RESEARCH
Open Access

The cross layer RMPA handover: a reliable mobility pattern aware handover strategy for broadband wireless communication in a high-speed railway domain

Marina Aguado[1*], Eduardo Jacob[1], Jasone Astorga[1], Nerea Toledo[1] and Marion Berbineau[2]

Abstract

Enhancing the handover process in broadband wireless communication deployment has traditionally motivated many research initiatives. In a high-speed railway domain, the challenge is even greater. Owing to the long distances covered, the mobile node gets involved in a compulsory sequence of handover processes. Consequently, poor performance during the execution of these handover processes significantly degrades the global end-to-end performance. This article proposes a new handover strategy for the railway domain: the RMPA handover, a Reliable Mobility Pattern Aware IEEE 802.16 handover strategy "customized" for a high-speed mobility scenario. The stringent high mobility feature is balanced with three other positive features in a high-speed context: mobility pattern awareness, different sources for location discovery techniques, and a previously known traffic data profile. To the best of the authors' knowledge, there is no IEEE 802.16 handover scheme that simultaneously covers the optimization of the handover process itself and the efficient timing of the handover process. Our strategy covers both areas of research while providing a cost-effective and standards-based solution. To schedule the handover process efficiently, the RMPA strategy makes use of a context aware handover policy; that is, a handover policy based on the mobile node mobility pattern, the time required to perform the handover, the neighboring network conditions, the data traffic profile, the received power signal, and current location and speed information of the train. Our proposal merges all these variables in a cross layer interaction in the handover policy engine. It also enhances the handover process itself by establishing the values for the set of handover configuration parameters and mechanisms of the handover process. RMPA is a cost-effective strategy because compatibility with standards-based equipment is guaranteed. The major contributions of the RMPA handover are in areas that have been left open to the handover designer's discretion. Our simulation analysis validates the RMPA handover decision rules and design choices. Our results supporting a high-demand video application in the uplink stream show a significant improvement in the end-to-end quality of service parameters, including end-to-end delay (22%) and jitter (80%), when compared with a policy based on signal-to-noise-ratio information.

Keywords: Handover, Railway, IEEE 802.16, Broadband, Handover policy, Handover algorithm, High speed, Reliability, HHO, Mobile WiMAX, IMT-advanced

* Correspondence: marina.aguado@ehu.es
[1]Faculty of Engineering, University of the Basque Country (UPV/EHU), Alameda de Urquijo s/n, Bilbao 48013, Spain
Full list of author information is available at the end of the article

Introduction

The unique features of a railway scenario have traditionally generated many research initiatives mainly aimed at fostering new transportation services. One of the key railway research areas focuses on the introduction of innovative communication technologies, i.e., mobile broadband communication technologies able to support newer, safer, and richer IT railway services at affordable cost. To achieve this goal, open technology and standards-based equipment should be promoted, avoiding proprietary solutions.

In this context, and in the telecom arena, the International Telecom Union Radio Section (ITU-R) proposed the International Mobile Telecommunications (IMT) Advanced technical requirements specification. IMT-Advanced systems or 4 G technologies identify those mobile communication systems with capabilities that go further than those of IMT-2000. One of the most demanding and challenging environments covered by the IMT-Advanced long-term endeavor is a high-speed vehicular scenario that corresponds precisely with the railway scenario documented in this article. The IMT-Advanced specification defines a very stringent constraint regarding handover latency. IMT-Advanced technologies are expected to satisfy a maximum value of 27 ms for intra-frequency handover interruption time [1].

In fact, when broadband wireless technologies migrate from a nomadic scenario to a high-speed vehicular usage scenario, such as a high-speed railway domain, handover processes are the norm and not a rare exception. By increasing the speed, the dwell time within the cell decreases and the time the mobile node is involved in the handover process compared with normal operation increases. High-speed scenarios represent a higher handover rate for the same radio coverage. Consequently, end-to-end quality of service (QoS) indicators, either delay time or data loss, are significantly much affected by the handover performance [2].

Another important issue to take into consideration is that wireless transmission is always vulnerable to unexpected message loss over the radio channel. Moreover, handover normally takes place across cell boundaries where the signal strength from both the serving base station (BS) and the target BS is poor and occasionally inconsistent [3]. In lossy channels [4], as characterized by the channel when the mobile is close to the cell boundary, the signal quality declines and the probability of missing a signaling handover message increases.

A complete survey on current deployments and research studies on broadband communication technologies in the railway domain in general (not specifically high speed), focused mainly on Internet services for passengers, is covered in [5]. In all the broadband communication architectures studied, it is possible to identify four different networks: the train network, the access or train to ground or vehicle-to-infrastructure network, the aggregation network, and the core network.

In this context, the most challenging issue is to address mobility management in the access network properly. As such, a heavy burden is placed on the performance of the sequence of handover procedures taking place along the train route.

The study in [5] also identifies several implementations based on switched Ethernet in an aggregation network and the trend of using 4 G technologies such as IEEE 802.16 in access networks. Other alternatives such as IEEE 802.11 are too costly owing to the overhead of installing sufficient access points at the trackside. Satellite communication links are not suitable for broadband access or real-time services.

In January 2012, ITU, after a detailed evaluation of stringent technical and operational criteria, formally declared the WirelessMAN-Advanced specification (Mobile WiMAX Release 2 or IEEE Std 802.16.1) and LTE-Advanced to be IMT-Advanced technologies [6].

Both communication technologies satisfy a high-speed mobility scenario. Owing to maturity reasons, in this study we focus on the IEEE 802.16 specification as a suitable candidate for access networks. Backing up this approach, there are some studies on real testbed deployments [7] showing that Mobile WiMAX is a promising candidate for large-scale deployment in environments such as the railway industry.

At present, despite the vast number of research initiatives [8], the mobility management problem has not yet clearly been solved in heterogeneous vehicular communication architectures. In fact, the performance of network mobility protocols declines drastically in rapid mobility scenarios with a high handover rate. Most of the current ongoing proposals include a signaling overload that increases the difficulty of satisfying the performance requirements for handover delay specified in IMT-Advanced.

At the same time, it is also worth pointing out that normally railway companies are in charge of their own communication architectures and, for economic reasons and in terms of expertise, only one communication technology is deployed over the entire railway network.

In [4], simulations using external real trace files on IEEE 802.16e network deployments with no specific handover enhancement and taking into account the signaling message loss showed an average handover latency of 162.25 ms for data traffic. This is clearly superior to the previously identified handover latency for 4 G technologies and justifies further research in this area.

Taking all of the above into consideration, we selected a specific scenario and present our contribution providing an intra-technology handover strategy that not only

satisfies the IEEE 802.16 standard specification for broadband wireless communication technology, but also benefits from the railway scenario.

In this article, we propose a new handover strategy for communication on trains, as well as other high-speed mobility scenarios, called the RMPA handover. The RMPA is a new Reliable Mobility Pattern Aware IEEE 802.16 handover scheme for the railway domain. Our performance simulation study shows that our proposed handover scheme presents better key performance indicators (KPIs) than traditional IEEE 802.16 standard-based implementations, which rely on received signal strength measurements.

This article is structured as follows. "Related work on handover enhancement techniques" section introduces related work on current handover enhancement techniques and on specific handover strategies in a high-speed domain, while "Overview of the RMPA handover strategy" section provides an overview of the RMPA strategy. "The RMPA handover procedure" section describes the RMPA handover design and related procedures. "RMPA scanning and handover policies" section focuses on the proposed RMPA handover policy. Since this policy makes use of position information, we propose a new location discovery strategy for the railway domain, the HTRU location technique, which is a hybrid triple-redundant uncorrelated source of location information. "RMPA handover performance analyses" section presents a performance analysis of the RMPA handover scheme. First, the variables involved in the RMPA handover are identified. Then, the most common metrics used to assess handover algorithm performance, handover latency, and data loss are evaluated as a function of the variables involved in the RMPA handover. "Performance comparison between RMPA handover versus an IEEE 802.16 HHO strategy SNR-based handover policy" section compares the RMPA handover strategy with a traditional hard handover (HHO) based on signal-to-noise ratio (SNR) information using a network simulation framework. This section finally presents the global end-to-end QoS and KPI values obtained when a closed circuit TV (CCTV) application is deployed on an MS in a communication architecture with and without implementing the RMPA handover mechanism. The final section presents our main conclusions.

Related work on handover enhancement techniques

The strategies and techniques found in the literature on handover enhancement techniques in IEEE 802.16 networks can be classified according to two complementary areas of research: the handover process itself and the efficient scheduling or timing of the handover process. In this section, we first introduce major initiatives in these two research areas and then we focus on handover enhancement techniques in the railway-specific arena.

Handover enhancement techniques focusing on the handover process

Regarding the handover execution process itself, more complex execution schemes are introduced in each new release of the standard, including establishing physical-layer connections with more than one BS at a time, fast BS switching (FBSS), and the fast ranging mechanism or enhanced execution technique built into IEEE 802.16 and referred to as the *seamless handover*. However, most of these enhancements are optional features in the WiMAX profile and consequently, selecting these approaches reduces multi-vendor compatibility and increases deployment cost.

There are many studies reported in the literature that aim to improve the handover process. In 2006, Lee et al. [9] and Wang et al. [10] highlighted some deficiencies in the IEEE 802.16 standard scanning process related to existing redundancy. When several neighboring BSs are chosen as target BSs for scanning or association, since only one BS can be selected as the target BS for performing the handover, this leads to redundant scanning processes. Boone et al. [11] introduced a strategy to enhance scanning process latency by reducing the number of frequencies checked during each scanning operation. This strategy incorporates the history of successful scanning frequencies to guide the MS in choosing frequencies for future scanning operations.

Rouil and Golmie [12] (NIST) proposed the adaptive channel scanning (ACS) algorithm to enhance the handover mechanism. This algorithm minimizes the disruptive effects of scanning on application traffic by using information about the QoS traffic requirements, the available bandwidth, and the number of concurrent scanning stations to define the set of parameters for the scanning configuration.

Choi et al. [13] introduced a fast handover scheme for real-time downlink services in IEEE 802.16 networks. They focused on reducing the service disruption during HHO for real-time services by allowing the MS to receive downlink data just after synchronization with a target BS and before establishment of the MS registration and authorization. This feature is called the Fast DL_MAP_IE HO scheme for real-time downlink services. However, it only supports downlink services. Moreover, this proposal was not adopted by later releases of the standard.

One of the necessary and fundamental stages in the execution of the handover process is the scanning phase. Information obtained during the scanning process can shorten the synchronization stage. However, scanning has a disruptive effect on application traffic. As already pointed out, the more time that is dedicated to the scanning process, the greater is the increase in buffering queues in the BS. Consequently, the end-to-end delay

increases and global throughput declines accordingly. There is also a clear dependence between the disruptive effect caused by the scanning process and the data load or traffic pattern. As the data load increases, less bandwidth is available to flush the buffered data before the delay increases. The *high scanning interactivity frequency problem* arises when, during the scanning, the medium access control (MAC) buffers packets to be sent to/ received by the MS and there is not enough time after each scanning period to empty the queues, especially before performing a handover.

Another deficiency in the overall handover process is the lack of a clear definition of the scanning operation schedule. Performing a scanning process far distant from the handover process itself may turn out to be invalid owing to changes in the channel quality of the neighboring BSs and this represents a waste of system resources. We can conclude that *the scanning process should take place as close as possible to the handover process itself.*

Finally, the disruptive effect of the scanning process may be minimized through an optimal scanning design configuration. The standard defines a set of parameters involved in the scanning process (channel scanning duration, duration between scanning iterations, and number of scanning iterations) and although some default values are proposed, definition of an optimum set of values for these parameters has been omitted from the standard. This has been left as an implementation decision for the handover designer.

In summary, after reviewing the literature and from a handover process execution point of view, a single neighboring BS scanning strategy with proper parameter configuration should be promoted. Moreover, from a scheduling of the scanning process perspective, the closer to the handover initiation process the scanning process takes place, the better.

Handover enhancement techniques focusing on scheduling the handover process

The second area of research regarding the handover process in IEEE 802.16 networks is the scheduling of the handover process. Link layer triggers provide this scheduling. One of the most important triggers is the predictive link going down (LGD) trigger that denotes that a broken link is imminent. The second one is the link down (LD) trigger denoting that no information is further decodable and MAC synchronization is over. Major improvements in handover performance are achieved when LGD triggers are involved in the handover process compared with a single LD trigger strategy [14]. The IEEE 802.16 standard does not cover link trigger generation; it is implementation dependent. The handover policies involved in the handover decision process stage play a decisive role in scheduling the handover process

efficiently. Handover policies are in charge of trigger generation. Link layer triggers play an even more important role in mitigating handover service disruptions in HHO or "break before make" methods [3,15]. Most LGD trigger algorithms are based on pre-defined thresholds associated with the received signal strength indicator (RSSI) or QoS metrics [16,17].

The authors of [18] introduced an interesting mixed approach using RSSI and location information, which also takes into account errors in location information. An important feature of this proposal and analytical demonstration in [18] is that the performance of the algorithm does not deteriorate significantly when the distance measurement error increases.

Regarding the scheduling of the handover process and Layer 2 trigger generation, it has clearly been shown that timely Layer 2 triggers are needed to provide a smooth Layer 2 handover with minimal latency.

As a basic rule to be followed, predictive events such as LGD need to occur timorously to prepare for a handover. The LGD trigger should be invoked before an actual LD event by at least the time required to prepare and to execute a handover (*anticipation factor or optimum threshold value*). Thus, one of the most relevant attributes for timely link triggering is prior knowledge of the required time for handover execution.

Handover enhancement techniques in the railway scenario

Next, we introduce specific research initiatives focused on the railway scenario. As previously identified, handover policies are mainly based on RSSI measurements, although these measurements in a high-speed scenario may turn out to be unreliable metrics. They can vary significantly as a result of free space, multipath, and Rayleigh fading. To obtain good predictive accurate instant values for the anticipation factor or threshold based on RSSI measurements, very complex techniques are usually applied. Most of these are not easy to implement in practical systems. Recently, some theoretical studies proposed introducing distance information in the handover policy in railway deployments [19,20].

In [19], the authors proposed enhancing handover triggering by means of positioning information and the deployment of relay stations (RSs) with power control. These RSs are installed in the middle of overlapping regions. The main drawback of this proposal is the investment cost due to deployment of RSs.

The theoretical approach presented in [20] also benefits from the fact that in the railway environment there is no ambiguity with respect to the target handover BS. Additionally, this approach proposes equipping the train with a GPS device, so that the handover algorithm can benefit from the current geographical location and speed

information of the train. The scanning phase of the handover procedure is omitted. Although this algorithm results in a significant reduction in the handover operation time, it is still far from the required latency and involves a modification of the standard handover procedure defined by the IEEE 802.16e, thereby penalizing interoperability.

Also taking advantage of the fact that the target handover BS is unique, the authors of [21] proposed an enhanced handover algorithm specifically designed for a railway environment. The processes defined in the standard handover procedure for scanning and association of other neighboring BSs during the cell-reselection phase can directly be omitted. Additionally, to minimize the service disruption time during the handover process, this enhanced handover algorithm also proposes the use of Fast DL_MAP_IE messages to forward downlink data after synchronization with the target BS. An important drawback of this approach is that it sacrifices compatibility in favor of performance, as it modifies the standard handover procedures. Additionally, it does not deal with the appropriate timing of the handover process. In other words, it does not define any handover policy for selecting the best instant at which to initiate handover procedures.

