-MT}}), & \text{for the scenario without MRs,} \end{cases}$$

(7)

where the distance is a function of the time and train speed. All the aforementioned formulas are generalized for the frequency range 2–6 GHz [24].

4.3. Channel gain calculation

The channel gain in dB between transmitter t and receiver r over subcarrier x is a function of PL, shadowing, and fading as follows:

$$H^{(x)}_{t,r,db} = -\text{PL}^{(x)}_{t,r} - \xi_{t,r} + 20 \log_{10}(P^{(x)}_{t,r}),$$

(8)
where \( \xi_{t,r} \) is the log-normal shadowing between the transmitter and receiver with a standard deviation \( \sigma_x \), and \( r_{t,r}^{(m,x)} \) corresponds to fast fading between the transmitter and receiver. 

In the channel model, spatial shadowing correlation is taken into account since shadow fading values depend on the fixed location of obstacles [25]. Spatial correlation can be described as a measure of how fast the local mean power evolves as the train moves along a certain route [26]. We will apply the correlated shadowing model of [26,27] where the shadowing correlation is expressed as

\[
\Lambda_{\xi}(\Delta d_r, \Delta d_t) = \exp \left( -\frac{\Delta d_r + \Delta d_t}{d_{cor}} \ln 2 \right), \tag{9}
\]

where \( \exp(\cdot) \) denotes the exponential function, \( \ln(\cdot) \) the natural logarithm, \( \Delta d_r \) and \( \Delta d_t \) represent the movements of receiver and transmitter, respectively, and \( d_{cor} \) is the decorrelation distance (equal to 20 m in the vehicular test environment; see [28]). As the train moves, the model of (9) can be applied to determine the correlation between shadowing values at the different train positions, in addition to the correlation between shadowing values of the train on the link with the BS (by setting \( \Delta d_r = 0 \) since the BS is fixed).

Consequently, we can use (9) to determine the shadowing values as follows:

\[
\xi_{t,r}(d_r + \Delta d_r, d_t + \Delta d_t) = \xi(d_r, d_t) \Lambda_{\xi}(\Delta d_r, \Delta d_t) + (1 - \xi(\Delta d_r, \Delta d_t)) \xi_{t,r}^{new}, \tag{10}
\]

where the shadowing value at positions \( (d_r + \Delta d_r, d_t + \Delta d_t) \) has a component equal to the shadowing value at positions \( (d_r, d_t) \), with the amount of similarity determined by \( \Lambda_{\xi}(\Delta d_r, \Delta d_t) \), in addition to an independent shadowing component corresponding to the new location \( \xi_{t,r}^{new} \). In our system model, we assume that \( \Delta d_r = 0 \) since the BS has a fixed position.

As for fast fading, it is a topic of ongoing research, in order to derive accurate models for railroad scenarios. For example, in [29], empirical measurements between a GSM-R BS and an antenna placed on the train showed that a Rician distribution is suitable to model fast fading effects. However, the Rician parameter \( K \) varies along different sections of the track, and it was found in [29] that most values occur between \( K = 0 \) and \( K = 4 \). Rician fading is suitable in scenarios where a strong dominant signal component is present, e.g. as in LoS scenarios. This would correspond to the BS–MR links and MR–MT links. In the scenario without relays, an LoS is not achievable between the MTs and BS, and thus Rayleigh fading would be more suitable. Rayleigh fading is a special case of Rician fading with the Rician parameter \( K \) set to \( K = 0 \). It is known to lead to deeper fades and worse propagation scenarios [30]. Since the value of the Rician parameter generally varies along the train track, we will consider Rayleigh fading in this paper. It will correspond to a reasonable representation for the scenario without relays, but to a worst case scenario to the case with relays. Although this will lead to a lower bound on the performance for the proposed approach compared to the case without relays, significant gains are still reached as shown in Section 7. A block fading model is considered, where the fast fading remains constant for a fixed time \( T_{dec} \) which is the channel de-correlation time. Then the channel conditions change and remain constant for another \( T_{dec} \), and so on.

