Journal of Electrical and Computer Engineering

LTE/LTE-Advanced Cellular Communication Networks

Guest Editors: Cyril Leung, Raymond Kwan, Seppo Hämäläinen, and Wenbo Wang

LTE/LTE-Advanced Cellular Communication Networks

Journal of Electrical and Computer Engineering

LTE/LTE-Advanced Cellular Communication Networks

Guest Editors: Cyril Leung, Raymond Kwan, Seppo Hämäläinen, and Wenbo Wang

Editorial Board

The editorial board of the journal is organized into sections that correspond to the subject areas covered by the journal.

Circuits and Systems

Communications

Signal Processing

Contents

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2010, Article ID 659295, 2 pages
doi:10.1155/2010/659295

Editorial
LTE/LTE-Advanced Cellular Communication Networks

Cyril Leung,[1] Raymond Kwan,[2] Seppo Hämäläinen,[3] and Wenbo Wang[4]

[1] *Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada V6T 1Z4*
[2] *Centre for Wireless Network Design, University of Bedfordshire, Bedfordshire LU1 3JU, UK*
[3] *NSN Research, Nokia Siemens Networks, 02610 Espoo, Finland*
[4] *Key Laboratory of Universal Wireless Communication, Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China*

Correspondence should be addressed to Cyril Leung, cleung@ece.ubc.ca

Received 29 September 2010; Accepted 29 September 2010

From humble beginnings in the 1980s with bulky 1G analog handsets, the wireless cellular communications industry now enjoys a major presence in most countries, with nearly 5 billion subscribers worldwide and revenues in the one trillion USD range. Consumer demand for services is expected to grow rapidly in the next few years, fuelled by new applications such as mobile web-browsing, video downloading/streaming, online gaming, and social networking. The commercial deployment of 3G networks began with 3GPP UMTS/WCDMA in 2001 and has evolved into current UMTS/HSPA networks. To maintain a competitive edge, 3GPP UMTS networks need to support higher bit rates, improved spectrum efficiencies, and lower delays, all at a reduced cost. A well-planned and natural evolution to 4G networks is considered essential. Long-term evolution (LTE) and LTE-Advanced are important steps in this transition. LTE technology demonstrations began as early as 2006 and commercial LTE networks are starting to be deployed by wireless carriers worldwide.

This special issue consists of five papers. The first three papers focus on LTE: there is a survey paper on resource scheduling and interference mitigation techniques, a paper on resource scheduling for the provision of different services, and a paper on a method for improving the tradeoff between service quality and radio coverage for enhanced multimedia broadcast and multicast service (E-MBMS). The other two papers are focused on LTE-A: one paper discusses distributed algorithms for solving two self-configuration problems and the other examines the performance of decode-and-forward relaying.

Resource scheduling and interference mitigation will be instrumental in the success of LTE. In the first paper, "*A survey of scheduling and interference mitigation in LTE,*" Kwan and Leung present an overview of downlink and uplink resource scheduling techniques, including tradeoffs involved in selecting a channel quality indicator reporting scheme. A variety of methods which have been proposed for reducing intercell interference (ICI) in order to improve the link quality for cell-edge users are also described. The paper includes a brief discussion of the challenges for LTE-Advanced network designers.

The issue of joint radio resource allocation for multiple services is challenging, as each service is associated with different characteristics and quality of service requirements. In the second paper, "*Scheduling for improving system capacity in multiservice 3GPP LTE,*" Lima et al. address the issue of scheduling algorithms for multiple services in LTE, in which resources are jointly allocated. Their proposed scheme takes into account different traffic characteristics as well as current achieved satisfaction level of each service. Results show that the proposed scheme can provide significant capacity improvement over conventional approaches.

Video broadcast and multicast services are expected to play a very important role in future mobile communications. One of the key techniques for improving performance is the use of hierarchical modulation (HM). In HM, users with poorer channel qualities can access data via a low-rate but more error-protected "base-layer" (BL). On the other hand, users with better channel qualities can access the high-rate "enhanced layer" (EL) associated with less error protection.

The aim is to provide a good design tradeoff between service quality and coverage. In the third paper, "*Hierarchical modulation with vector rotation for E-MBMS transmission in LTE systems,*" Zhao et al. propose the use of vector rotation to improve the performance of the ELs in the context of LTE multicast/broadcast networks. Simulation results show that the proposed method can provide significant benefits for video quality and coverage compared to conventional schemes.

In LTE-Advanced systems, the number of available primary component carriers (PCCs) and physical cell IDs (PCIs) are limited and their assignment can be formulated as graph coloring problems. In the fourth paper, "*Distributed graph coloring for self-organization in LTE networks,*" Ahmed et al. investigate algorithms for solving these problems in a distributed fashion, considering both real-valued and binary pricing of interference. The number of component carriers needed to select conflict-free PCCs and the number of PCIs needed to assign confusion-free PCIs are estimated from simulations.