In the implementation field, some patents can be found that deal with enhancing handover in railway systems. For example, a system for determining the optimal moment at which to perform the handover is described in [22]. This system uses installed sensors along the train track and in those areas estimated to be the optimal point for changing BSs, i.e., the dividing line between pre-calculated cells. A series of sensors are used to determine precisely the position of the vehicle, and therefore, of the mobile station. This method has drawbacks in that it requires the installation of a specific infrastructure and is independent of the incidental conditions of the network (e.g., velocity or quality of signals), and therefore, it is unable to adapt optimally to a changing scenario.

Another system is described in [23] for improving communications in a railway network by means of the installation of a series of additional BSs for those mobile stations moving at high velocity, thereby dividing users of the communications network according to their velocity. Again, the installation of a specific infrastructure is required, and despite providing a specific service to high-velocity mobile stations, this system does not optimize the handover process, which is essential to assure the stability of the services provided by the system.

Finally, the system in [24] focuses on minimizing the time required to perform a handover process by installing equipment in the MS and BS, thereby reducing the number of messages necessary to perform the process. It does not, however, give any consideration to how to optimize the moment at which the essential handover is performed. In addition, the system lacks specialization as it focuses on IP communications in a bullet train ("shinkansen" in Japanese) environment.

Based on all these previous studies, we can conclude that prior knowledge of contextual information, data traffic profiles, mobility patterns, and location information, present and predicted may represent an important contribution in determining the optimized dynamic threshold values at a much lower computational cost. A good approach, considering that location techniques are becoming increasingly accurate, is to make use of location information in the handover policy.

Overview of the RMPA handover strategy

After the research carried out we can conclude that no current proposal covers both areas of research in enhancement techniques on the IEEE802.16 handover process: the optimization of the handover process itself and the efficient timing of the handover process in a high-speed railway scenario by making use of an optimum handover policy.

Our proposal is that the RMPA handover strategy is to be supported by the BS and MS nodes included in a 4 G communication architecture such as that depicted in Figure 1. In this figure, the train or MS carries out a sequence of handover processes when moving from a PoA or BS to another BS while different railway communication services such as train control, VoIP, or maintenance services are established with the control center. In this context, the vehicular-to-infrastructure (V2I) architecture is a fully bridged WiMAX/IEEE 802.16 using Ethernet as the end-to-end transport technology. Suitability of this broadband architecture to the very specific and highly demanding railway domain has already been established in [25]. The goal is to enhance the end-to-end and handover QoS performance indicators of the consecutive number of intra technology, intra subnet, and intra domain Layer 2 handovers that the train or MS performs along the entire railway layout.

The RMPA handover consists of:

- *a set of configuration choices in the different stages of the handover process* taking into consideration a high-speed context,
- making use of pre-emptive link layer events such as LGD trigger and the new SCI trigger,
- *a handover decision-making engine which fires these link layer events in accordance to a* novel *specific railway handover and scanning policy,*
- a new location discovery technique: the Hybrid, Triple, Redundant and Uncorrelated source of location information. This HTRU technique will provide a reliable source of location information to the handover decision-making engine.
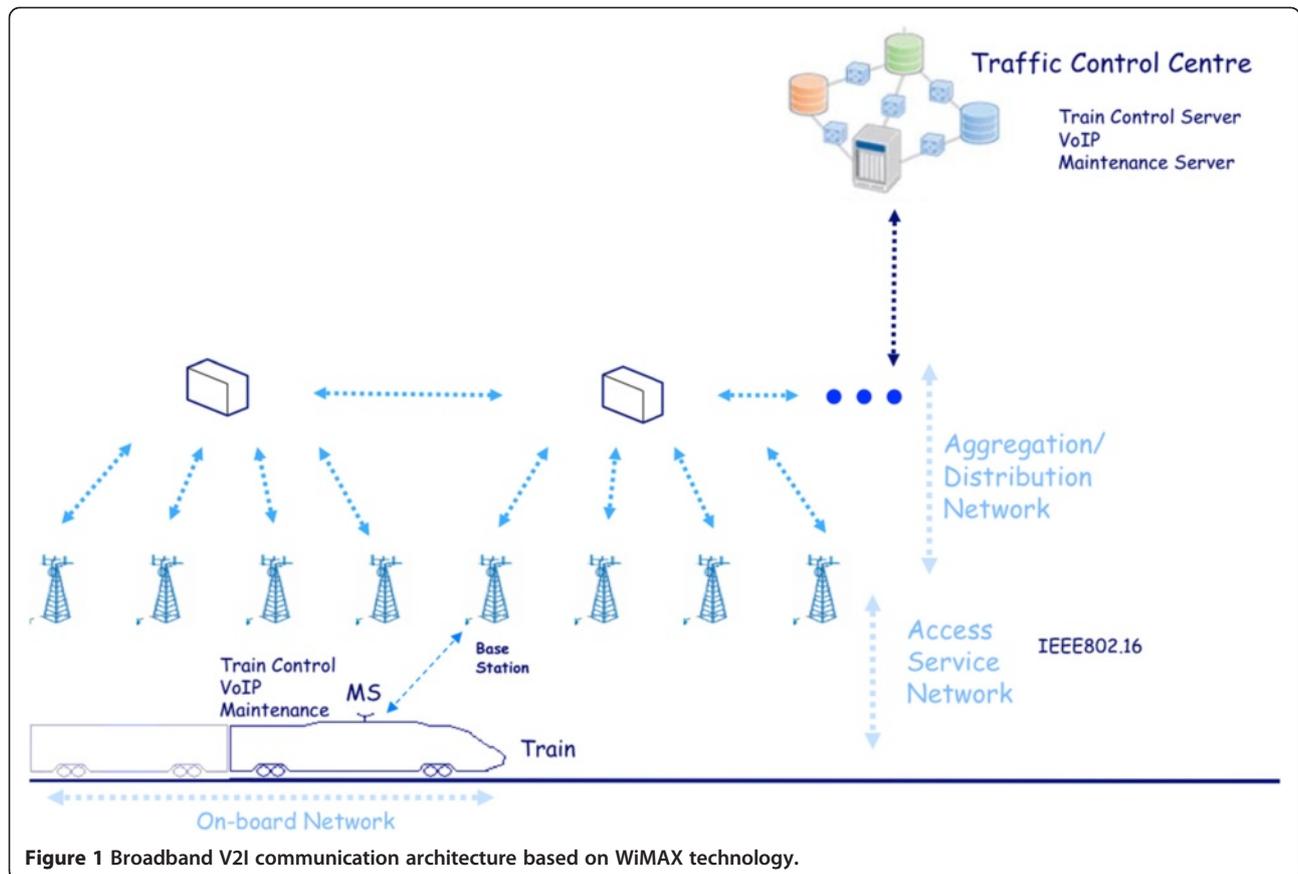
Figure 1 Broadband V2I communication architecture based on WiMAX technology.

RMPA's contributions are in areas that are left unspecified in the standard; that is, they are implementation dependent or left to the handover designer to implement. The RMPA handover strategy satisfies the IEEE 802.16 handover process standard specification. In the design specification and for compatibility purposes, only the compulsory features are selected. Consequently, it can be deployed on standards compliant multivendor equipment.

RMPA proposes a cost-effective and standards-based solution.

Tables 1 and 2 list the variables involved in the RMPA handover strategy. Table 1 details the handover configuration parameters for the MS and their respective value ranges according to the standard specification. Table 2 gives the variables related to the data traffic profile supported, the topology information, and the QoS configuration in the MS and BS involved in the RMPA handover.

In the next sections, we specify the RMPA handover procedure by describing, first, the specific procedures that take place during each stage in the RMPA handover process and, then, the new RMPA scanning and handover policies used for efficient scheduling of the handover process.

Table 1 Variables related to configuration parameters involved in the RMPA handover

Configuration parameters	Range
Frame size	2–20 ms
Neighbour advertisement interval	<Cell radio/ $(5*Ms(V)*T_{frame})$
Ranging parameters	
Backoff window size	
Ranging backoff start	0–15
Ranging backoff end	0–15
$T3$ = timeout value for receiving a valid ranging code	0–200 ms
N_{cs} = contention area[a]	>2*6
Number of retries to send contention ranging requests ($T33$)	>16
Scanning parameters	
Scan duration (N)	0–255 frames
Interleaving interval (P)	0–255 frames
Scan iterations (T)	0–255 frames
Start frame (M)	0–15 frames
$T44$ Scan request retransmission timer	0–100 ms

[a]Number of symbol times >2 per number of subchannels >6) or number of slots per frame in single carrier PHY model.

Table 2 Variables related to data traffic profile supported, topology information, and QoS configuration in the MS and BS involved in the RMPA handover

Dynamic information	MS speed
	Distance between BS serving and MS
	Distance between BS target and MS
	CINR BStarget
	CINR BStarget2
MS data traffic profile	Minimum jitter and latency from application profile
	Traffic load
QoS configuration in the MS and BS	Service class and schedulers used
	Downlink and uplink service flows (admitted and active)
	Classifier table used
	Queuing/buffering per connection
	Buffer size associated with service class and service flow
	Initial modulation chosen for each service flow

The RMPA handover procedure

This section explains the RMPA handover procedure. As previously mentioned, the RMPA handover strategy satisfies the IEEE 802.16 handover specification. Consequently, the RMPA handover strategy considers the same set of clearly identified stages in the IEEE 802.16 handover specification: network topology advertising, network topology acquisition, scanning, handover decision process, and handover execution. A concise description of the different stages in the IEEE 802.16 handover procedure, how the different stages are related to one another, and their order of execution is given in the Appendix I. Here, we present the details of the RMPA strategy in each stage of the IEEE 802.16 handover procedure.

Network topology advertising mechanism in the RMPA strategy

The RMPA network topology advertising strategy establishes the following configuration with respect to the standard IEEE 802.16 network topology advertising strategy:

- Each BS involved in the RMPA scheme is configured to belong to three different adjacent and overlapping neighborhoods, as shown in Figure 2. For example, BS3, the serving BS, is configured to belong to neighborhood A, B, and C. In the same way, BS4 belongs to neighborhood B, C, and D. The MOB_NBR_ADV message provides information
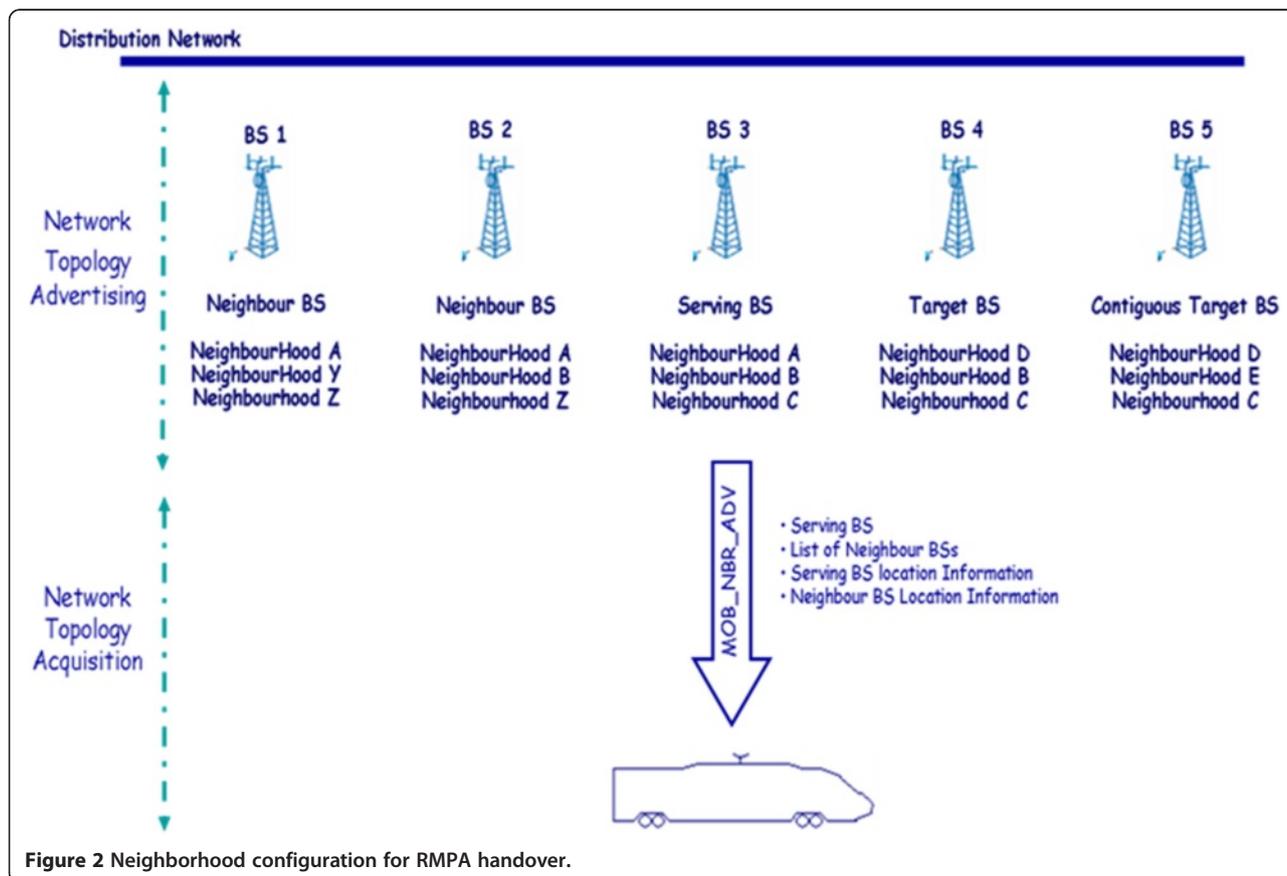

Figure 2 Neighborhood configuration for RMPA handover.

related to each of the neighbor BS in each neighborhood. This configuration is for redundancy purposes.

- Each BS acquires the channel and location information of its neighboring BSs over the backbone.
- A neighbor's advertising message (MOB_NBR-ADV) in the RMPA handover introduces location information from the serving BS.
- The maximum value of the MOB_NBR-ADV interval is initially set to be less than a fifth of the expected dwell time within a cell. The standard specification sets the maximum value of the MOB_NBR-ADV interval parameter to 30 s (*nominal time between transmission of MOB_NBR-ADV messages*). In the RMPA handover, we consider that to be effective, this interval must be smaller than the time the MS takes to reach the point at which to perform the handover or to cross over a cell. Therefore, in the RMPA scheme the maximum value of the MOB_NBR-ADV interval is set accordingly.

In this way in the RMPA strategy an MOB_NBR-ADV unconfirmed message is received *at least five times* by the MS before the handover process begins. This configuration promotes redundancy and reliability in the RMPA. In the RMPA scheme, the MOB_NBR-ADV interval depends on the MS speed and radio coverage or distance between two adjacent BSs.

Regarding the network topology acquisition strategy in the RMPA handover, when the MS receives an MOB_NBR-ADV message, it updates its neighboring BS list to include the location information of the next two target BSs. The MS then processes the information in the RMPA MOB_NBR-ADV message of the adjacent target BS. The MS double-checks its neighboring BS list by comparing the next scheduled BS target information with the current list obtained from its train program schedule. It also verifies the MS trajectory direction. In this way, the number of target BSs to be scanned during the scanning interval is reduced to just two.