5. Effective rate calculations with the relay-based approach

We denote by \( \mathcal{R} \) the set of relays, and by \( \mathcal{M}_k \) the set of MTs served in train car \( k \), with \( |\mathcal{M}_k| = M_k \), where \( |\cdot| \) represents set cardinality. In addition, we denote by \( R_{t,r} \) the transmission rate on the LR links to receiver \( r \), and by \( R_{S,k} \) the achievable transmission rate on the SR links between relay \( k \) and MT \( j \).

The time needed to transmit (receive) one data bit to (from) any MT \( j \in \mathcal{M}_k \) in the DL (UL) is given by

\[
D_{\text{relay},jk}^{\text{bit}} = \frac{1}{R_{L,k}} + \frac{1}{R_{S,\bar{k}j}}, \tag{11}
\]

where the first term corresponds to the time needed by the relay to receive (send) the data bit from (to) the BS and the second term corresponds to the transmission from (to) relay \( k \) to (from) the MT \( j \in \mathcal{M}_k \) in the DL (UL). With each MR acting as a femto BS inside its train car, control information can be exchanged separately on two levels: on the SR inside the train cars, the MR handles control communications with the MTs similarly to a femto BS; on the LR outside the train car, control information is exchanged between the BS and the MR as it would be between the BS and a normal mobile user. This allows better mobility management, as explained in Section 8.1.

Considering data communications, the effective data rate of MT \( j_k \in \mathcal{M}_k \) when relays are used can be expressed as the inverse of (11), or

\[
R_{\text{eq},jk} = \frac{R_{L,k} \cdot R_{S,\bar{k}j}}{R_{L,k} + R_{S,\bar{k}j}}, \tag{12}
\]

The expression in (12) implicitly assumes that relay \( k \) is only relaying the data of MT \( j_k \). When multiple MTs in the train car served by relay \( k \) are communicating simultaneously with the BS, the LR rate \( R_{L,k} \) of relay \( k \) would be used to carry the aggregated information corresponding to all MTs in its train car. Thus, it should be replaced by \( \alpha_{jk} R_{L,k} \), where \( \alpha_{jk} \) corresponds to the fraction of the data dedicated to MT \( j \) in the aggregated data carried by relay \( k \). Hence, we have \( \sum_{j_k} \alpha_{jk} = 1 \). Consequently, (12) becomes

\[
R_{\text{eq},jk} = \frac{\alpha_{jk} R_{L,k} \cdot R_{S,\bar{k}j}}{\alpha_{jk} R_{L,k} + R_{S,\bar{k}j}}, \tag{13}
\]

For example, in order to ensure a fair allocation of resources, \( \alpha_{jk} \) can be set to

\[
\alpha_{jk} = \frac{R_{S,\bar{k}j}}{\sum_{j_k \in \mathcal{M}_k} R_{S,\bar{k}j}}. \tag{14}
\]

It should be noted that (13) is general and applicable to both UL and DL communications. In addition, in the absence of relays, we simply have \( R_{\text{eq},jk} = R_{L,k} \).

6. Problem formulation and proposed solution

This section presents the problem formulation to maximize the number of satisfied users, in addition to a low
complexity heuristic solution that leads to good performance.

6.1. Problem formulation

The objective is to serve the largest number of mobile users in the high speed train. We consider that each user $j_k$ needs to achieve a target data rate $R_{th,j_k}$ in order to satisfy its QoS requirements. This threshold or target data rate can be set for the UL, for the DL, or for both (in the case of certain applications, e.g. mobile video chat). Therefore, we define an indicator variable, $\delta_{j_k}$, as follows:

$$\delta_{j_k} = \begin{cases} 1 & \text{if } R_{eq,j_k} \geq R_{th,j_k}, \\ 0 & \text{if } R_{eq,j_k} < R_{th,j_k}. \end{cases}$$

In other words, $\delta_{j_k}$ is set to one if user $j_k$ is successfully served, and it is set to zero otherwise. Consequently, the problem can be formulated as follows:

$$\max_r \ p_{rx} \sum_{k \in \mathcal{K}} \sum_{j_k \in \mathcal{R}_k} \delta_{j_k},$$

subject to:

$$\sum_{x=1}^{N_r} p_{tx}^{x,t} \leq p_{t,\max} \ \forall t.$$  \hspace{1cm} (17)

The maximization is a mixed integer program, since the variables $\delta_{j_k}$ are binary (zero-one), whereas the transmit powers $p_{tx}^{x,t}$ are non-integers. In Section 6.2, we present a low complexity heuristic solution for the above problem, using a two step LTE radio resource management approach.