Decode-and-forward relaying is likely to play an important role in LTE-Advanced networks. In the fifth paper, "*On the coverage extension and capacity enhancement of inband relay deployments in LTE-Advanced networks,*" Saleh et al. investigate the coverage extension capabilities and capacity enhancements achievable with in-band Type 1 and Type 1b relaying within the LTE-Advanced framework in different propagation scenarios. System level simulation results show that the gains from employing these two types of relays vary greatly, depending on the deployment environment.

It is evident that this special issue does not cover many other important areas of relevance to LTE/LTE-Advanced networks. Nonetheless, we hope that readers will find the information presented to be interesting and useful. We thank all the authors who responded to the call for papers. We also wish to acknowledge all the reviewers for their dedicated efforts in ensuring a high standard for the selected papers.

Cyril Leung
Raymond Kwan
Seppo Hämäläinen
Wenbo Wang

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2010, Article ID 273486, 10 pages
doi:10.1155/2010/273486

Research Article

A Survey of Scheduling and Interference Mitigation in LTE

Raymond Kwan[1] and Cyril Leung[2]

[1] *Centre for Wireless Network Design, University of Bedfordshire, Park Square,, Luton LU1 3JU, UK*
[2] *Department of Electrical and Computer Engineering, University of British Columbia, 2332 Main Mall, Vancouver, BC, Canada V6T 1Z4*

Correspondence should be addressed to Raymond Kwan, raymond.y.c.kwan@gmail.com

Received 10 April 2010; Accepted 30 May 2010

Academic Editor: Wang Wenbo

Among the goals of 3GPP LTE networks are higher user bit rates, lower delays, increased spectrum efficiency, support for diverse QoS requirements, reduced cost, and operational simplicity. Resource scheduling and interference mitigation are two functions which are key to achieving these goals. This paper provides a survey of related techniques which have been proposed and shown to be promising. A brief discussion of the challenges for LTE-Advanced, the next step in the evolution, is also provided.

1. Introduction

The demand for cellular communication services is expected to continue its rapid growth in the next decade, fuelled by new applications such as mobile web-browsing, video downloading, on-line gaming, and social networking. The commercial deployment of 3G. Cellular network technologies began with 3GPP UMTS/WCDMA in 2001 and has evolved into current UMTS/HSPA networks. To maintain the competitiveness of 3GPP UMTS networks, a well-planned and graceful evolution to 4G networks [1] is considered essential. LTE is an important step in this evolution, with technology demonstrations beginning in 2006. Commercial LTE network services started in Scandinavia in December 2009 and it is expected that carriers worldwide will shortly be starting their upgrades.

The main design goals behind LTE are higher user bit rates, lower delays, increased spectrum efficiency, reduced cost, and operational simplicity. The first version of LTE, 3GPP Release 8, lists the following requirements [2]. (1) peak rates of 100 Mbps (downlink) and 50 Mbps (uplink); increased cell-edge bit rates; (2) a radio-access network latency of less than 10 ms; (3) two to four times the spectrum efficiency of 3GPP Release 6 (WCDMA/HSPA); (4) support of scalable bandwidths, 1.25, 2.5, 5, 10,15, and 20 MHz; support for FDD and TDD modes; smooth operation with and economically viable transition from existing networks. In order to meet these demanding requirements, LTE makes use of multiantenna techniques and intercell interference coordination.

In this paper, we provide a survey of radio resource scheduling and interference mitigation in LTE. Both are widely recognized as areas which can greatly affect the performance and spectrum efficiency of an LTE network. The paper is organized as follows. In Section 2, a brief overview of some aspects of LTE, necessary to discuss scheduling in a meaningful way, is given. Radio resource scheduling methods proposed are discussed in Section 3. Methods based on interference mitigation to improve the QoS of cell-edge users are described in Section 4. Section 5 provides a brief discussion of LTE-Advanced, as the next step towards a 4G network.

2. Overview of LTE PHY

A brief description of LTE Physical Layer features, necessary to discuss scheduling algorithms, is now provided. It should be noted that scheduling decisions are made at the base station (eNB in 3GPP parlance) for both downlink and uplink radio transmissions. More details can be found in [3–5].

2.1. Downlink. LTE radio transmission on the downlink is based on OFDM, a modulation scheme that is used in a variety of wireless communication standards. OFDMA, a

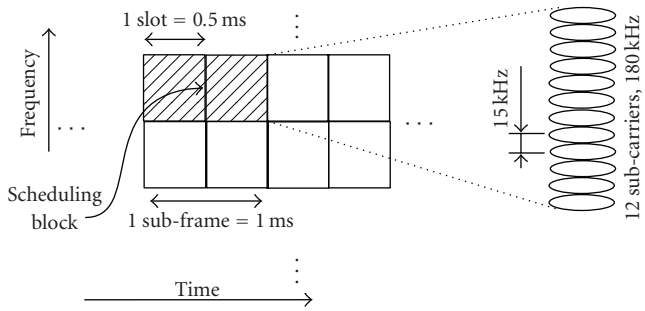


Figure 1: Illustration of a scheduling block in LTE downlink.

variant of OFDM which allows several users to simultaneously share the OFDM subcarriers, is employed in LTE in order to take advantage of multiuser diversity and to provide greater flexibility in allocating (scheduling) radio resources. Specifically, the OFDM subcarriers are spaced 15 kHz apart and each individual subcarrier is modulated using QPSK, 16-QAM, or 64-QAM following turbo coding. AMC is used to allow the optimal MCS to be chosen, based on current channel conditions.