Scanning mechanism in the RMPA handover

The standard does not specify the parameters for the scanning interval duration, interleaving interval duration or frequency. It merely defines certain maximum values for the parameters specifying these intervals. Determining these values is left to the handover designer.

The RMPA handover strategy fixes these values taking into consideration the railway context. It promotes a bi-neighbor scanning strategy without association. This means that the MS only scans two BSs, the target BS and the next adjacent target BS. The idea is to shorten

the time normal data traffic is disrupted while unnecessarily scanning neighboring BSs other than the target BS. Scanning the next adjacent target BS is included for the sake of redundancy and consequently, reliability.

The MS carries out a MAC synchronization enhanced through uplink channel descriptor (UCD) and downlink channel descriptor (DCD) information provided by the MOB_NBR-ADV message with the target BS during the scanning stage, but it does not perform the initial orthogonal frequency division multiplex (OFDM) ranging.

The RMPA scheme uses a simplified version of the ACS algorithm proposed in [14] to identify the correct scanning parameters (N,P,T). The goal is to correctly configure the scanning parameters so that the necessary measurements can be obtained, but without losing significant throughput.

The maximum scanning process duration, N times the frame size, T_{frame}, is established and set to be the minimum of all jitter and latencies supported by the MS data traffic profile. These values are obtained from the required QoS or KPI values for each application supported by the MS. The most common railway service values currently supported by the V2I communication architecture and their maximum values for latency and jitter are given in Table 3.

Figure 3 shows the message exchange in the RMPA scanning strategy. In this figure, M, N, and P parameters are represented. The recommendation for the N times T_{frame} in the RMPA varies from the minimum value corresponding to the necessary duration of the scanning process, T_{SCN}, to the minimum value of all jitter and latencies expressed in Table 3 for the different application profiles. In the case under investigation and considering Table 3, the maximum value for N is 4 if we take T_{frame} equal to 5 ms, and considering that the minimum value of all jitter and latencies in Table 3 is 20 ms. This value satisfies the main railway services currently supported by the V2I communication architecture.

The interleaving interval duration (P) in the RMPA handover is expected to be equal to the scanning process duration. The number of iterations is set to 2 for the sake of redundancy. For each station, the buffered data, and thus the buffer size, must be equal to the scanning process duration multiplied by the data rate.

RMPA handover promotes the use of the lowest value possible for parameter N. The second iteration ($T = 2$) in the scanning stage is configured merely for redundancy purposes. Under normal conditions, and if the handover decision is taken as expected, the scanning stage should take place only once.

Execution stage in the RMPA handover

Although handover execution obviously takes place after the handover initiation stage, the RMPA handover

Table 3 Maximum latency and jitter supported for each application scenario

|	Maximum end-to-end transfer delay	Packet delay variation within a flow jitter	Throughput
Train control application	<0.5 s		>0.2 kbps
VoIP[a]	<60 ms preferred < 200 ms limit	< 20 ms	>64 kbps
Video surveillance	<60 ms	<20 ms	>384 kbps

[a]VoIP requirements for a 4 G access technology.
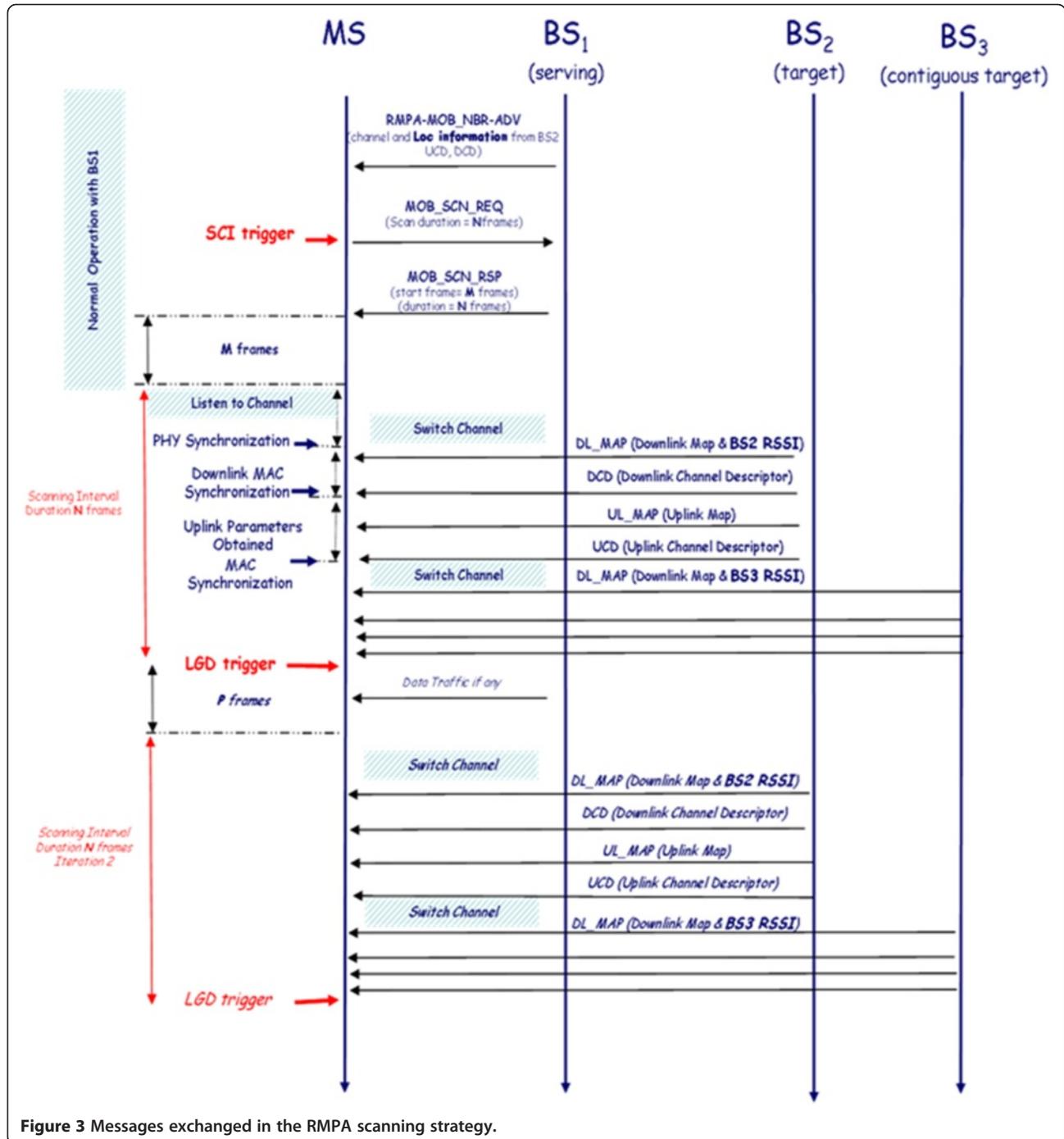
Figure 3 Messages exchanged in the RMPA scanning strategy.

execution strategy is introduced at this point because in the RMPA strategy, the handover latency, T_{HO} (time to execute the handover process), is one of the measurements considered when designing the handover algorithm. The handover algorithm is the policy that governs the handover initiation stage.

Different handover execution modes are specified in the standard: the HHO and two soft handover modes, namely, FBSS and Macro Diversity handover (MDHO). While HHO is the default handover procedure, FBSS and MDHO are useful alternatives. The HHO mechanism, unlike similar ones used in 3 G technologies such as HSDPA [8], is highly efficient and has the potential to minimize handover overhead and achieve a handover delay of less than 50 ms. Nevertheless, soft handover strategies obviously offer significantly better handover performance compared with HHO. Their main drawback is that their deployment cost is considerably greater since they require a larger number of BSs within a specified area, perfect synchronization of the active or diversity set of BSs, and accurate sharing of the same frequency. All these issues are difficult to implement in a long distance deployment such as a high-speed railway.

Taking all the above into consideration, the RMPA handover execution strategy is an HHO strategy. Thus, we focus on optimizing the worst case scenario. The next step describes the proposed control plane actions or procedures in the RMPA handover execution stage. Figure 4 illustrates the messages exchanged during the RMPA HHO process. RMPA execution process is a standard IEEE802.16 HHO handover process, consequently the sequence of message exchanged is as in any HHO process. The scanning and handover signaling latency are also represented. As novelty, Figure 4 shows the three link triggers involved in the RMPA handover: scanning process initiation (SCI) trigger, LGD trigger, and link UP trigger.

Service interruption, and consequently handover delay, in the RMPA handover execution stage is initiated when the MS switches channels to a new target BS. Since this target BS has information related to MAC synchronization, previously recorded during the scanning process, the MS directly initiates the ranging process.

The total RMPA handover execution delay, T_{HO}, is calculated as

$$T_{HO} = T_{RANGING} + T_{SBC} + T_{REG} + T_{DSA}$$

where $T_{RANGING}$ is time required for MS to perform the ranging process, T_{SBC} the time required for MS to inform on basic capabilities, SBC-REQ and SBC-RSP message exchange, T_{REG} the time required for MS registration with target BS, REG-REQ and REG-RSP message exchange, T_{DSA} the time required for the DSA-

REQ and DSA-RSP message exchange to provision service flows.

Figure 4 shows the complete message exchange during the RMPA HHO process. According to Figure 4, and considering that these messages are exchanged sequentially between MS and BS and that each message delay is T_{frame}

$$T_{handover} = T_{ranging} + 6T_{frame}$$

As indicated in [26], the maximum initial ranging latency for association level 1, before any retry, is given by where

$$T_{ranging} = \left[\frac{2^{B_{exp}} - 1}{N_{cs}} \right] x \, T_{frame+} T_3$$

B_{exp} is the backoff exponent, T_{frame} the frame duration, T_3 the timeout value for receiving a ranging response (50–200 ms), N_{cs} the number of slots per frame in Single Carrier PHY model or contention area in multi access OFDM (OFDMA) PHY profile.

It is worth pointing out that in a high-speed railway scenario, more than one train simultaneously performing a handover to the same BS is not a common use case. A good communication network design and deployment should plan the handover process to take place in a double line stretch and far away from the yards or crossing junctions. Moreover, most of the time the mobile node does not need to compete for a ranging channel.

Handover interruption time starts when the MS switches channel to target BS. However, it is interesting to take into account the time for handover preparation, or handover signaling latency and the scanning latency. Moreover, due to the disruptive effect this scanning latency represents in the global performance indicator. Another reason to assess this value is that the RMPA handover policy makes use of these two values for properly scheduling the handover process and the triggers involved.

The MS, in the RMPA scanning strategy, needs to select a channel; channel information through DCD and UCD messages has been provided via neighboring message; second, it has to listen for a preamble message. Once the preamble is found, it has to wait for a DL_MAP message and to record CINR measurements from target BS that will be used in the RMPA handover policy. The estimated minimum value for T_{SCNmin} in the RMPA handover equals to $2T_{frame}$. This value is also in accordance with the minimum estimated value in [14].

$$T_{SCN\,min} = 2 * T_{frame}$$

$$T_{SCN\,max} = 20ms$$

T_{SCNmax} should correspond to the minimum jitter from the most demanding application in the railway context, as indicated in Table 3.
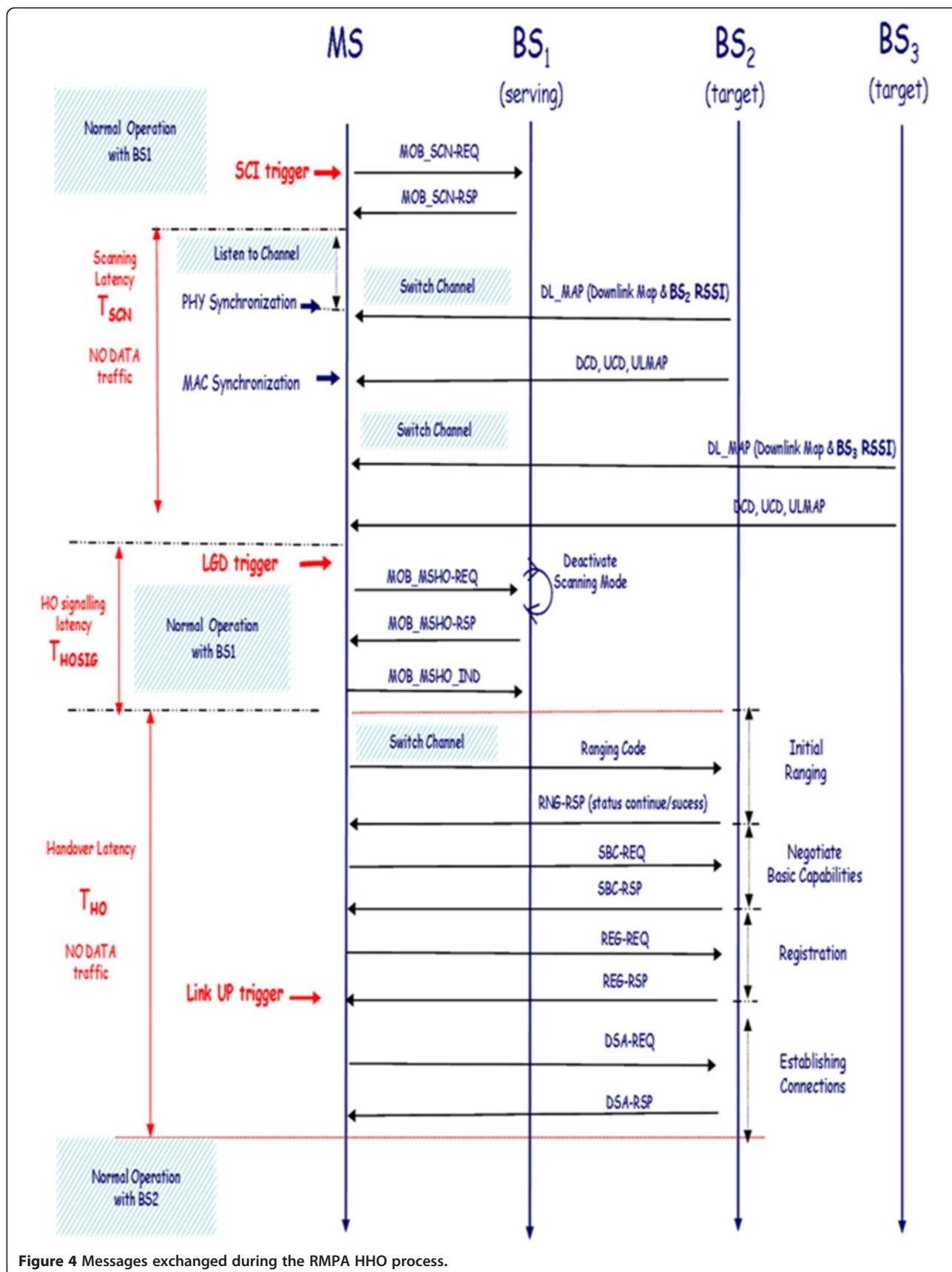
Figure 4 Messages exchanged during the RMPA HHO process.

RMPA scanning and handover policies

There are a variety of different concepts involved in the handover decision-making process: the triggers or events from the different layers, the handover policy, and the decision-making process. Combined these are commonly referred to as the handover policy—a set of rules that contribute to shaping the handover decision for a mobile node. Although it plays a pivotal role in handover performance, the handover decision policy itself is not within the scope of the standard. It is implementation dependent.