6.2. Low complexity LTE radio resource management solution

In this section, we present the proposed low complexity LTE radio resource management approach. We denote by $N_r$ the number of receivers, and by $R_{eq,j}^{(r)}$ the achievable rate on the link between a transmitter $r$ and a receiver $j$ over RB $y$, which is the sum of the rates on the individual subcarriers that constitute RB $y$. The proposed algorithm is applicable on the three types of links studied in this paper: BS–MR, MR–MT, and BS–MT links. In the scenario with relays, we have a two-hop resource allocation approach, where the algorithm is applied on the BS–MR links and MR–MT links in each car. The details are presented in Algorithm 1, where we consider that one RB is allocated to each transmitter–receiver link. In the DL, RBs are allocated to receivers. In the UL, they are allocated to transmitters. The notation in Algorithm 1 corresponds to the DL scenario. This approach allocates RBs to receivers in a way to maximize performance by selecting the best RB for each receiver. To apply it in the UL, the same algorithm is implemented with the index $r$ replaced by $t$.

6.3. Comparison to the Hungarian algorithm

We compare the proposed algorithm to the Hungarian method leading to the optimal solution of the assignment problem [31]. It allows assigning $n$ RBs to $n$ receivers (MRs or MTs, depending if LR or SR LTE links are considered) in an optimal cost minimizing solution. The “cost” considered

Algorithm 1: LTE resource allocation with unicasting

1: Initially, assume equal power transmission over the subcarriers (this step is needed to compute the data rates required in Step 3 of the algorithm):

$$P_{tx}^{r} = \frac{P_{t,x}^{r,\max}}{N_{\text{sub}}}.$$  \hspace{1cm} (18)

2: while RBs are available and all receivers have not been assigned an RB do

3: Find the pair $\{\text{Receiver } r^*, \text{RB } y^*\}$ such that:

$$\{r^*, y^*\} = \arg \max_{r,y} R_{eq,j_k}^{(r)}.$$  \hspace{1cm} (19)

4: Mark RB $y^*$ as occupied and

5: Mark receiver $r^*$ as served

6: Set $R_{t,x} = R_{eq,j_k}^{(r^*)}$

7: Repeat (19) for the remaining RBs and receivers

8: until all receivers are served or all RBs are allocated

9: Divide the transmitter power equally over the allocated RBs and subcarriers and update the data rates accordingly (i.e. the power is subdivided equally over the allocated subcarriers, which do not necessarily correspond to all available subcarriers as initially assumed in Step 1)

10: end while

11: in our implementation is the transmit power to achieve the target rate. Hence, if a receiver has a higher channel gain on a given RB, it will need less transmit power to reach $R_{th}$ (as opposed to Algorithm 1 which uses equal power allocation). The remaining power can be used to serve more receivers. Thus, this approach would maximize the number of served receivers. However, in the scenario of this paper, the number of receivers (whether relays or MTs) is generally different than the number of RBs (whether at the BS or at the relays). Thus, in order to solve the Hungarian algorithm, we assume that the assignment matrix is expanded to become of size $max(N_{RB}^{(r)}, N_r) \times max(N_{RB}^{(r)}, N_r)$. Very high values for the power are entered in the fictitious entries in order to lead to dummy assignments that do not affect the actual resource allocation of real RBs to real receivers. For example, the power can be set to $10P_{t,\max}$ (to avoid using $+\infty$ values in the Matlab simulations). This will lead to the allocation of $\min(N_{RB}^{(r)}, N_r)$ RBs to the $\min(N_{RB}^{(r)}, N_r)$ receivers consuming the least power in order to achieve their target rates. This will allow serving the largest number of MTs with the available transmit power by reaching an optimal allocation.

The complexity of the Hungarian algorithm is known to be $O(n^3)$, but it can be implemented with a complexity of $O(n^2)$ [31]. In the scenario of this paper, this corresponds to $O((max(N_{RB}^{(r)}, N_r))^2)$.