A major difference between packet scheduling in LTE and that in earlier radio access technologies, such as HSDPA, is that LTE schedules resources for users in both TD and FD, whereas HSDPA only involves TD. This additional flexibility has been shown to provide substantial throughput and coverage gains. In order to make good scheduling decisions, a scheduler requires knowledge of channel conditions. Ideally, at each scheduling time, the scheduler should know the channel gain for each sub-carrier and each user. However, due to limited signalling channel resources, subcarriers are grouped into RBs, each consisting of 12 adjacent subcarriers. Each RB has a time slot duration of 0.5 ms, which corresponds to 6 or 7 OFDM symbols depending on whether an extended or normal cyclic prefix is used. The smallest resource unit that a scheduler can allocate to a user is an SB, which consists of two consecutive RBs, spanning a subframe time duration or TTI of 1 ms and a bandwidth of 180 kHz (see Figure 1).

2.2. Uplink. Except for the use of SC-FDMA, a precoded version of OFDM, in place of OFDMA, the PHY for the uplink is similar to that for the downlink, especially for the FDD mode. SC-FDMA is preferred over OFDMA since SC-FDMA has a smaller PAPR which allows a reduced power consumption. This is an important consideration for the UE (user/mobile terminal) and results in improved coverage and cell-edge performance. Another advantage is a simpler amplifier design, which translates to a lower-cost UE.

3. Scheduling

Due to the central role of the scheduler in determining the overall system performance, there have been many published studies on LTE scheduling [6–41]. Simply stated, the scheduling problem is to determine the allocation of SBs to a subset of UEs in order to maximize some objective function, for example, overall system throughput or other fairness-sensitive metrics. The identities of the assigned SBs and the MCSs are then conveyed to the UEs via a downlink control channel.

To reduce complexity, most schedulers operates in two phases: TDPS followed by FDPS. The TD PS creates a SCS which is a list of users which may be allocated resources in the current scheduling period and the FD PS determines the actual allocation of SBs to users in the SCS. Since the TD PS does not concern itself with the actual allocation of SBs, it typically uses only average full-band subcarrier CQI and not individual subcarrier CQI. The TD and FD packet schedulers can be designed using possibly different metrics depending on the characteristics desired of the scheduler, for example, high throughput, fairness, low packet drop rate, and so forth.

We discuss scheduling on the downlink in Section 3.1 and scheduling on the uplink in Section 3.2.

3.1. Downlink Scheduling. By taking advantage of the time and frequency channel quality variations of the SBs associated with different users, a scheduler can greatly improve average system throughput compared to a round-robin scheduler which is blind to these variations. The basic idea is to allocate each SB to the user who can best make use of the SB according to some utility function. Scheduling with a single-carrier-based downlink access technology, such as 3GPP Release 6 HSDPA, can only be time-opportunistic. In contrast, OFDMA schedulers can be both time- and frequency-opportunistic.

The performances of different aspects of FDPS in a variety of scenarios have been studied in [6–14]. Results in [7], obtained using a detailed link simulator, indicate that a PF FD PS can provide average system throughput and cell-edge user bit rate gains of about 40% compared to a time-opportunistic scheduler which does not use subcarrier CQI information. As to be expected, the magnitudes of the gains depend on many factors including CQI accuracy, CQI frequency resolution and the number of active users. A CQI error standard deviation of 1 dB, a frequency resolution similar to the channel coherence bandwidth and a SCS size of at least 5 are found to be adequate.

In [9], the spectral efficiencies and cell-edge user throughputs for three different packet scheduler combinations, namely, TD-BET/FD-TTA, TD-PF/FD-PF, and TD-MT/FD-MT, are compared with those for a reference round-robin scheduler with one user scheduled per SP. The benefits of multiuser diversity and channel-dependent scheduling are illustrated. The throughput performances of an optimal and a suboptimal FD PS are studied as a function of subcarrier correlation in [10]. An interesting observation is that the throughput tends to increase with correlation. A scheduler which takes into account the status of UE buffers, in addition to channel conditions, is proposed in [11] to reduce PLR due to buffer overflow while maintaining a high system throughput and good fairness among users. Significant improvements are reported relative to the RR, PO, and PF schedulers. The design of a low-complexity queue- and channel-aware scheduler which can support QoS for

different types of traffic is presented in [20]. The use of decoupled TD and FD metrics for controlling throughput fairness among users is investigated in [18]. Simulation results show that a large gain in coverage can be obtained at the cost of a small throughput decrease. The authors suggest the use of FD-PF in LTE, complemented by TD-PSS or FD metric weighting depending on the number of active UEs in the cell. In [12], the impact of downlink CC signaling overhead and erroneous decoding of CC information by UEs on scheduler performance is examined.