This section details the RMPA's contribution in efficiently scheduling the handover process, that is, the RMPA scanning policy and the RMPA handover policy. One of the contributions of the RMPA scheme is the proposal of adequate policies to guide the handover designer when deploying an IEEE 802.16 network in the railway context. As previously mentioned, handover policies control the firing of the link layer, which in turn, governs the handover process scheduling, i.e., when the handover process takes place. Figure 4 shows the three link triggers involved in the RMPA handover: SCI trigger, LGD trigger, and link UP trigger.

RMPA scanning policy

In the RMPA handover, the MS initiates the scanning procedure. The decision to initiate the scanning procedure is also taken by the MS. This decision is governed by the scanning policy, which likes the handover policy, is omitted from the standard and is implementation dependent. In most current implementations, the scanning policy is based on RSSI measurements (either in the MS or BS).

The scanning process, in the standard specification and in general, although disruptive, is a necessary step prior to execution of the handover process. Most often, in power strength-based algorithms, RSSI measurements taken during this process feed the handover decision-making process, firing the LGD trigger, and initiating the handover execution.

In most implementations, an MS requests (by sending an MOB_SCN-RSP) to enter into scanning mode if its uplink measurements fall below a fixed "scanning threshold". It is also possible to specify multiple scanning thresholds and to configure a set of scanning intervals associated with each threshold. In this way, the MS can initiate different "scanning behaviors" according to the measurements received. However, a better approach is to use a dynamic or adaptive threshold strategy. This means being able to increase the scanning interactivity as the MS gets closer to the cell boundary.

In the RMPA scanning initiation strategy, the adaptive threshold strategy has been taken to the extreme. It has been turned into a single scan interactivity period strategy and the generation of a scanning request from any MS is postponed as much as possible.

In the RMPA scheme, the scanning policy is based on the following rule: *"The scanning process should take place just once and the closer to the handover initiation process it occurs, the better"*.

To achieve this, the RMPA approach is based on a tight timing mechanism for trigger delivery that involves chained or concatenated triggers.

The RMPA scanning policy follows the RMPA handover policy with an anticipation factor equal to the scanning process duration in frames, N. This scanning process latency has previously been established to range between the following two values:

- It is higher than the time required to perform MAC synchronization.
- It is lower than the minimum of all latencies and jitter from the different application KPIs supported in each scenario. Consequently, it is closely related to the MS data traffic and QoS profile.

For the RMPA strategy, we propose a new predictive SCI trigger, defined to initiate active scanning of neighboring BSs and which occurs earlier than the LGD trigger. This approach is in keeping with the IEEE 802.21 Media Independent Handover (MIH) initiative group. The SCI trigger initiates the scanning procedure. As illustrated in Figure 4, the next step involves the MS sending an MOB_SCN_REQ to the serving BS. Upon receipt of the MOB_SCN_RSP message, the BS responds with an MOB_SCN-RSP. The relationship between the normal operation state, the scanning process, and the handover process can be observed in the flowchart represented in Figure 5. As it can be observed, the decisions that lead to jump from normal operation to the scanning stage or to the handover stage are ruled by the reception of SCI and LGD triggers.

LGD trigger generation in the RMPA scheme follows the handover policy detailed in the next section. It is not generated based on signal degradation alone.

To summarize, in the RMPA strategy a new SCI trigger is proposed, the scanning policy is the policy in charge of triggering this SCI, and this policy is aligned with the handover policy.

RMPA handover policy

The handover process begins with the decision for an MS to hand over from the serving BS to a target BS. This decision may take place in either the MS, the serving BS, or the network. The RMPA handover initiation strategy could be a network-initiated handover: this design choice would allow a railway operator to control the target network selected by the mobile node.
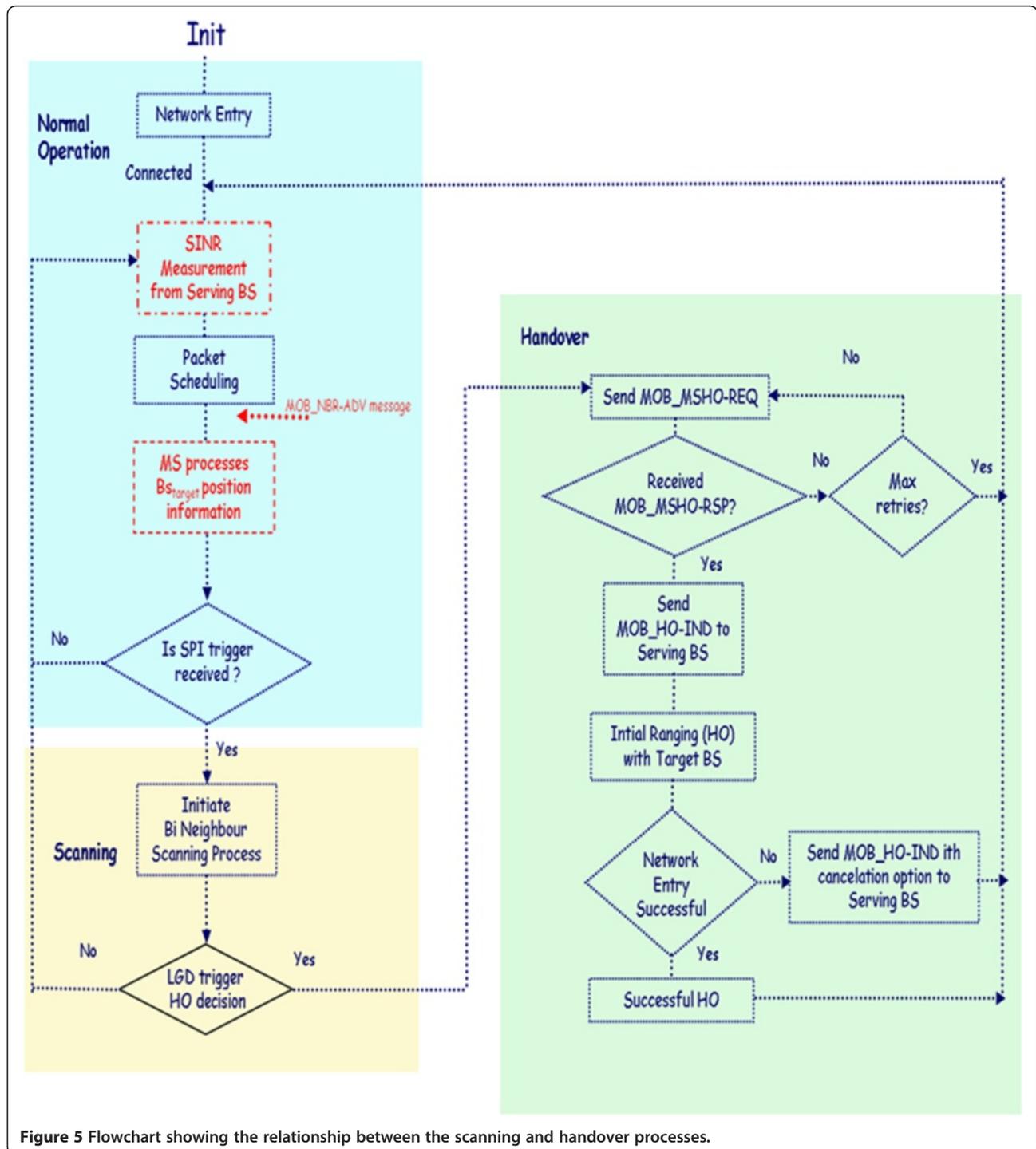
Figure 5 Flowchart showing the relationship between the scanning and handover processes.

However, this option has less redundancy and consistency. Since the handover policy is based on measurement taken in the MS (e.g., location information or RSSI), the RMPA handover supports an MS-initiated handover strategy. The handover decision is taken in the MS.

According to the conclusions reached in "Related work on handover enhancement techniques" section, the RMPA

handover policy for a high-speed railway scenario is based on prior knowledge of the mobile node's mobility pattern, the time required to perform handover signaling, the neighboring network conditions or contextual information, the data traffic profile, the received power signal in this high-speed scenario wireless channel, and the current mobile node location and speed information. All these variables and information come from different layers

(PHY, MAC, and application layer). The handover policy engine contained in the MS node and all the variables involved are represented in Figure 6. The RMPA handover decision engine contained in the MS integrates all of these in a cross-layer interaction. We make use of MIH function and services to generate the necessary triggers and events in our architecture.

The LGD in the RMPA handover is triggered according to the RMPA handover policy mechanism embedded in the MS policy engine. It is not fired merely as a result of signal power degradation, information on distance to

target BS is also relevant. Rule 3 in this section specifies when LGD trigger is fired. Once the LGD is fired, the MS sends the MOB_MSHO-REQ as shown in Figure 4 and initiates the handover execution process.

The following variables from the different layers and represented in Figure 6 take part in the handover policy:

1. Mobile node type of train. Each specific type of train has a predefined train program, that is, source and destination, and route to be followed. As such, the sequence of expected BSs along the full trajectory of
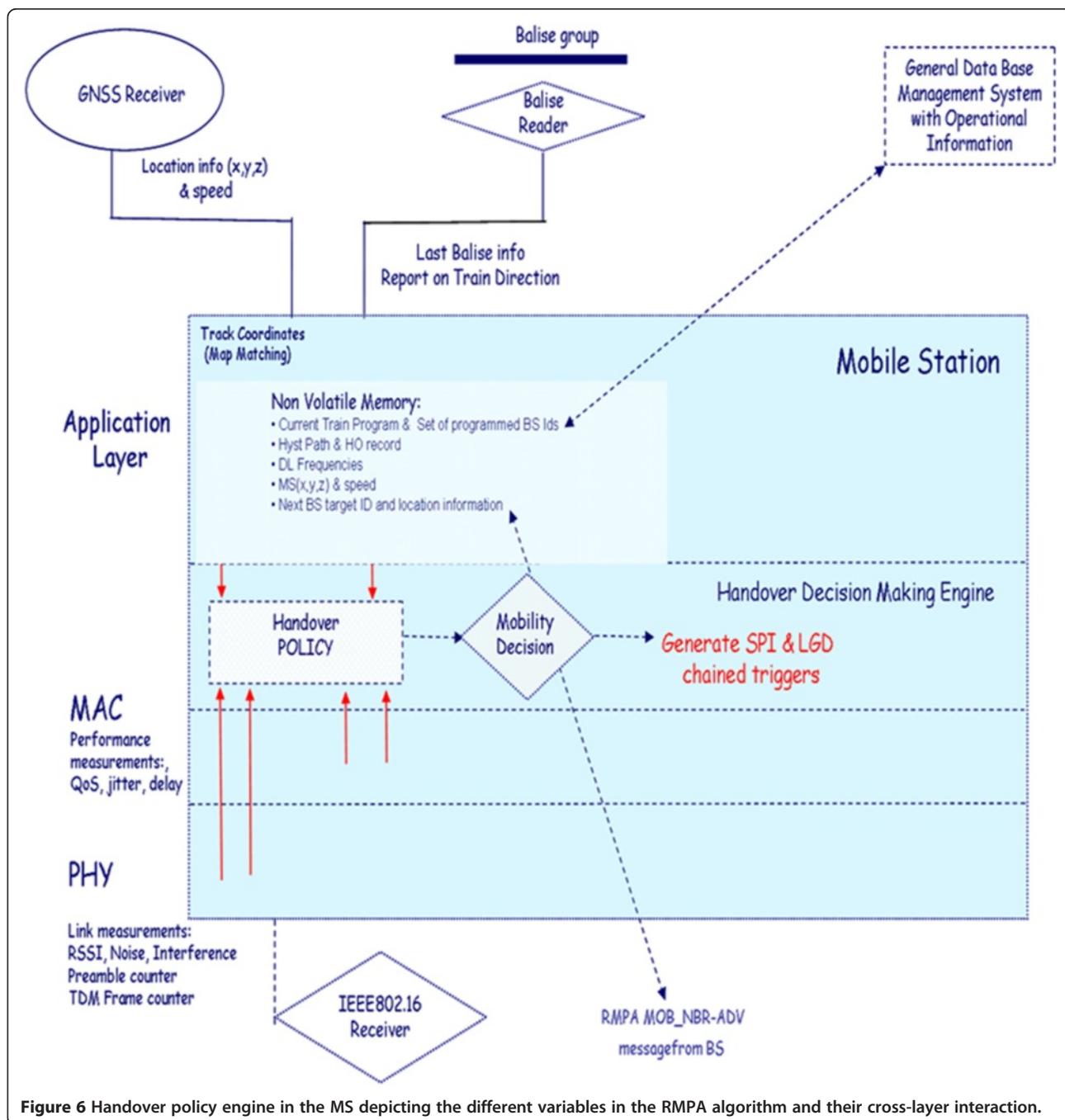

Figure 6 Handover policy engine in the MS depicting the different variables in the RMPA algorithm and their cross-layer interaction.

the MS is known. As it is indicated in Figure 6, this information is queried from the General Data Base Management System.

2. Historical information regarding previously undertaken routes for this specific type of train. This information may be obtained from the non-volatile memory unless it is the first train for this train program. This information may also be queried from the General Data Base Management System.

3. MS current location and speed. This information is obtained through the proposed HTRU location technique detailed in "Location discovery technique in the RMPA handover: the HTRU strategy" section.

4. The next BS's target id, channel information, DL frequencies, and location information obtained from the last received RMPA MOB_NBR-ADV message from $BS_{serving}$.

5. Data traffic QoS specific requirements for the specific profile supported by the MS. This information is obtained from the MAC layer configuration.

6. RSSI information from both the serving and target BSs reported by the PHY layer. RSSI measurements from the serving BS are recorded in the MS once per received frame when the MS is in Normal Operation status. These measurements are normally elicited from the downlink preambles, which occur once per frame. RSSI measurements from the BS target are received during the scanning interval.

7. The expected handover signaling time T_{HOSIG}, which is also known in advance.

Some considerations are necessary. The train type defines the source and destination and consequently the route of the train and the sequence of BSs in that route. By checking the MS type of train and its current location, the next expected target BS id can be determined. This value is double checked with the information provided by the RMPA MOB_NBR-ADV message.

Figure 6 shows the different variables involved in the RMPA handover algorithm and their emplacement in different layers.

The following rules were used to guide the RMPA handover policy design:

Rule 1
The scanning process should take place immediately before the handover process. The handover process then benefits from considering valid target BS information obtained from the last MAC synchronization of the MS with the target BS in the preceding scanning process.

Figure 7 represents the downlink SNR received by the train from the serving and target BS across the train route. It can be observed the message exchanged and the triggers involved. It is also represented the distance between the MS and the serving BS and between the MS and the target BS ($MSBS_{target}$ and $MSBS_{serving}$). As shown in Figure 7, the following relationship is defined between the two link triggers that govern the scheduling of the RMPA handover scheme:

$$t_{SCI} < t_{LGD} < t_{LD}$$

where t_{SCI}, t_{LGD}, and t_{LD} are the time instants at which the SCI, LGD, and LD triggers are fired.

The SCI trigger must be launched soon enough so that the MS has sufficient time to execute the single neighbor RMPA scanning strategy (T_{SCN}) and to exchange the MOB_MSHO-REQ/RSP and MOB_MSHO-IND messages (T_{HOSIG}).

The LGD trigger should timeously be fired prior to an expected LD event by at least the time required to prepare and execute a handover (T_{HOSIG} and T_{HO}), as illustrated in Figure 7.