6.4. Complexity analysis of the proposed algorithm

Algorithm 1 allocates each RB after performing a linear search on the receivers and RBs in order to find the receiver–RB pair that maximizes the throughput. Hence, the complexity to allocate the first RB is $O(N_{RB}^{(r)}N_r)$, the complexity to allocate the second RB is $O((N_{RB}^{(r)} - 1)(N_r - 1))$. The complexity to allocate the following RBs is $O(N_r - 1)$, since the algorithm will search for the remaining $N_r - 1$ receivers among the allocated receivers. Therefore, the complexity of the algorithm is $O(N_r(N_r - 1))$, which is less than $O(n^2)$, the complexity of the Hungarian algorithm. This is because we do not have enough RBs to allocate as many RBs as receivers, and we need to use the second step of the algorithm to allocate the remaining RBs to the remaining receivers.
1), and so on. Consequently, the total complexity of the algorithm is

$$
\theta \left( N_{\text{RB}}^{(i)} N_r + (N_{\text{RB}}^{(i)} - 1)(N_r - 1) + (N_{\text{RB}}^{(i)} - 2)(N_r - 2) \right)
\times \left[ N_r - (\min(N_{\text{RB}}^{(i)}, N_r) - 1) \right]
\leq \Theta \left( N_{\text{RB}}^{(i)} N_r \left( \min(N_{\text{RB}}^{(i)}, N_r) - 1 \right) \right)
\leq \Theta \left( \max(N_{\text{RB}}^{(i)}, N_r) \right). \tag{20}
$$

Consequently, the proposed algorithm has a lower complexity than the Hungarian Algorithm, and thus could be easily implemented in real-time. The performance comparison in terms of the number of served MTs is presented in Section 7.

### 7. Results and analysis

In this section, we compare the performance of the scenarios with and without relays. In the simulations, we assume that a high speed train moves along a straight track with a speed of 350 km/h. The train consists of $L = 10$ cars each with 10 m length, 5 m width, and 2.5 m height. A relay is placed on the middle of the ceiling of each car. We assume that accurate channel state information can be available at the transmitter. At high speeds, this can be achieved by using, for example, the predictor antenna method described in [32]. We consider that BSs use a directional antenna of gain $G_{BS} = 14$ dBi. We assume that the relays and MTs are equipped with omnidirectional antennas and hence $G_{MR} = G_{MT} = 0$ dB.

The time where fading is considered constant is taken to be $T_{\text{dec}} = 1$ ms. This is in line with the guidelines discussed in [33, (Section V–F)] based on channel coherence time. In fact, at a speed of 350 km/h and a carrier frequency of 2.6 GHz, this would correspond to a channel coherence time of around 1.18 ms. In addition, given the fact that the lowest granularity in LTE resource allocation takes place at 1 ms TTIs, it would be logical to select $T_{\text{dec}} = 1$ ms. The simulation parameters are summarized in Table 1. The simulations are implemented in Matlab, where a Monte-Carlo-based analysis has been carried out, by averaging the results over 1000 iterations.

Fig. 3 shows the number of served MTs versus the distance to the BS for $R_{\text{th}} = 0.5$ Mbps, considered to be the same for all MTs. The distance is measured between the nearest BS and the first car in the train. Clearly, the presence of mobile relays (MRs) leads to significant enhancements compared to the case without MRs. In addition, the proposed algorithm leads to results that are very close to the optimal solution obtained by the Hungarian algorithm. From this figure, it can be seen that with the proposed algorithm, around 200 passengers can be served simultaneously in the train with an inter-BS distance $d_{\text{BS}} = 6$ km in the model of Fig. 1.

Fig. 4 shows that as $R_{\text{th}}$ increases, the number of MTs that can be served simultaneously decreases, as expected. In addition, the closer the train to a neighboring BS, the better the performance. The proposed low complexity algorithm performs very closely to the optimal Hungarian algorithm in all scenarios. Furthermore, the scenario with MRs leads to significantly better performance than the scenario without MRs, up to a certain limit of $R_{\text{th}}$, after which the number of served MTs decreases rapidly. This limit is around 0.8 and 0.5 Mbps for a distance of 100 and 2000 m to the BS, respectively.