MBMS provides a means for distributing information from the eNB to multiple UEs. A group of MBMS subscribers listen to a common channel, sharing the same SBs. In order to meet the BLER target, the MCS has to be selected based on the UE with the poorest SB channel quality in the group. A MBMS FD PS is proposed in [17] which attempts to allocate SBs so as to improve the channel quality of the poorest quality UE in the group. Simulation results are presented to illustrate the effectiveness of the proposed scheduler.

There are few studies on schedulers designed to operate in combination with MIMO techniques [21, 22]. A MIMO aware FDPS algorithm, based on the PF metric, is presented in [21]. Both single-user and multiuser MIMO cases are treated. In the single-user case, only one user can be assigned to one time-frequency resource element whereas in the multiuser case, multiple users can be assigned on different streams on the same time-frequency element. A finite buffer best effort traffic model is assumed and a 1×2 maximal ratio combining case is used as reference. Simulations results show that in a macrocell scenario, the average cell throughput for MIMO without precoding does not improve relative to the reference case; with precoding, a gain of 20% is observed. The throughput gains in a microcell scenario are significantly larger due to the larger SINR dynamic range.

3.2. Uplink Scheduling. Similar to OFDMA schedulers used on the DL, SC-FDMA schedulers for the UL can be both time- and frequency-opportunistic. An important difference between DL and UL scheduling is that SB CQI reporting is not needed for UL since the scheduler is located at the eNB which can measure UL channel quality through SRS.

Scheduling algorithms for the LTE UL have been discussed by many authors. In [24], a heuristic localized gradient algorithm is used to maximize system throughput given the constraint that subcarriers assigned to a specific user must be contiguous. This work is extended to dynamic traffic models and finite UE buffers in [25]. A variety of channel-aware schedulers with differing optimization metrics and computational complexities are described in [26–30, 32, 34, 35].

A scheduler which minimizes transmit power subject to a bound on average delay is treated in [31]. The scheduler chooses the power and transmission rate based on channel and UE buffer states. It is observed that substantial reductions in transmit power can be achieved by allowing small increases in average delay.

To provide the QoS differentiation needed to support diverse services such as VoIP, web browsing, gaming, and e-mail, a combined AC/PS approach is proposed in [36]. Both AC and PS are assumed to be QoS- and channel-aware. The AC decides to accept or deny a new service request based on its channel condition and whether its QoS can be satisfied without compromising the required QoS of ongoing sessions. The PS dynamically allocates resources to ongoing sessions to meet their required QoS. Simulation results are presented to illustrate the effectiveness of the AC/PS in meeting the QoS requirements of different types of users in a mixed traffic scenario.

3.3. VoIP. Even though the major growth will be in data services, telephony will continue to represent a significant portion of the traffic carried in LTE networks. Voice applications will be supported using VoIP. Scheduling for VoIP is discussed in [37–41].

The nature of the carried traffic plays an important role in how scheduling should be done. VoIP users are active only half of the time and VoIP traffic is characterized by the frequent and regular arrival of short packets. VoIP service also has tight packet delay and PLR requirements. While *dynamic scheduling* based on frequent downlink transmit format signalling and uplink CQI feedback can exploit user channel diversity in both frequency and time domains, it requires a large signalling overhead. This overhead consumes time-frequency resources, thereby reducing the system capacity.

In order to reduce signalling overhead for VoIP-type traffic, *persistent scheduling* has been proposed. The idea behind persistent scheduling is to preallocate a sequence of frequency-time resources with a fixed MCS to a VoIP user at the beginning of a specified period. This allocation remains valid until the user receives another allocation due to a change in channel quality or an expiration of a timer. A big disadvantage of such a scheme is the lack of flexibility in the time domain which may result in a problematic difference between allocated and actually needed resources. To reduce the wastage when a VoIP call does not make full use of its allocated resources, a scheduling scheme is proposed in [40] which dynamically allows two "paired" VoIP users to share resources.

In [37], it is noted that meeting the QoS requirements of VoIP packets by giving these packets absolute priority may negatively impact other multimedia services and degrade the overall system performance. A scheduling scheme which dynamically activates a VoIP packet priority mode (to satisfy voice QoS requirements) and adjusts its duration based on VoIP PLRs (to minimize system performance degradation) is proposed. Simulation results are provided to demonstrate the effectiveness of the scheme.

Semipersistent scheduling, which represents a compromise between rigid persistent scheduling on the one hand, and fully flexible dynamic scheduling on the other, has been proposed by several authors. In [38], initial transmissions are persistently scheduled so as to reduce signalling overhead and retransmissions are dynamically scheduled so as to provide adaptability. Another semipersistent scheduling scheme based on that in [40] is studied in [41].

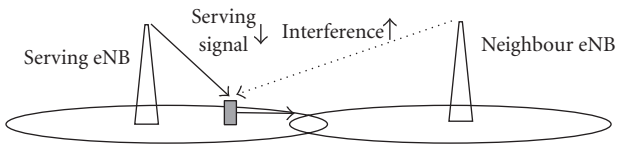


Figure 2: Illustration of UE moving away from the serving eNB.