Rule 2
To minimize general packet loss during the handover process, and in particular, handover signaling packet loss, the RMPA handover policy should search for the highest power signal received from each BS. As shown in Figure 7, the maximum theoretical power signal area is obtained if the LGD trigger is scheduled so that the channel switch occurs at the midpoint. This strategy for promoting the channel switch at the midpoint provides a balance between two commonly used handover metrics: successful handover rate and average number of handovers.

Rule 3
Location discovery techniques have evolved significantly over the last few years and particularly in the railway domain. It can also be concluded from the handover enhancement survey carried out in "Introduction" section that a policy involving dynamic thresholds and a mixed handover policy with location and RSSI information are the most suitable for a high-speed railway domain.

The rule for launching the SCI trigger is based on distance information:

$$M\bar{S}B\bar{S}_{target} < \left(M\bar{S}BS_{serving} + Distance\bar{}_{Threshold1} \right)$$

where

$$M\bar{S}B\bar{S}_{taegtr} = distance\,between\,MS\,and\,BS_{target}$$

$$M\bar{S}B\bar{S}_{serving} = distance\,between\,MS\,and\,BS_{serving}$$

The SCI trigger is fired when the distance from the MS to the target BS is lower than that from the MS to

Figure 7 Tight timing mechanism policy for trigger delivery.

the serving BS plus a specific distance given by Distance_Threshold1.

The rule for launching the LGD trigger is based on:

$$MSBS_{target} < \left(MSBS_{serving} + Distance_{Threshold}2 \right) \; \grave{U}$$

Where $MSBS_{serving}$ = distance between MS and $BS_{serving}$, $CINR_{BStarget}$ = CINR from BS_{target}, $CINR_{BStarget}$ = CINR from BS_{target}, $CINR_{BStarget2}$ = CINR from $BS_{target2}$.

The LGD trigger is fired when the distance from the MS to the target BS is lower than that from the MS to the serving BS plus a specific distance, given by Distance_Threshold2 and the carrier-to-interference-plus-noise ratio (CINR) from the target BS is higher than that from the next target BS (denoted as $BS_{target2}$). The latter condition ensures that the target BS is ready. Otherwise, the handover initiation should not go ahead. It should wait

until the distance from the MS to the next target BS is lower than that from the MS to the serving BS.

Rule 4

From the RMPA procedure described previously, it should be understood that when determining the consecutive decision-making stages for SCI and LGD trigger generation, there is enough information to allow the MS to calculate the expected handover signaling latency. In other words, that the expected T_{HOSIG} and T_{SCN} are known, as well as the MS's speed, and consequently the values for Distance_Threshold1 and Distance_Threshold2.

Distance_Threshold1 is equal to minus twice the scanning latency plus the handover latency multiplied by the MS speed.

$$Distance_{Threshold}1 \; = \; -2 \left(T_{SCN} + T_{HOSIG} \right) * M\bar{S}(v)$$

Distance_Threshold2 is equal to minus twice the handover signalling latency multiplied by the MS speed.

$$\text{Distance } \bar{\text{T}}\text{hreshold2} = -2\,(\text{T}_{\text{HOSIG}}) * \bar{\text{MS}}(v)$$

As calculated in "Execution stage in the RMPA handover" section the range of T_{SCN} is defined as

$$\text{T}_{\text{SCNmin}} = 2 * \text{T}_{\text{frame}} \quad \text{and} \quad \text{T}_{\text{SCNmax}} = 20 \text{ ms}$$

Here, 20 ms corresponds to the minimum jitter from the most demanding application in the railway context.

The handover signaling latency T_{HOSIG} represents the required time to exchange the MOB_MSHO-REQ, MOB_MSHO-RSP, and MOB_MSHO-IND messages.

$$\text{T}_{\text{HOSIG}} = 3 * \text{T}_{\text{frame}}$$

Figure 8 presents a flowchart of the algorithm that governs the RMPA handover decision-making engine. It can be observed from Figure 8 that after the initialization stage, when the non-volatile area of memory in the MS is first updated with location, speed, train program, train direction, and historic information, the decision-making engine in the MS starts the decision making loop, in which updated information from the previously indicated variables is verified sequentially together with the two necessary conditions for SCI and LGD trigger generations. The second rule for generating LGD trigger employs the CINR measurements from the target BS and next target BS obtained during the scanning process.

Location discovery technique in the RMPA handover: the HTRU strategy

This section details our proposed location discovery technique in the RMPA handover scheme: the HTRU strategy, which is a hybrid, triple, redundant, and uncorrelated source of location information.

Efficient handover timing depends mostly on on-time trigger generation. This generation depends on a proper handover policy, which, at the same time, relies on accurate measurements from the different sources of information. Consequently, one of the *RMPA handover design requirements* is to make use of accurate information for all the variables involved in the decision-making process. Thus, the RMPA handover proposal makes use of sources of location information commonly available in the railway scenario, such as Global Navigation Satellite Systems (GNSS) and balises, and adds specific resources from the time division multiplexed (TDD) communication system technology chosen.

GNSS coordinates typically represent a fixed accuracy and precision depending on the manufacturer and the reception environment. In open areas and with classic receivers an accuracy of 10 m with a precision of 95% is achieved. This means that there is a 95% probability that the user is located in an area with radius 10 m around the calculated GNSS coordinates. In this study, the "rounded or spherical probability" can be reduced by introducing track GNSS locations previously recorded. The three-dimensional problem is therefore reduced to a two-dimensional equation in the railway domain. The mobile node performs only track-based movement. A map-matching algorithm can be used to increase the GNSS performance. In major railway networks in the USA, a differentiated GPS (DGPS) system is also used. In general terms, the railway reception environment decreases GNSS performance (availability and accuracy). This is due to the fact that, in these contexts, GNSS signals experience multipaths and lack of visibility in the vicinity of the antenna, particularly in cutting or urban environments. For satellite-based positioning systems, performance depends on satellite visibility. Reflections or other multipath effects introduce pseudo-range errors that degrade GNSS performance. These pseudo-range errors generally follow a non-Gaussian distribution as demonstrated in [27].

Given the hardware constraints with respect to the frequency at which these GNSS coordinates are updated (of the order of 1 s or even 100 ms), it is necessary to implement extrapolation techniques. In TDD systems such as the chosen IEEE 802.16 profile, a frame has a fixed duration. So, keeping track on the number of frames received since the last GNSS measurement and being aware of MS speed and the direction (information obtained from the last balise group information), it is possible to predict the MS location information prior to the next GNSS record. This location information is computed for each frame that arrives.

As another uncorrelated source of information, balise groups located every 500 m provide information on the last registered mobility pattern record. The mobile node reports its calculated position to the network (via basic connection messages) for redundancy and to feed the global Management Information System. The communication architecture can then monitor the trajectory of the mobile host to determine with high probability to which BS the mobile host should hand over. There are several benefits involved: advanced radio resource management information can be obtained from the imminent target BS and buffers can be prepared in advance. Figure 9 depicts the different sources of location information involved in the HTRU location strategy. It can be observed that the GNSS receiver provides the GNSS location information, the IEE802.16 receiver provides information regarding the TDM frame counter, and the balise reader equipment provides information on the last balise and consequently the train direction. Apart from this input, the mobile station is also aware of
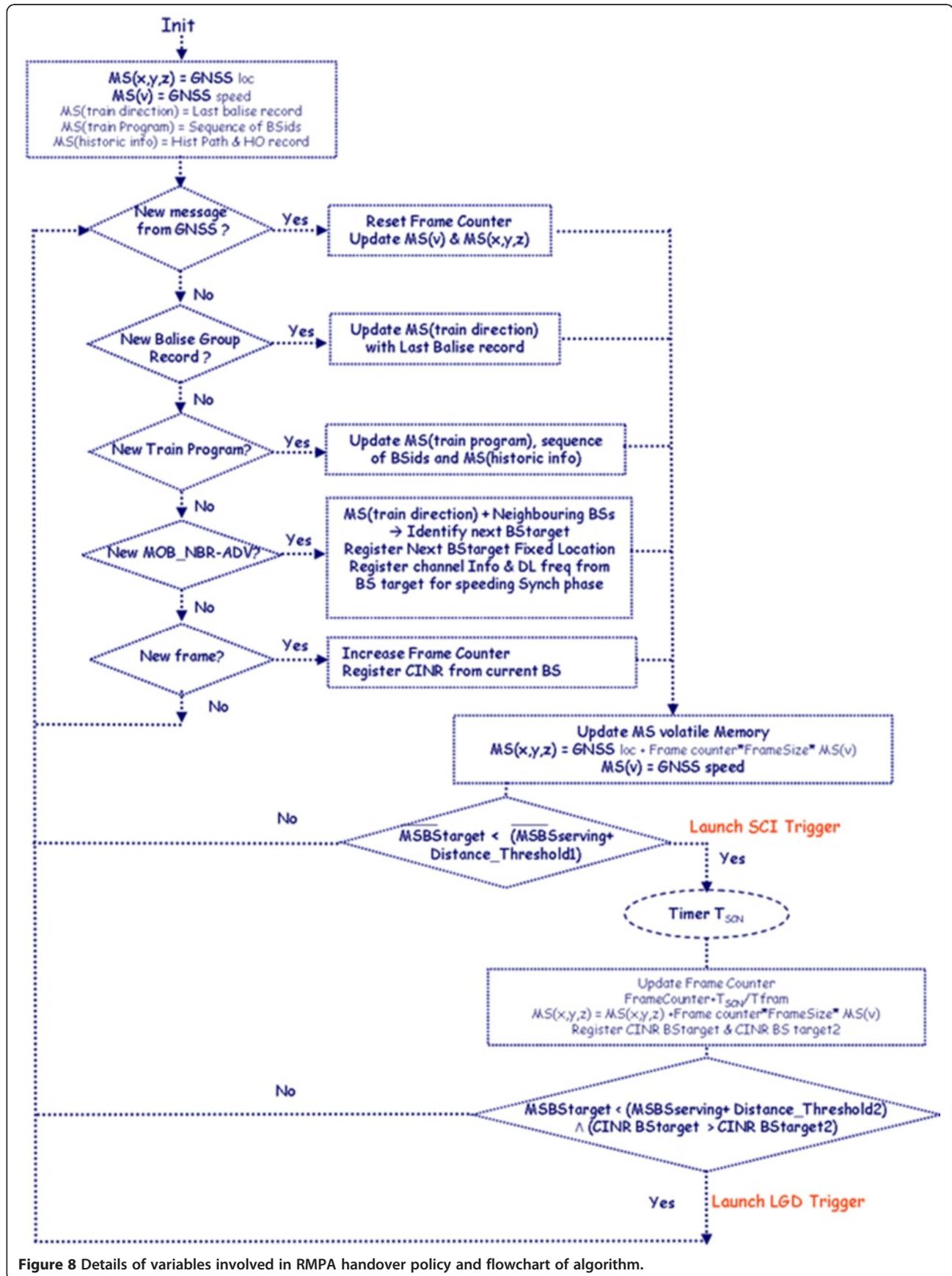
Figure 8 Details of variables involved in RMPA handover policy and flowchart of algorithm.
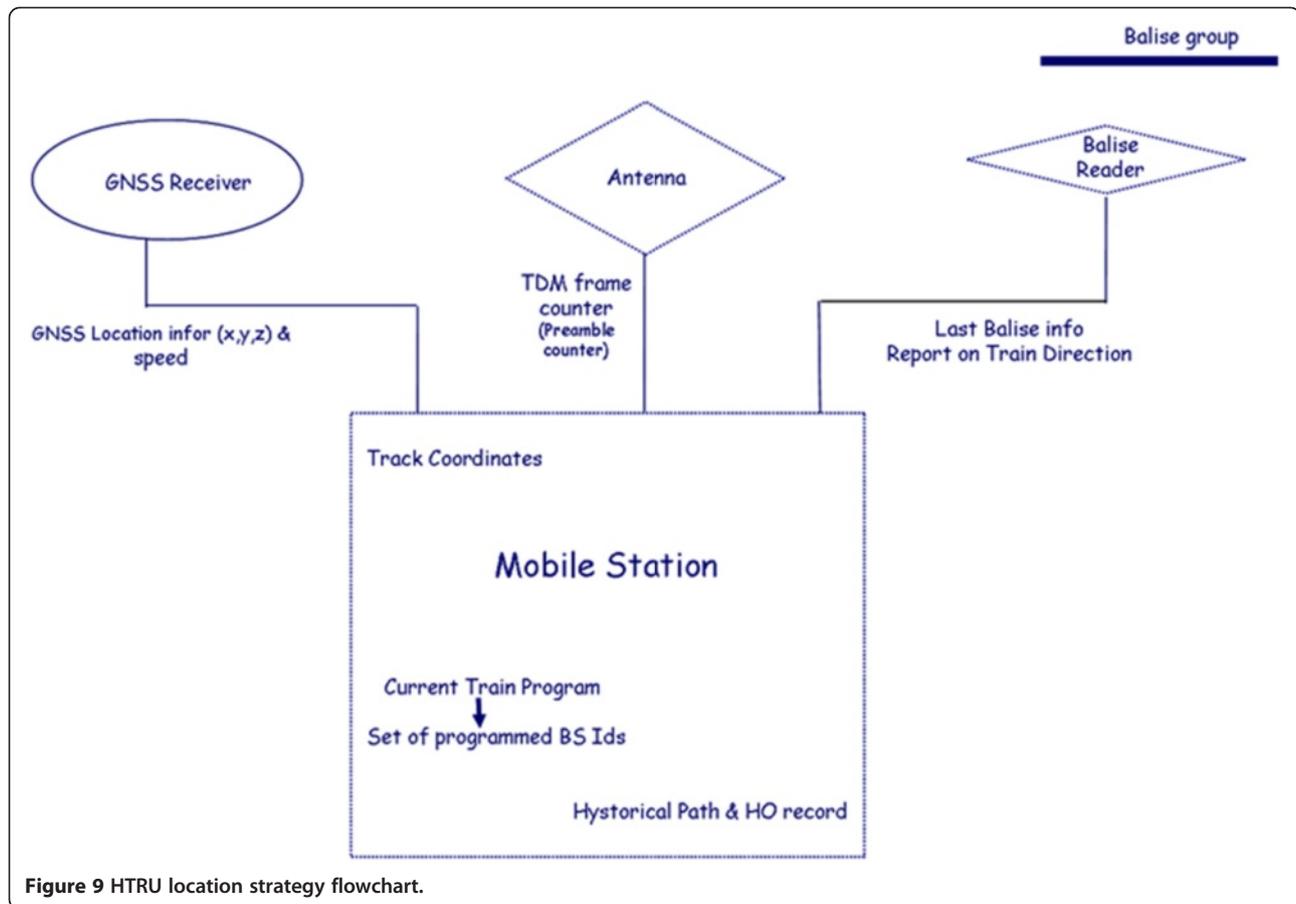
Figure 9 HTRU location strategy flowchart.

the current train program and the set of programmed BS Ids.

The main aim of the proposed specific location mechanism is to increase availability and accuracy using location information. The handover decision-making stage makes use of and relies on this information. A flowchart of the HTRU algorithm is given in Figure 10.

In this figure, it is represented that, initially, the MS updates its location and speed information from the GNSS information. The MS direction is double checked with the information obtained from the last balise-registered information. The MS updates its location information when a new GNSS message is received, information relative to a new balise group is obtained, or when a new value is reached in the preamble frame counter. The updated location value is the frame counter preamble value times the frame size times the mobile node speed. The mobile node speed is considered constant during two GNSS messages.