In practice, different users might be requesting different services. For example, lower $R_{\text{th}}$ might be required for users performing web browsing, compared to users performing video streaming. Thus, it is unlikely that all passengers will be requesting simultaneously a high $R_{\text{th}}$, which would allow serving a larger number of users in practice.

### 8. Limitations and extensions

In this section, we describe certain limitations of the approach presented in this paper, and outline some future research directions.
8.1. Handover issues

Handover can be a significant problem in high speed train networks. In fact, in the case without MRs, when the train moves from the coverage of one BS to that of another BS, a large number of MTs have to be handed over in a very short time. When MRs are deployed on the train cars, the process is simplified by performing group handover: the MTs served by a single MR can be considered as a single group, which limits the probability of handover failures [34]. Furthermore, considering MRs as femto BSs, each in its respective car, would mean that MTs remain under the coverage of the same femto BS, unless they move from one car to another, which would happen at low mobility (pedestrian speed) and would not pose challenges. However, mobility support and mobility management for MRs remain an open challenge in this case, as they will have to be handed over from one BS to another as the train moves along the track. Several handover options with moving relays were discussed and compared in [9], where several challenging problems are highlighted as needing further investigation.

8.2. Doppler effect

Doppler shift is an important problem in high speed cellular communications. It also affects the handover process. The Doppler effect was not considered in this paper. However, efficient techniques to overcome this problem and perform LTE handover in high speed railway systems have been recently investigated [35]. Furthermore, the Doppler effect is being mitigated for LTE communications occurring at higher speeds than in the scenario of this paper, e.g., see [36,37]. In addition, there is no significant Doppler effect in the MR–MT links: their relative movement is too small since both are inside the train car. Furthermore, the Doppler in the BS–MR links is easier to mitigate in the railway communication scenario than in other high speed scenarios, since the speed and track of the train are known and predictable, in addition to the locations of the BSs along the rail track. Thus, advanced techniques can be applied both at the BSs and relays to reduce the impact of the Doppler shift. The effect of the Doppler shift will be more significant on the BS–MT links in the absence of relays. Therefore, neglecting it in this paper is in favor of the traditional scenario. Taking it into account would increase the gains of the proposed method in the presented results.

8.3. Radio resource management

In this paper, we considered a 20 MHz system bandwidth that is subdivided into 100 RBs. In loaded scenarios where up to 250 MTs are assumed to be accessing the network simultaneously, allocating no more than one RB per MT at a given TTI helps ensuring a certain degree of fairness by avoiding situations where some MTs are allocated several RBs whereas others are not allocated any RBs. In addition to ensuring fairness, imposing this constraint allows Algorithm 1 to be applicable as is to UL and DL since the allocation of one RB per MT respects the RB contiguity constraint in single carrier FDMA (SCFDMA), a modified form of OFDMA used in LTE UL. Algorithms that allocate more than one RB per user in the DL (where OFDMA is used and the RBs allocated in the DL do not have to be consecutive) should be modified for UL implementation in order to enforce the RB contiguity constraint of SCFDMA. Moreover, the allocation of one RB per MT at a given TTI lends itself to a fair comparison with the Hungarian algorithm, which solves optimally the assignment problem.

Other algorithms can be implemented with the two-hop relay-based railroad communication scenario, including the known max C/I, proportional fair (PF), and Max–Min algorithms (e.g. see [38]), with PF and Max–Min providing more fairness. It would be interesting to investigate two-hop versions of these algorithms, or certain combinations that can be suitable for the railroad scenario. For example, one can compare, in terms of throughput and fairness, the implementation of PF scheduling at the BS–MR links then at the MR–MT links, with the implementation of Max–Min scheduling at the BS–MR links then at the MR–MT links, or a hybrid implementation of Max–Min scheduling at the BS–MR links then PF scheduling at the MR–MT links.