3.4. Channel Quality Indicator. It is well documented that an accurate knowledge of channel quality is important for high-performance scheduler operation. From the perspective of downlink scheduling, the channel quality is reported back to the eNB by the UE (over the uplink) using a CQI value. If a single "average" CQI value is used to describe the channel quality for a large group of SBs, the scheduler will be unable to take advantage of the quality variations among the subcarriers. This may lead to an unacceptable performance degradation for frequency-selective channels. On the other hand, if a CQI value is used for each SB, many CQI values will need to be reported back, resulting in a high signalling overhead.

Several CQI reporting-schemes and associated trade-offs are discussed in [10, 15, 16]. Two broad classes of reporting schemes can be identified. In the first class, schemes report a compressed version of the CQI information for all SBs to the eNB. In the second class, schemes report only a limited number of the SBs with the highest CQI values.

The dependence of the system throughput on the number of SBs whose CQI values are reported by UEs is discussed in [10]. In [15], a "distributed-Haar" CQI reporting scheme is proposed and simulation results are reported which show that it can provide a better trade-off between throughput performance and CQI feedback signaling overhead, compared to a number of other CQI-compression schemes. The effect of varying FD and TD granularity in CQI reporting is discussed in [16]. It is concluded that a CQI measurement interval of 2 ms and a frequency resolution of 2 SBs per CQI is adequate in most cases. The spectral efficiencies for four different CQI reporting schemes are also compared in [16] and it is suggested that the *Best-M average* and *Threshold-based* schemes provide a good trade-off between system performance and UL signaling overhead.

4. Interference Mitigation

ICI is a major problem in LTE-based systems. As the cell-edge performance is particularly susceptible to ICI, improving the cell-edge performance is an important aspect of LTE systems design.

As a UE moves away from the serving eNB, the degradation in its SINR can be attributed to two factors. On the one hand, the received desired signal strength decreases. On the other hand, ICI increases as the UE moves closer to a neighbouring eNB, as illustrated in Figure 2.

One approach for reducing ICI from neighbouring eNBs is to use *enhanced frequency reuse* techniques. To motivate such techniques, we first take a look at conditions under which *conventional* frequency reuse can be beneficial.

Let $\gamma_{i,j}^{(n)}$ be the SINR of UE i which is served by eNB j, based on an FRF n configuration. The resulting spectral efficiency is given by

$$C_n = \frac{1}{n}\log_2\left(1+\gamma_{i,j}^{(n)}\right). \tag{1}$$

To assess the benefits of conventional frequency reuse, it is convenient to define the spectral efficiency improvement factor as

$$\Delta_n \triangleq \frac{C_n}{C_1}. \tag{2}$$

Thus, Δ_n is a measure of the spectral efficiency improvement that is obtainable with a FRF n configuration compared to full reuse, that is, $n = 1$. For the FRF n configuration to have a higher spectral efficiency than the full reuse case, we require that $\Delta_n > 1$. Using (1) and (2), it can be shown that this requirement is equivalent to

$$\gamma_{i,j}^{(n)} > n\gamma_{i,j}^{(1)} + \binom{n}{2}\left(\gamma_{i,j}^{(1)}\right)^2 + \cdots + \left(\gamma_{i,j}^{(1)}\right)^n. \tag{3}$$

In the case of a UE which is close to its serving eNB, hereafter referred to as a *cell-center* UE, the SINR value is usually high, that is, $\gg 1$, and (3) can be approximated as

$$\gamma_{i,j}^{(n)} > \left(\gamma_{i,j}^{(1)}\right)^n. \tag{4}$$

Thus, for cell-center UEs, an SINR increase in the order of the nth power of $\gamma_{i,j}^{(1)}$ must be achieved in order to justify a FRF n scheme. On the other hand, for cell-edge UEs for which $\gamma_{i,j}^{(1)} \ll 1$, (3) can be approximated by

$$\gamma_{i,j}^{(n)} > n\gamma_{i,j}^{(1)}. \tag{5}$$

While it is possible to obtain a large SINR improvement for cell-edge UEs by eliminating ICI, the SINR improvement will be smaller for cell-center UEs, as ICI is not as dominant for these UEs. Also, since a cell-center UE SINR is typically much higher than unity, the spectral efficiency increases only logarithmically with SINR. However, for a cell-edge UE with SINR value much less than unity, the spectral efficiency increases almost linearly with SINR. From (4) and (5), we can conclude that, in general, only cell-edge UE spectral efficiency will improve from using a high FRF value.

4.1. Fractional Frequency Reuse. The above discussion indicates that it is not bandwidth-efficient to use the same FRF value for the entire cell. One way to improve the cell-edge SINR while maintaining a good spectral efficiency is to use an FRF greater than unity for the cell-edge regions and an FRF of unity for the cell-center regions [42]. Such an FFR scheme is illustrated in Figure 3. In this example, the entire bandwidth is divided into four segments, that is, $\{f_1, f_2, f_3, f_4\}$. The idea is to apply a unity FRF scheme using f_1 for cell-center regions, whereas f_2, f_3, and f_4 are used for an FRF of 3 configuration for the cell-edge regions.