Thus, we propose a triple, mostly uncorrelated, source of location information. The GNSS (DGPS + map-matching) equipment, the balise reader and frame (*preamble*) counter take part in the HTRU strategy. This approach takes into consideration the downlink

preambles that arrive one per frame. The RMPA approach also takes into account that during the scanning process there are no arrivals of downlink preambles from the serving BS.

RMPA handover performance analyses

Having introduced the RMPA strategy for IEEE 802.16 handovers in railways, in this section we present a performance analysis of the proposed RMPA handover scheme. We used a simulation framework, OPNET Modeler, to model our RMPA handover, and compared it with the default SNR-based implementation. In all the simulation graphs here represented and obtained directly through the simulation platform Opnet Modeler, the *X*-axis denotes simulation time and the legend above the graph denotes the variable in the *Y*-axis.

The first step in the analysis identifies the most commonly used metrics for assessing handover algorithm performance (handover latency and data loss) and shows their interdependence and the relationship between the variables or handover design parameters in the RMPA handover strategy. In the second step, we validate the RMPA handover policy and design rules in the simulation platform by measuring how the scheduling of the
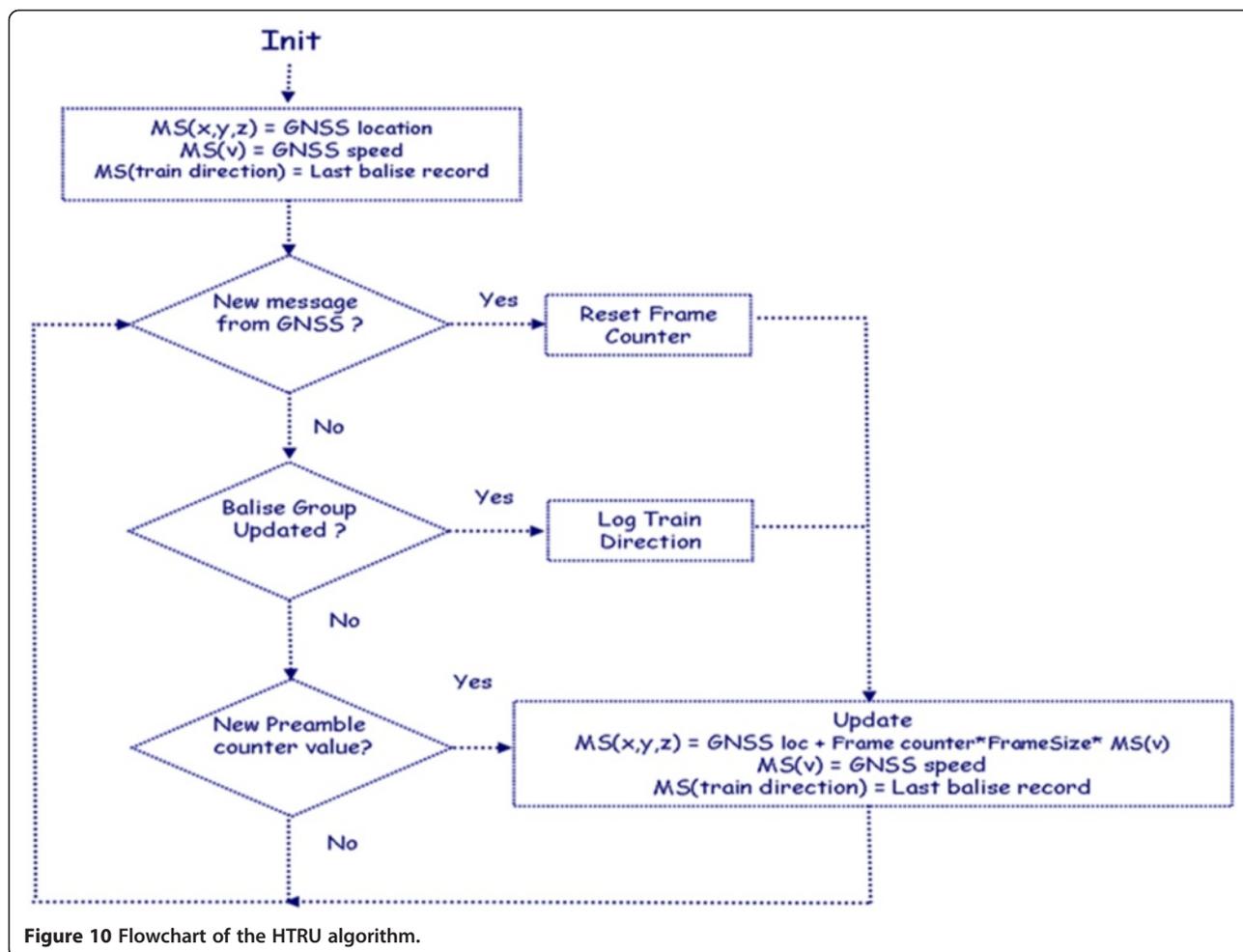
Figure 10 Flowchart of the HTRU algorithm.

handover process affects the performance indicators (data loss during handover). The third stage in the analysis compares the RMPA strategy with an IEEE 802.16 HHO strategy that uses a traditional power-based handover policy.

Regarding the simulation scenario used, some authors when analyzing handover performance consider the influence of all six adjacent BSs in the hexagonal cell model. However, in our case, considering the communication architecture topology deployment and taking into account that a great deal can be learnt about handover performance metrics merely by analyzing its fundamental performance in a link level setting, a simplified scenario was considered. This deployment scenario is the scenario proposed by the IEEE 802.16 m in its documentation regarding handover evaluation methodology. It is a *single user mobility* scenario, consisting of three BSs and one MS. The effect of other adjacent BSs is omitted; this simplifies our performance analysis. Figure 11 depicts the simulation scenario implemented in the OPNET Modeler tool. Additional file 1 provides detail on simulation scenario configuration in this tool.

Performance metrics: handover interruption time and data loss in the RMPA handover

Here, handover interruption time, also known as handover latency or handover delay, represents the time duration that an MS cannot receive service from any BS during a handover. It is defined as the time interval from when the MS disconnects from the serving BS to the start of transmission of the first data packet from the target BS. Latency is a key metric in evaluating and comparing various handover schemes as it has a direct impact on application performance as perceived by the user.

As stated in "Execution stage in the RMPA handover" section, the total RMPA handover execution delay, T_{handover}, is defined as

$$T_{\text{handover}} = T_{\text{ranging}} + 6T_{\text{frame}}$$

The following dependencies are identified. Regarding the backoff window size: a higher value of this parameter in a high-density scenario results in higher values for handover delay. The time out value for receiving a

Figure 11 Simulation scenario implemented in OPNET Modeler tool.

ranging response (default 50 ms) outlines the maximum value for the ranging activity delay. The higher the number of initial ranging opportunities per frame, the lower is the handover delay.

The handover interruption time is independent of handover policy, unless the handover policy results in very late scheduling of the handover process and consequently, handover failure. Handover interruption time is dependent on the other performance metric: data loss. A high data loss during the handover process increases the possibility of signaling handover message loss and consequently, timer expirations and retransmission could markedly increase handover latency.

Data loss is the second handover performance criterion commonly used in the literature. Average packet loss or bounded packet loss during handover is defined as the number of packets that fail to reach the MS during link layer handover. A more global performance indicator is the information loss or packet error rate (PER), which takes into account the global end-to-end packet loss and not only the handover packet loss.

Directly related to this data loss parameter is the common measurement PER, or equivalently block error rate (BLER) or frame error rate. All these measurements refer to the probability that at least one bit is in error in a block of L bits. This is the more relevant measure, since the detection of a single bit error in a packet by the cyclic redundancy check causes the packet to be discarded by the receiver. When narrowband systems are considered, the bit error rate probability of quadrature amplitude modulation systems in additive white Gaussian noise (AWGN) channels increases when SNR decreases [28].

However, when considering a multi-carrier system with frequency selective fading such as the OFDMA system, classic AWGN curves relating the SNR in the

channel with the resulting BLER do not apply directly; the OFDMA system can have a different SNR in each subcarrier.

When system level simulators are involved, it turns out to be infeasible and impractical to simulate the instantaneous performance of a wireless link in real-time particularly with channels that are frequency selective and where the signal is being impaired by interfering signals and thermal noise. An effective SNR based on an exponentially effective SNR map is usually considered a better metric than average SNR.

The OPNET simulation tool used models the following characteristics of the OFDMA WiMAX system: channel model effects (multipath fading effects based on the power delay profile of the channel, time correlation due to Doppler effects, path loss, log-normal fading) and co-channel interference (matrix of subchannel-to-subchannel overlaps for different PermBasis).

Data loss or PER during handover is thus related to the modulation and coding scheme, the block length, and the effective SNR during the handover process. Figure 12 depicts the downlink SNR values (in dB) for packet transmissions through the WiMAX physical layer for $BS_{serving}$ (blue) and BS_{target} (red) obtained in the simulation platform.

By zooming in to the identified green area, as shown in Figure 13, the variability in the SNR profile close to the handover area can be seen.

Validation of RMPA design rules
To minimize the data loss during the handover process, the maximum value for the SNR measured at the mobile node during the full handover process must be investigated. The effective SNR received by the mobile node during the handover process from each BS depends on the moment that the LGD is triggered, and consequently
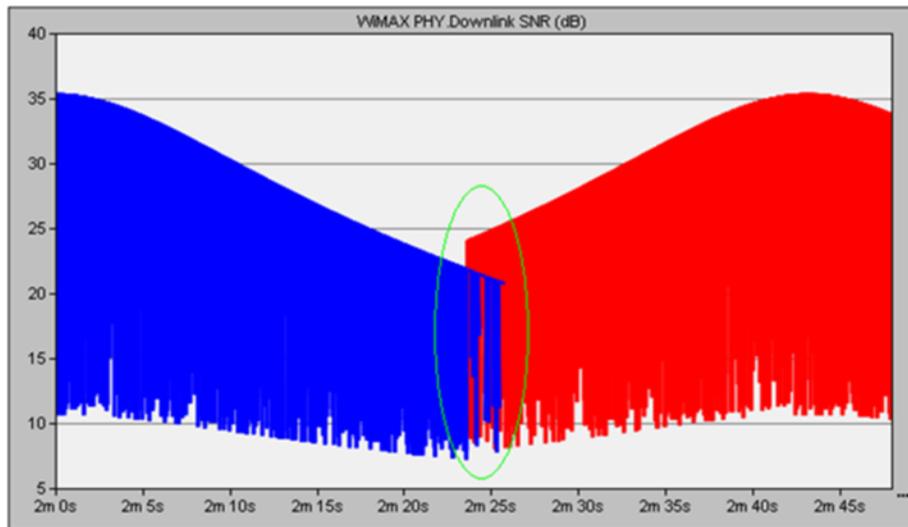
Figure 12 Downlink SNR from serving and target BSs.

on the handover policy. As shown in Figure 13, the highest area value under the SNR curve is achieved if the MS switches channels at the midpoint.

Thus, aiming to minimize the data loss, the RMPA handover policy schedules the handover so that the SNR received by the MS from both BSs during the handover process is maximized. To achieve this, it schedules the LGD so that the handover occurs at the midpoint.

To confirm that the data loss decreases if the handover takes place closer to the midpoint, a distance-based handover policy with varying values for Distance_Threshold1 (scanning threshold) and Distance_Threshold2

(handover threshold) was used. The configuration values for SCI and LGD trigger are shown in Table 4.

Figure 14 shows how the integration function of the BLER measured at the MS nodes for all packets arriving from the BS decreases as both the distance handover and scanning handover thresholds decrease. Each of the colored lines represents the results obtained from one scenario configuration. For example, the dark blue line, legend RMPA_1_DES1 represents the results obtained for scenario configuration 1 in Table 4. It can be observed that the lowest value for the downlink BLER curve is obtained for scenario number 11. In this
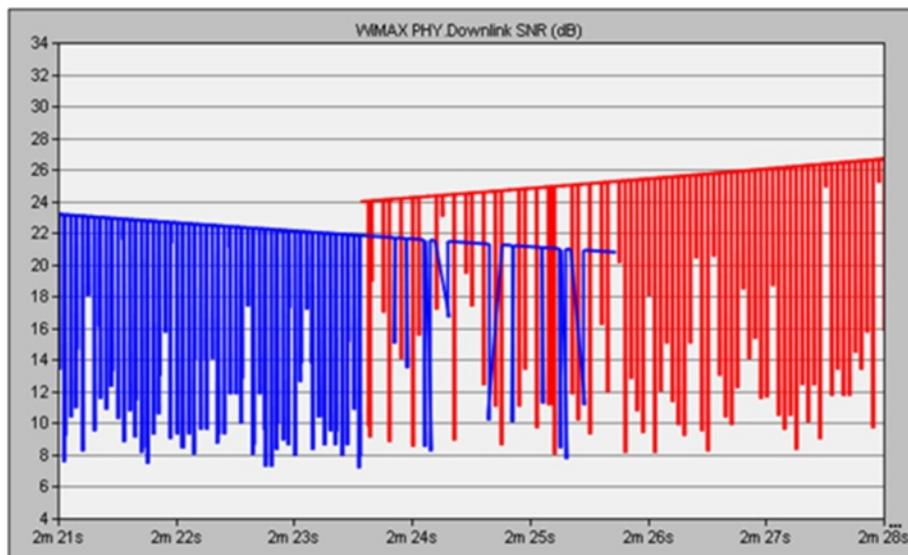
Figure 13 Downlink SNR from serving and target BSs.

Table 4 SCI and LGD trigger values for each simulation scenario

Scenario	1	2	3	4	5	6	7	8	9	10	11
SCI—Distance_Threshold1 scanning threshold (m)	200	182	165	147	130	112	95	77	60	42	25
LGD—Distance_Threshold2 handover threshold (m)	100	90	80	70	60	50	40	30	20	10	1

scenario configuration SCI and LGD trigger are fired the closest possible to the midpoint.

To verify the disruptive effect of the scanning process, we consider the following configuration. In this case, the SCI trigger initiates 200 m from the midpoint in each run, while the distance for the handover threshold decreases. The configuration values for SCI and LGD trigger are shown in Table 5. Figure 15 shows that the integration function of the BLER measured in the MS node for all packets arriving from the BS increases as the distance between the SCI and LGD triggers increases. The colored lines represent the results obtained for each configuration scenario indicated in Table 5. When the distance between the SCI trigger and LGD triggers reaches the highest value, scenario 11, light yellow colored line, the downlink BLER reaches the maximum value too.

These results serve to validate the RMPA approach in terms of the need to trigger the SCI link trigger as close as possible to the LGD trigger.

Performance comparison between RMPA handover versus an IEEE 802.16 HHO strategy SNR-based handover policy

In this section, we compare the performance of the RMPA handover and an IEEE 802.16 HHO strategy with a handover policy based on SNR, by measuring all packets arriving from the BS.

The mobile node travels at 100 m/s (360 km/h) along its trajectory. An error is considered in the distance source information as can be observed from Table 6.

As previously explained, the SCI trigger is calculated as $T_{SCNmax} = 20ms + T_{HOSIG} = 20ms + 3 * T_{frame}$.

Taking into consideration the GNSS classic receiver precision, we considered a 5-m error. The simulation scenario was run with different seeds.