8.4. Inter-cell interference issues

The scenario presented in Fig. 1 considers BSs covering the rail track for the purposes of ensuring connectivity inside the trains and/or to the mobile relays on top of the trains. The BSs are considered to be equipped with directional antennas pointing in the direction of the track. Thus, interference was neglected in this case for the cellular network concerned with the railroad coverage. In fact, trains in different portions of the track would be served by different BSs. Due to the directivity of the antennas, the interference signal in the direction of the other trains would be negligible. When the trains are close enough to be served by the same BS (e.g. two trains traveling in opposite directions and passing next to each other on parallel railroad segments), the BS will allocate orthogonal resources to the MRs and/or MTs of both trains and hence no interference will occur. However, in this paper, we consider
“background” interference from the “traditional” cellular networks ensuring cellular coverage outside of the track (for subscribers in nearby towns, villages, etc.). Thus, modeling accurately the impact of this interference depends on the distance from the deployed BSs and population concentrations to the railroad track, which would vary along the track depending on the areas traversed by the train. In this paper, we treat this interference as a noise-like signal with a level of $-10 \text{ dBm}$ as indicated in Table 1. Using more detailed interference models corresponding to different outdoor environments, and also to model interference inside the train across different compartments, constitutes an important direction for future research in the area of railroad communications.

9. Conclusions

Two-hop relay-based LTE radio resource management in high speed train systems was investigated. Mobile relays positioned at the ceiling of each train car communicate with LTE BSs using an outdoor antenna, and with mobile terminals inside the train using an indoor antenna. A high mobility environment with correlated shadowing was considered. A low complexity resource allocation approach for maximizing the number of served users was presented. Relays were shown to lead to significant enhancements concerning the number of successfully served mobile terminals. Furthermore, the proposed approach was shown to perform closely to the Hungarian algorithm.

Acknowledgments

The authors would like to thank the anonymous Reviewers and the Editor for their comments that helped in significantly enhancing the clarity and quality of the paper. This work was made possible, in part, by NPRP grant # 4-347-2-127 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[28] ETSI TR 101 112 v3.2.0, Universal mobile telecommunications system (UMTS): selection procedures for the choice of radio transmission technologies of the UMTS, UMTS 30.03.


Elias Yaacoub received the B.E. degree in Electrical Engineering from the Lebanese University in 2002, the M.E. degree in Computer and Communications Engineering from the American University of Beirut (AUB) in 2005, and the Ph.D. degree in Electrical and Computer Engineering from AUB in 2010. He served as a Research Assistant in the American University of Beirut from 2004 to 2005, and in the Munich University of Technology in Spring 2005. From 2005 to 2007, he worked as a Telecommunications Engineer with Dar Al-Handasah, Shair and Partners. Since November 2010, he has been a Research Scientist at the QU Wireless Innovations Center (QUWIC), rebranded to QMIC (Qatar Mobility Innovations Center) in July 2012. His research interests include Wireless Communications, Resource Allocation in Wireless Networks, Inter-cell Interference Mitigation Techniques, Antenna Theory, Sensor Networks, and Bioinformatics.

Rachad Atat received his B.E. degree in Computer Engineering from Lebanon American University, Byblos, Lebanon, in 2010. He obtained M.Sc. in Electrical Engineering from King Abdullah University of Science and Technology (KAUST), Saudi Arabia in 2012 after being awarded the KAUST Discovery Scholarship Award. His research interests are in the general area of wireless communications with emphasis on energy and delay efficient networks, LTE network, cooperative communications, relays, heterogeneous networks and cognitive radios. He is a member of the order of Engineers in Beirut, and a member of IEEE Communications Society. He is currently pursuing his Ph.D. at the University of Kansas, Lawrence, KS.

Ahmad Alsharoa was born in Irbid, Jordan. He received the B.Sc. degree from Jordan University of Science and Technology (JUST), Irbid, Jordan in 2011, and the M.Sc. degree from King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia in 2013, both in Electrical Engineering. His research interests include cooperative relay networks, cognitive radio networks, green communication, and energy-efficient/high-performance algorithms for wireless communication systems. He is currently pursuing his Ph.D. at KAUST.

Mohamed-Slim Alouini was born in Tunis, Tunisia. He received the Ph.D. degree in electrical engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998. He was with the department of Electrical and Computer Engineering of the University of Minnesota, Minneapolis, MN, USA, then with the Electrical and Computer Engineering Program at the Texas A&M University at Qatar, Education City, Doha, Qatar. Since June 2009, he has been a Professor of Electrical Engineering in the Division of Physical Sciences and Engineering at KAUST, Saudi Arabia, where his current research interests include the design and performance analysis of wireless communication systems.