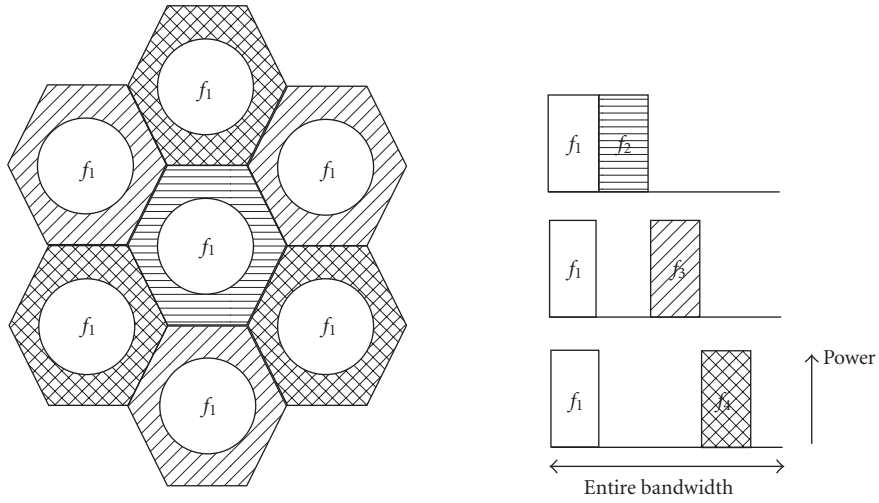


Figure 3: Illustration of FFR scheme.

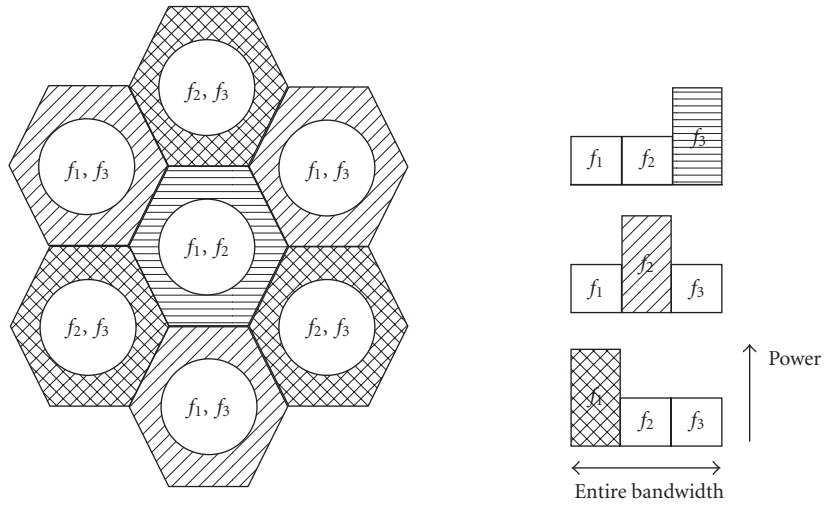


Figure 4: Illustration of SFR scheme.

4.2. Soft Frequency Reuse. In FFR, adjacent cell-edge regions do not share the same frequency segments, thereby resulting in a lower ICI. However, this strict no-sharing policy may under-utilize available frequency resources in certain situations. In order to avoid the high ICI levels associated with a unity FRF configuration, while providing more flexibility to the FFR scheme, an SFR scheme has been proposed, in which the entire bandwidth can be utilized [42]. As in FFR, the available bandwidth is divided into orthogonal segments, and each neighboring cell is assigned a cell-edge band. In contrast to FFR, a higher power is allowed on the selected cell-edge band, while the cell-center UEs can still have access to the cell-edge bands selected by the neighbouring cells, but at a reduced power level. In this way, each cell can utilize the entire bandwidth while reducing the interference to the neighbours. An example SFR scheme is illustrated in Figure 4. Note that this scheme can improve the SINR of the cell-edge UEs using a greater than unity FRF, while degrading the SINR of the cell-center UEs. This degradation is due to the overlap in frequency resources between the cell-edge band of the neighbouring cells, and the cell-center band of the serving cell. However, as mentioned earlier, the cell-edge performance improvement is almost linear while the degradation to the cell-center UEs is logarithmic. In SFR, the power ratio between the cell-edge band and the cell-center band can be an operator-defined parameter, thereby increasing the flexibility in system tuning. The parameters for both FFR and SFR can be varied semistatically, depending on the traffic loads and user channel qualities.

Variations can be made to the above reuse schemes. One example is given in [43–45]. Instead of using a fraction of the bandwidth as shown in Figure 4, the full bandwidth can be utilized for the cell-center UEs, at the expense of higher ICI to the cell-edge UEs. An example of a reuse 3/7 scheme [46, 47] is shown in Figure 5, where frequency reuse is applied to the cell-edge UEs of the respective sectors.