Figure 16 illustrates the values for the instant at which the MS hands over the session to a new BS in an SNR policy-based scenario. The different colors represent the different seed values for a unique handover threshold value equal to 0.4 dB (default value in the specification). Owing to the variability in the downlink SNR profile shown in Figure 13, the handover takes place at a different place in each run.

However, Figure 17 depicts the moment at which the MS hands over the session to a new BS based on the RMPA handover policy. As can be seen, the instant values when using the SNR policy are more greatly dispersed. As a main conclusion, when the RMPA handover policy is used, and even when considering an error in the distance value, it is possible to force the LGD trigger to the midpoint, thereby reducing packet loss.

Figures 18 and 19 show the values obtained for different QoS performance KPIs, more specifically downlink packet dropped and throughput, with a traffic data load of 2048 K for both handover policies. In these figures,
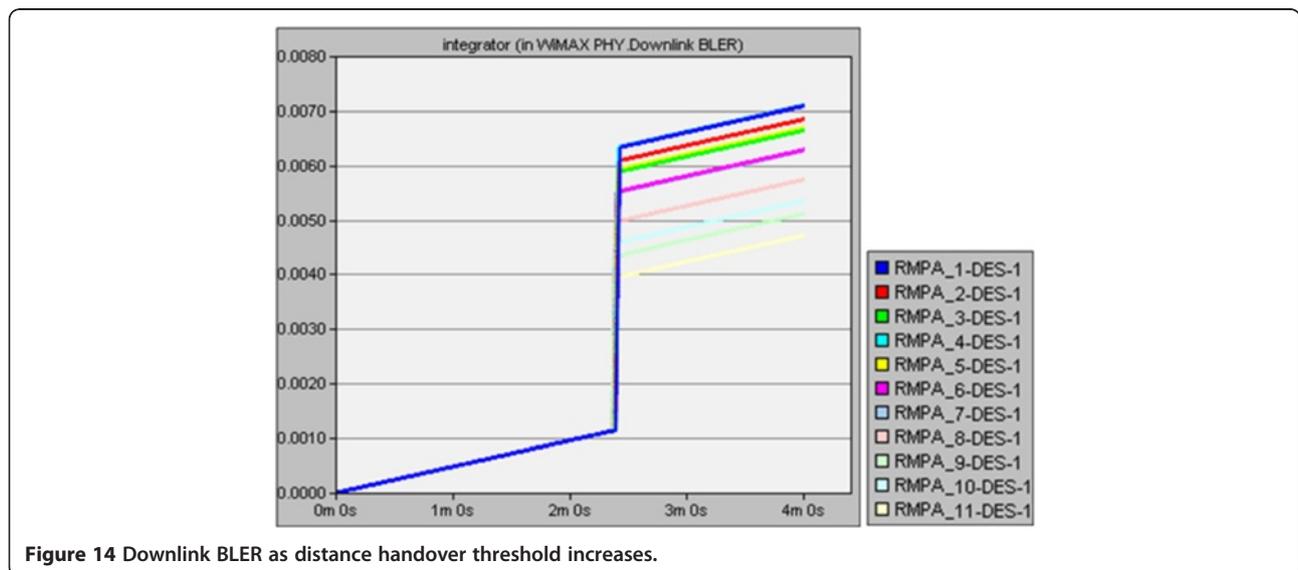

Figure 14 Downlink BLER as distance handover threshold increases.

Table 5 SCI and LGD trigger values for each simulation scenario

Scenario	1	2	3	4	5	6	7	8	9	10	11
SCI—Distance_Threshold1 scanning threshold (m)	200										
LGD—Distance_Threshold2 handover threshold (m)	100	90	80	70	60	50	40	30	20	10	1

the red values correspond to the RMPA handover results. As it can be seen, the throughput parameter is almost independent of the handover policy while the packet drop benefits from our RMPA policy.

To conclude the comparison of handover policies, we analyzed numerically the end-to-end KPIs obtained when the MS supports a CCTV video application in the uplink stream and in a longer trajectory with a higher amount of handover processes involved. The results are given in Table 7. Additional file 2 provides detailed information on statistic information for each simulation set. The MS speed was 360 km/h. It can be observed that practically all the KPIs improve when the RMPA policy is applied. The most outstanding enhancement is for the jitter or packet delay variation and the end-to-end delay parameter.

As a graphic representation of the previous results, we plotted the packet end-to-end and packet delay variation graphs in Figure 20, where the X-axis denotes simulation time. The red and blue lines represent the RMPA and SNR-based handover policies, respectively.

Conclusions

The handover process in railway communication systems is the most challenging process in end-to-end communication—especially for high-speed travel.

This article proposed a new handover strategy for the railway domain, namely, the RMPA handover, which is an IEEE 802.16 handover, specially "customized" for a high-speed mobility scenario. The stringent high-mobility feature is balanced with three other common features in a high-speed context: mobility pattern awareness, different sources for location discovery techniques, and a previously known traffic data profile.

The RMPA handover satisfies the design goals related to a cost-effective and standards-based solution. Compatibility with standards-based equipment is guaranteed since major contributions of the RMPA handover are in areas that have been left to the handover designer's discretion.

In terms of handover policies, the RMPA handover policy represents a step further when compared with traditional communication technologies currently in use for safety applications in the railway domain such as GSM-R. In these communication architectures, the handover decision is taken by the network and based on CINR measurements. In contrast, the RMPA handover policy takes into account the latest advances in location discovery techniques and makes use of a triple source of uncorrelated location information—the HTRU location technique. The decision to initiate handover is taken in the MS.

Taking into consideration safety railway services (train control applications) that may be supported by the communication architecture, redundancy features in the RMPA are considered in several aspects, i.e., a triple
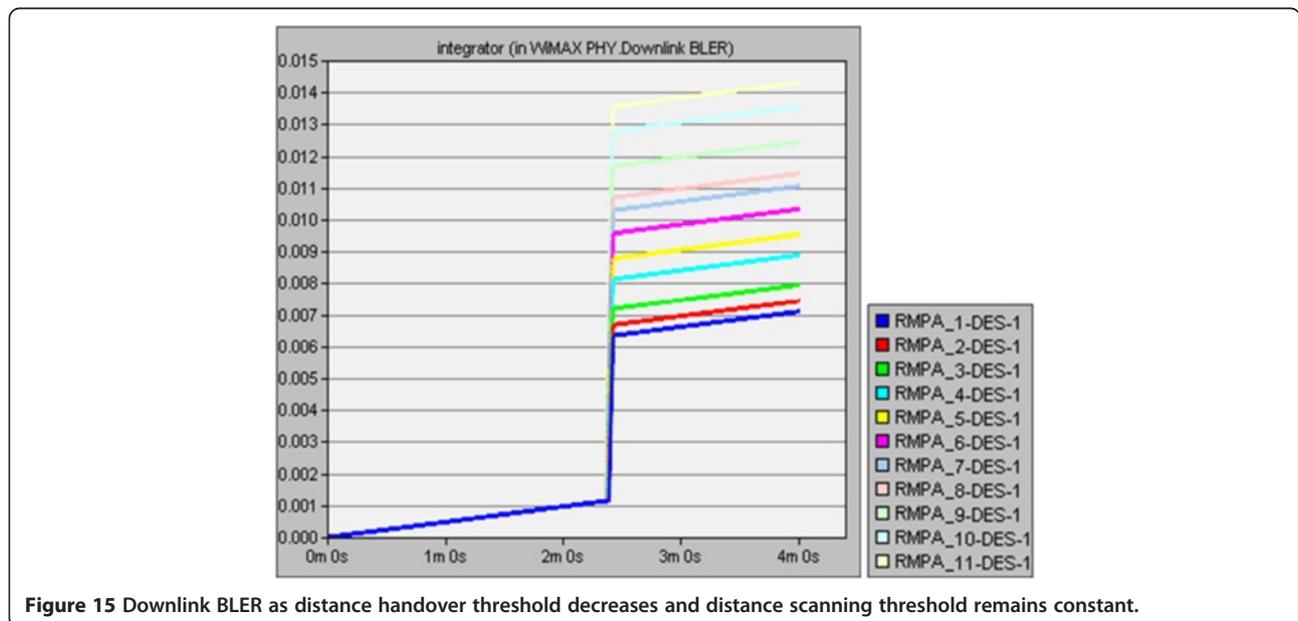

Figure 15 Downlink BLER as distance handover threshold decreases and distance scanning threshold remains constant.

Table 6 Details of values for threshold on RMPA and SNR

	RMPA handover	SNR handover
SCI—scanning threshold	35 ms*MS(v) + error = 345 m	50 dB
LGD—handover threshold	15 ms*MS(v) + error = 145 m	0.4 dB (default value)

neighborhood configuration, a neighboring advertising frequency conscious of the mobile node's dwell time, a bi-neighboring scanning policy with double iteration, and so on. Even though some steps are redundant, the results obtained show higher efficiency than traditional HHO strategies based on SNR measurements.

The simulation analysis carried out validates the handover decision rules used in the RMPA handover policy and its superior handover performance when compared with an SNR-based policy approach. Considering that the handover performance, particularly in high-speed environments, directly affects the end-to-end QoS KPIs of the applications supported by the MS, a highly demanding application such as a CCTV application with video in an uplink stream was used to further compare and validate the RMPA approach. The results show a significant improvement in the end-to-end parameters, including the end-to-end delay (22%) and jitter (80%), when compared with standards-based SNR approach.

To the best of the authors' knowledge, there is no IEEE 802.16 handover enhancement scheme that covers both research areas addressed in this study, that is, the optimization of the handover process itself and the efficient timing of the handover process in a high-speed scenario and, at the same time, represents a cost-effective and standards-based solution. Our RMPA strategy, with its configuration rules and handover policy

engine, the variables of which merge in a cross-layer interaction, achieves this goal.

Appendix A. IEEE 802.16 Handover process
This section describes the mechanism and procedures underlying the IEEE 802.16 handover process and the different stages in the handover process itself.

A.1. The network topology advertising and acquisition mechanisms
Before to perform a handover process, the mobile node has to be aware of the network topology so to be able to identify the possible BSs that are available and able to provide the access service. IEEE802.16 standard provides two complimentary mechanisms to achieve this.

A.2. Neighboring advertising strategy
BSs in IEEE802.16 standard are configured to belong to one or various neighborhoods within the network topology. Each BS may receive channel information related to their neighboring BSs over the backbone network. Each BS, similar to access point balises in wireless LAN, broadcasts the network topology relative to its neighborhood using the MOB_NBR_ADV message. A BS may belong to more than one neighborhood.

The MS receives this MOB_NBR_ADV message and creates and updates a list of neighboring BSs. The mobile node
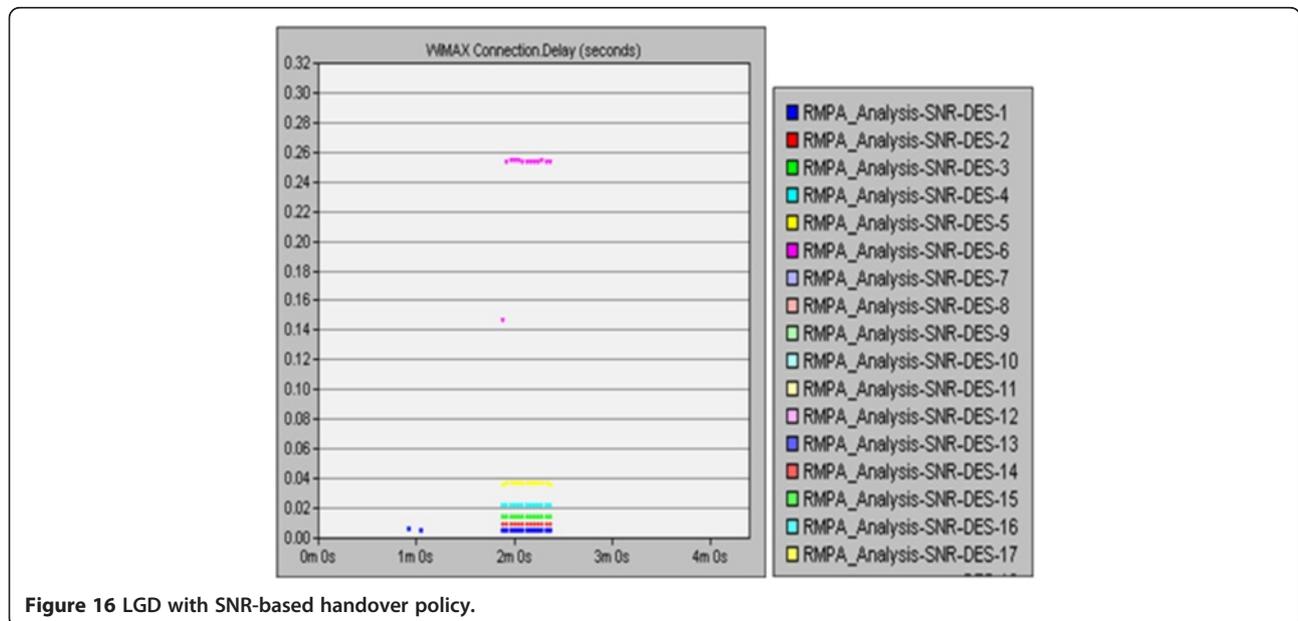

Figure 16 LGD with SNR-based handover policy.
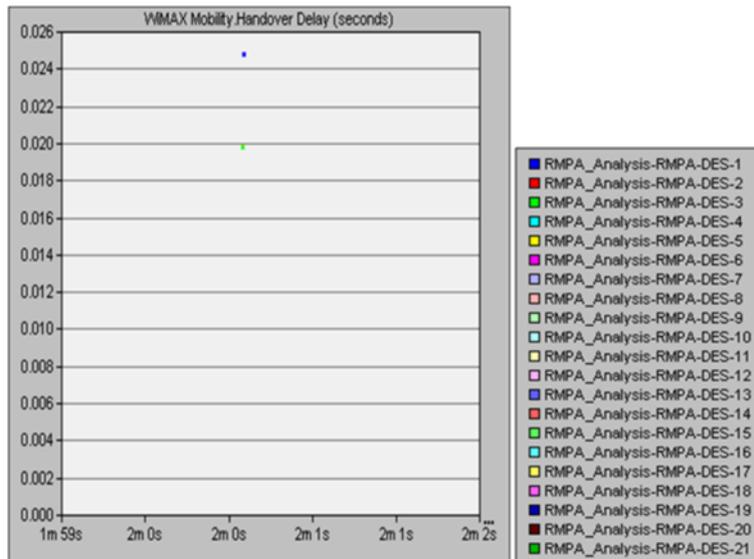
Figure 17 LGD with RMPA handover.

then will be aware of its target BSs for handover and scanning purposes. The information contained in MOB_N-BR_ADV message and relative to DCD and UCD settings for each neighbor BS facilitates MS synchronization with neighboring BS by removing the need to monitor transmission from the neighboring BS for DCD/UCD broadcasts messages, the most critical and time consuming step. The

fact of properly receiving on-time neighboring information enhances handover performance.