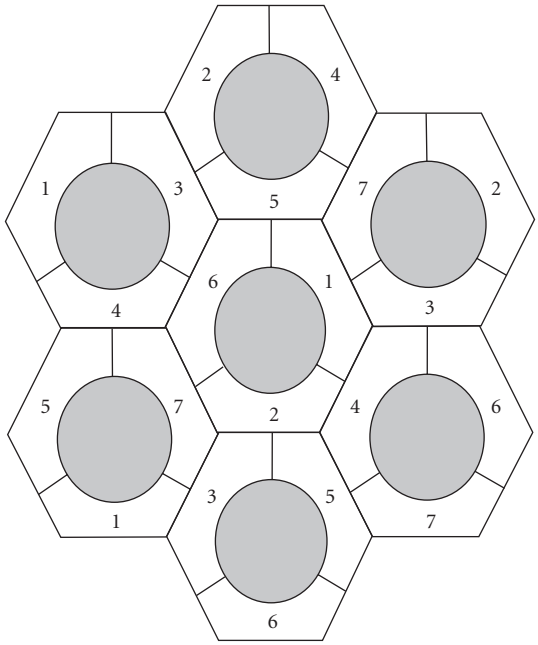


Figure 5: Illustration of SFR 3/7 scheme.

4.3. Adaptive Frequency Reuse Schemes. One of the main assumptions behind the FFR and SFR schemes is that the traffic load within each cell remains stable throughout the life-time of the deployment. However, in practice, this is not always the case, and further improvements may result in better resource utilization and system performance. In [48], simulation results suggest that a fixed FFR or SFR may not provide an improvement over the full reuse scheme, and that a more dynamic reuse scheme may be beneficial.

In [49], the entire bandwidth is divided into N subbands, and X subbands are used for the cell-edge UEs, and $N - 3X$ subbands are allocated for the cell-center UEs. The X subbands between neighbouring cell edge regions are orthogonal, while the $N - 3X$ subbands are the same for all cells. The value for X can be adjusted based on the traffic load ratio between the cell-edge and the cell-center regions. In [50], an improvement on the above scheme is made, which not only takes into the traffic ratio between the edge and the center regions of the serving cell, but also of the neighbouring cells. For example, when a cell detects that heavy traffic is present at the cell edge, but not for a subset $\mathcal{C}_n$ of the neighbouring cells, more frequency resource can be allocated to the edge UEs by borrowing some edge resources from cells within $\mathcal{C}_n$. For this scheme, some form of intercell communication may be required in order for a cell to assess the edge traffic loads of its neighbouring cells.

Recently, messages for inter-eNB communications have been standardized for LTE Release 8 to allow different eNBs to exchange interference-related information. For example, an Uplink HII message is introduced per PRB, to indicate the occurrence of high interference sensitivity of these PRBs as experienced from the sending eNB [51]. The receiving eNB(s) would then try to avoid allocating these PRBs to the cell-edge UEs. In [52], the cell-edge bandwidth is allocated dynamically according to the traffic load. In this scheme, each cell maintains two disjoint sets of PRBs, one for the cell-edge region and one for the cell-center region. When a cell-edge region load is high, the cell compares the identities of its own reserved cell-edge PRBs and those included within the neighbouring HII(s), in order to determine the "borrowable" cell-edge PRBs, and select the PRBs that would minimize ICI. In order to avoid intracell interference, the borrowed PRBs are not assigned to the cell-center UEs.

In [53], a flexible frequency reuse scheme is proposed, in which the entire bandwidth is divided into the cell-center band and the cell-edge band(s). In this scheme, the cell-center band can "borrow" a fraction of the cell-edge bands, thereby providing a way to adapt the bandwidth according to the nonhomogeneous nature of the traffic load within a cell. A similar load-based frequency reuse scheme is discussed in [54].

In [55], a decentralized, game-theoretic approach is proposed, whereby each eNB iteratively selects a set of PRBs that minimizes its own perceived interference from its neighbours. At each allocation instance, a new set of PRBs is selected with a certain probability if the perceived interference can be reduced by a specified threshold.

4.4. Graph-Based Approach. The problem of interference mitigation can be formulated as an interference graph in which the UEs correspond to the nodes and relevant interference relations between UEs correspond to the respective edges. To minimize interference, connected UEs should not be allocated the same set of resources. Such a problem is directly related to the graph coloring problem in which each color corresponds to a disjoint set of frequency resources. The goal is for each node in a graph to be assigned a color in such a way that no connected nodes are assigned the same color. In [56], a centralized graph coloring approach is proposed, where a "generalized" frequency reuse pattern is assigned to UEs at the cell edge by a centralized coordinator. In order to generate the interference graph, UEs are required to measure the interference and path losses, and report back to their respective eNBs. Subsequently, every eNB sends the necessary information to the central coordinator, which then generates the interference graph and performs the necessary optimization. Note that the definition of the edges of a graph is model- and problem-specific. Other graph-based approaches can be found in [57, 58].