A.3. Cell reselection or scanning mechanism

The main purpose of this procedure is the MS node monitoring target BS nodes in the neighborhood and finding out if they are suitable for handover. The time
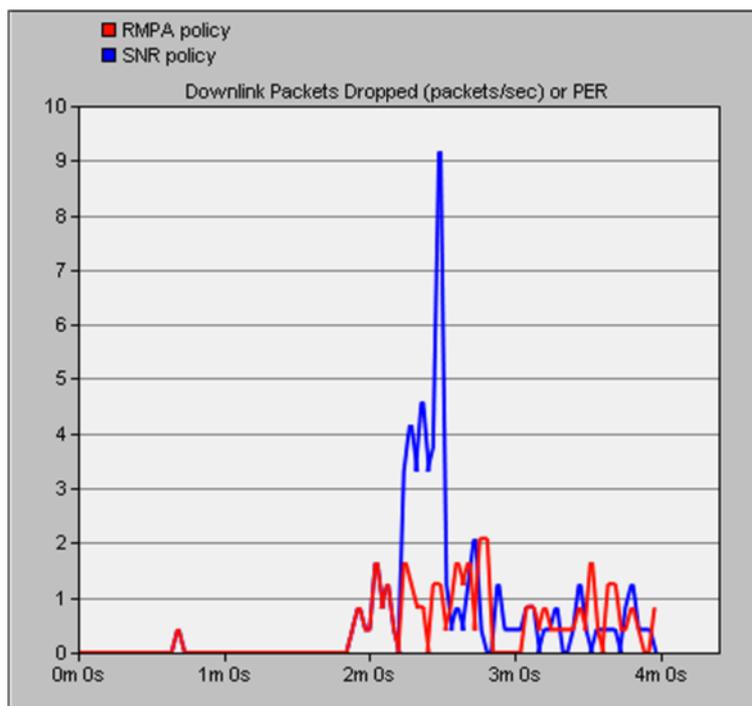
Figure 18 WiMAX delay and downlink packets dropped in both scenarios.
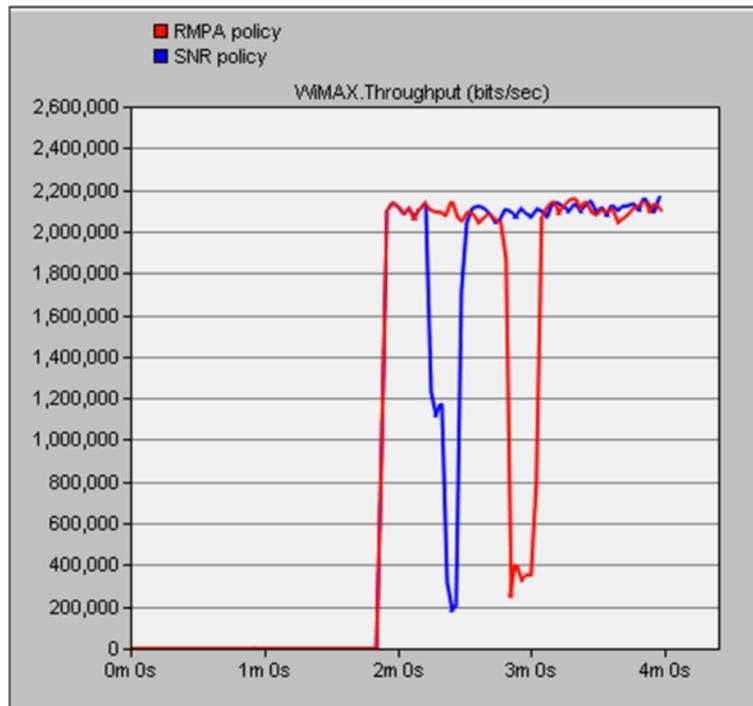
Figure 19 Throughput comparison.

during which the MS scans for available BS will be referred to as *scanning interval*.

Either the BS or the MS nodes may initiate this procedure. That is, a BS may allocate time intervals to MS for scanning purposes. In this case, a MOB_SCN_RSP is sent by the serving BS. This message includes information relative to: the start frame, the scanning process duration (N frames) the interleaving interval (P frames), normal data traffic, if any, is re-established, the iteration (T times).

Scanning interval and interleaving interval will repeat with the number of scan iteration.

A.4. Stages in the IEEE802.16 handover process

The IEEE802.16 handover process, in which a MS migrates from the air interface provided by one BS to the air interface provided by another BS, consists of the following

stages: normal or regular operation, cell reselection, handover decision, handover initiation, and handover execution.

### A.5. Stage 1: Normal or regular operation

The MS is connected to the serving BS and it is in packet scheduling process. The periodic ranging takes places during all the time the MS is connected to the serving BS. When an MS is assigned bandwidth, as in the ETCS application traffic profile, it is required to send messages even if there is no data to send. The BS can then monitor that the MS is responding to messages and can decide when it is considered out of the network.

A.6. Stage 2: Cell reselection

The main purpose of this stage is to collect information relative to the neighbor BSs. This information will contribute in the handover decision process. In this stage,

Table 7 Results obtained from comparing RMPA and SNR-based handover policies

CCTV (KPI)	KPI value	SNR handover 95% conf interval	RMPA handover 95% conf interval	(RMPA-SNR) mean/SNR mean%
Throughput (kbps)	>384 kbps	[4,404.17, 4,990.43]	[4,889.85, 5,597.67]	11.63
End-to-end delay (ms)	<60 ms preferred	[34.36, 35.59]	[28.58, 28.95]	21.58
Jitter	<20 ms	[0.5009, 0.5242]	[0.2807, 0.2891]	79.88
PER	<1%	[0.128, 0.134]	[0.128, 0.132]	0.4
WiMAX delay	<25 ms	[28.04, 29.58]	[23.83, 24.75]	18.89
Handover delay	<50 ms	[19.67, 24.61]	[19.17, 23.68]	3.36

Figure 20 End-to-end delay and jitter comparison.
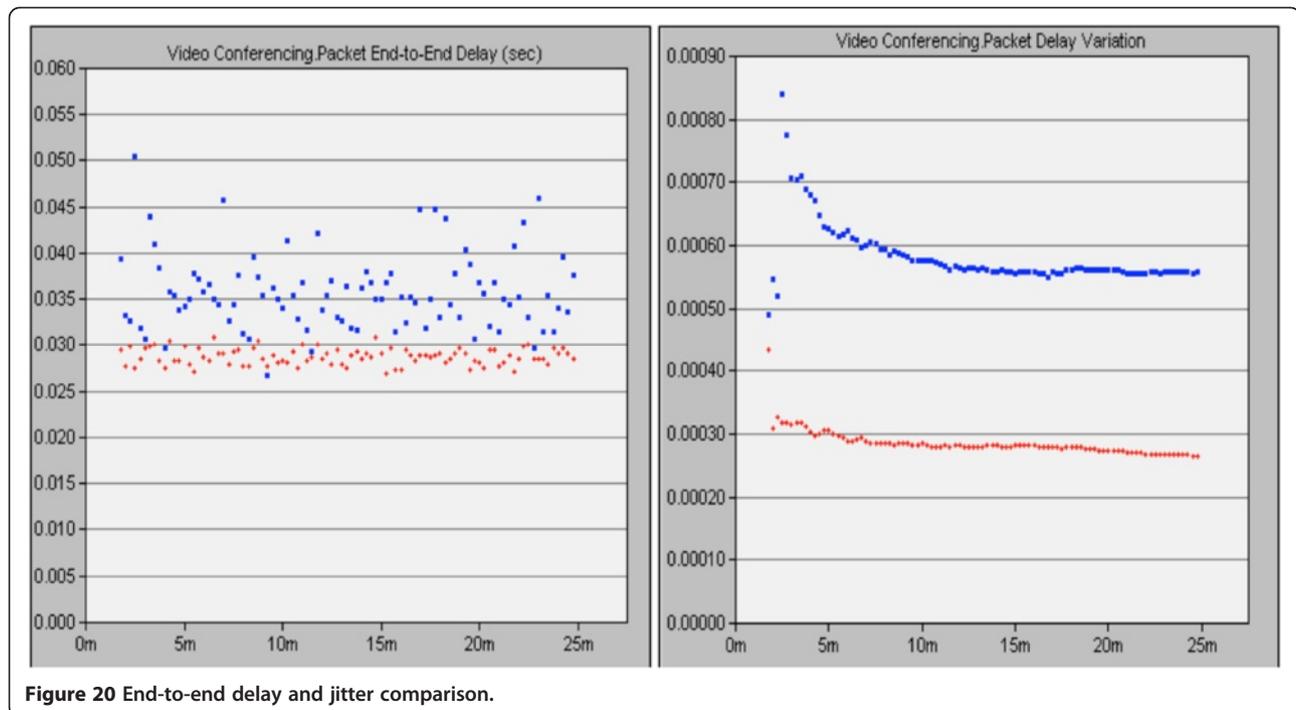
the MS uses neighboring BS information acquired from a decoded MOB_NBR-ADV message.

MS may also make a request to the serving BS to schedule scanning intervals for the purpose of evaluating MS's interest in handover to a potential target BS.

A.7. Stage 3: Handover decision

There are a set of the different concepts involved in the handover decision-making process: the triggers or events from the different layers, the handover policy, and the decision-making process. It is commonly understood as handover policies—a set of rules that contribute to shaping the handover decision for a mobile node. The handover decision policy itself, although it plays a pivotal role in handover performance, is not in the scope of the standard. It is implementation dependant.

A.8. Stage 4: HO initiation

In most IEEE802.16e implementations, the decision to initiate a handover process is typically taken by MS as a result of their own measurement of the quality of the signal from the neighboring BSs but it can also be initiated by the network (serving BS) under special circumstances. In accordance to the entity that takes the decision, this handover decision process consummates with a notification of MS intent to handover through MOB_MSHO-REQ message or MOB_BSHO-REQ message.

A.9. Stage 5: HO execution

Regarding the handover procedure or execution itself, the handover methods supported within the IEEE802.16 standard can be classified into hard and soft handover. The four supported methods are HHO, optimized HHO, FBSS, and MDHO. Out of these, just the HHO (break before make) is mandatory and it is the only type required to be implemented by mobile WiMAX, at the time of this writing.

Additional files

Additional file 1: Provides detail on simulation scenario configuration in this tool.

Additional file 2: Provides detailed information on statistic information for each simulation set.

Author details
[1]Faculty of Engineering, University of the Basque Country (UPV/EHU), Alameda de Urquijo s/n, Bilbao 48013, Spain. [2]University Lille Nord de France, F-59000, Lille IFSTTAR, LEOST, Villeneuve d'Ascq F-59650, France.

References
1. REPORT ITU-R M.2134, Requirements related to technical performance for IMT-Advanced radio interface(s). ITU-R (2008)
2. M Aguado, E Jacob, M Berbineau, I Lledó, *QoS and security challenges in the high mobility scenario: the handover process, in WiMAX Security and Quality of Service* (Providing and End-to-End Explanation, WILEY, 2010), pp. 213–239

3. SJ Yoo, D Cypher, N Golmie, Predictive link trigger mechanism for seamless handovers in heterogeneous wireless networks. Wirel. Commun. Mob. Comput (2008). Online at www3.interscience.wiley.com
4. K Park, K Jung, Modelling mobile WiMAX hand-over mechanisms and minimizing hand-over latency over lossy channels. (2007)
5. DT Fokum, VS Frost, A survey on methods for broadband internet access on trains. IEEE Commun. Surv. Tutor. **12**(2), 171–185 (2010)
6. ITU News Room, IMT-Advanced standards announced for next-generation mobile technology. (2012). http://www.itu.int/net/pressoffice/ press_releases/2012/02.aspx. Accessed March 2012
7. P Mahasukhon, H Sharif, M Hempel, T Zhou, T Ma, Distance and throughput measurements in mobile WiMAX test bed, in *Military Communications Conference.* MILCOM, 154–159 (2010)
8. I Al-Surmi, M Othman, B Mohd, Ali, Mobility management for IP-based next generation mobile networks: review, challenge and perspective. J. Netw. Comput. Appl. **35**(295–315) (2012)
9. DH Lee, K Kyamakya, UP Umondi, Fast handover algorithm for IEEE 802.16 e broadband wireless access system. IEEE Proceedings of the first International Symposium on Wireless Pervasive Computing (2006)
10. Y Wang, PHJ Chong, L Qiu, L Chen, E Lee, LC Seck, D Cheung, Research and software development of TETRA & TETRAPOL networks models for IP-based data services using OPNET. Proceedings of OPNETWORK (2006)
11. P Boone, M Barbeau, E Kranakis, Strategies for fast scanning and handovers in WiMAX/802.16, in *IEEE Proceedings of the 2nd International Conference on Access.* Networks **07**, 1–7 (2007)
12. R Rouil, N Golmie, *Adaptive channel scanning for IEEE 802.16e* (Proceedings of 25th Annual Military Communications Conference (MILCOM 2006), Washington, DC,), pp. 23–25. October 2006, pp. 1–6
13. S Choi, GH Hwang, T Kwon, AR Lim, DH Cho, *Fast handover scheme for real-time downlink services in IEEE 802.16 e BWA system*, 3rd edn. (IEEE 61st Vehicular Technology Conference VTC, 2005)
14. R Rouil, N Golmie, *Effects of IEEE 802.16 link parameters and handover performance for select scenarios*, 802nd edn. (IEEE,), pp. 21–26
15. SJ Yoo, D Cypher, N Golmie, LMS predictive link triggering for seamless handovers in heterogeneous wireless networks, in *IEEE Military Communications Conference.* MILCOM **2007**, 1–7 (2007)
16. S Woon, N Golmie, YA Sekercioglu, Effective Link Triggers to improve Handover performance. 2006 IEEE 17th International Symposium, Personal, Indoor and Mobile Radio Communications, 1–5 (2006)
17. D Lee, Y Han, J Hwang, QoS-based vertical handoff decision algorithm in heterogeneous systems. 2006 IEEE 17th International Symposium, Personal, Indoor and Mobile Radio Communications, 1–5 (2006)
18. KI Itoh, S Watanabe, JS Shih, T Sato, Performance of handoff algorithm based on distance and RSSI measurements. IEEE Trans. Veh. Technol. **51**(6), 1460–1468 (2002)
19. H Huang, W Hu, A fast handover scheme based on GPS information for IEEE 802.16e on high-speed railway. 2011 International Conference on Electronics, Communications and Control (ICECC), 2408–2412 (2011)
20. L Lu, X Fang, M Cheng, C Yang, W Luo, C Di, *Positioning and relay assisted robust handover scheme for high speed railway*, 73rd edn. (IEEE Vehicular Technology Conference, VTC – Spring, 2011)
21. TC Chen, JH Chiou, SW Wei, TT Lin, A handover algorithm for broadband wireless access system on railway. Mech. Electron. Eng. (ICMEE) **1**, 98–102 (2010)
22. H Suzuki, NEC Corporation, Japanese Patent 2007235541. **13** (2007)
23. S Willow, NEC Corporation. Japanese Patent **2007194754**, 2 (2007)
24. K Murakami, M Takenaka, T Kodama, Mitsubishi Electric Corporation. Japanese Patent **2007324635**, 13 (2007)
25. M Aguado, O Onandi, P Sáiz, M Higuero, E Jacob, *WiMAX on Rails: A Broadband Communication Architecture for CBTC Systems*, vol. 3, 3rd edn. (IEEE Vehicular Technology Magazine, New York, 2008), pp. 47–56
26. N Golmie, R Rouil, Seamless mobility in WiMAX, in *WiMAX Forum Conference*, 2007. (2007). www.antd.nist.gov/seamlessandsecure.shtml
27. J Marais, M Berbineau, M Heddebaut, Land mobile GNSS, availability and multipath evaluation tool. IEEE Trans. Veh. Technol. **54**(5), 1697–1704 (2005)
28. WiMAX System Evaluation Methodology 2.1. WiMAX Forum (2008)