4.5. Interactions with Scheduler. It is important to note that the potential benefits of the FFR or SFR schemes may not always be realized [45, 48, 59, 60]. On the one hand, ICIC can provide interference reduction for cell-edge users, thereby improving the cell-edge user bit rate. On the other hand, in a unity FRF system, the reduced SINR due to ICI can be compensated by allocating more bandwidth to these users. By compensating for the lower SINR with more bandwidth, that is, resource blocks, to suitable users using an intelligent scheduler, a good performance can still be obtained, even without ICIC. In this case, the benefit of ICIC may not justify the added complexity of intercell communications

[59]. However, it is pointed out in [60] that if a limit is imposed on the peak rate that can be used for a particular bearer, the benefit of bandwidth *compensation* is not fully realized. It is found that ICIC is useful mostly for low-to-moderate traffic loads. A similar observation is made in [61].

The frequency of CQI feedback provides a trade-off between temporal diversity gain and the effective interference reduction achieved by a frequency-selective scheduler. A more frequent CQI feedback provides a good temporal adaptation to the channel variation, thereby improving the spectral efficiency. However, the frequency of CQI feedback should also depend on the short-term interference due to the instantaneous scheduling decisions of the neighboring cells. Since this interference is weakly correlated with that in subsequent subframes, the usefulness of the CQI feedbacks, especially those from the cell-edge users, is greatly reduced. Thus, from the point of view of interference mitigation, a less frequent scheduling decision would be useful. It is suggested in [62] that CQI feedback filtering should be performed in order to average out the temporal variations of the ICI, thereby providing a more stable CQI value for the cell-edge users.

5. Conclusion

A high-level survey of works on resource scheduling and interference mitigation in 3GPP LTE was presented. These two functions will be key to the success of LTE. The next step in the evolution of LTE is LTE-A, a 4G system which promises peak data rates in the Gbps range and improved cell-edge performance. Important scheduling/interference mitigation-related technical issues which require further exploration include [63]:

(1) use of relaying techniques which can provide a relatively inexpensive way of increasing spectral efficiency, system capacity, and area coverage. Preliminary studies can be found in [23, 33, 64].

(2) DL and UL coordinated multipoint transmission/reception to improve high data rate coverage and cell-edge throughput. For DL, this refers to coordination in scheduling transmissions from multiple geographically separated transmission points. For UL, this involves different types of coordination in reception at multiple geographically separated points.

(3) support for UL spatial multiplexing of up to four layers and DL spatial multiplexing of up to eight layers to increase bit rates.

Another general area deserving attention is the design of low-complexity scheduling/interference mitigation schemes which provide near optimal performance.

List of Acronyms

3GPP: Third Generation Partnership Project
AC: Admission control
AMC: Adaptive modulation and coding
BLER: Block error rate
CQI: Channel quality indicator
CC: Control channel
eNB: Enhanced NodeB (3GPP term for base station)
E-UTRA: Evolved universal terrestrial radio access
FD: Frequency domain
FDD: Frequency division duplex
FDPS: Frequency domain packet scheduling
FFR: Fractional frequency reuse
FRF: Frequency reuse factor
HII: High-interference indicator
HSDPA: High-speed downlink packet access
HSUPA: High-speed uplink packet access
HSPA: Refers to HSDPA and/or HSUPA
ICI: Intercell interference
ICIC: Intercell interference coordination
PRB: Physical resource block, same as RB
LTE: Long-term evolution
LTE-A: Long-term evolution-Advanced
MBMS: Multimedia broadcast/multicast service
MCS: Modulation and coding scheme
OFDM: Orthogonal frequency division multiplexing
OFDMA: Orthogonal frequency division multiple access
PAPR: Peak-to-average power ratio
PF: Proportional fair
PLR: Packet loss rate
PO: Pure opportunistic
PS: Packet scheduler
PSS: Priority set scheduling
QoS: Quality of service
RB: Resource block, same as PRB
RR: Round robin
SB: Scheduling block
SC-FDMA: Single-carrier frequency division multiple access
SCS: Scheduling candidate set
SFR: Soft frequency reuse
SINR: Signal to interference and noise ratio
SRS: Sounding reference signal
TD: Time domain
TDD: Time division duplex
TDPS: Time domain packet scheduling
TTI: Transmission time interval
UE: User equipment (3GPP term for mobile terminal)
UMTS: Universal Mobile Telecommunications System
UTRA: UMTS terrestrial radio access
UTRAN: UMTS terrestrial radio access network
VoIP: Voice over Internet Protocol
W-CDMA: Wideband code division multiple access.

Acknowledgments

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada

under Grant OGP0001731 and by the UBC PMC-Sierra Professorship in Networking and Communications.

References

[1] "Requirements related to technical performance for IMT-Advanced radio interface(s)," Tech. Rep. ITU-R M.2134, International Telecommunication Union (ITU), Geneva, Switzerland, 2008.

[2] "Requirements for evolved UTRA (E-UTRA) and evolved UTRAN (E-UTRAN) (Release 9)," Tech. Rep. TR 25.913 V9.0.0, 3rd Generation Partnership Project (3GPP), 2009.

[3] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, Academic Press, Amsterdam, The Netherlands, 2008.

[4] "Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (Release 8)," Tech. Rep. 3G TS36.211, 3rd Generation Partnership Project, New Orleans, La, USA, September 2007.

[5] D. Astély, E. Dahlman, A. Furuskär, Y. Jading, M. Lindström, and S. Parkvall, "LTE: the evolution of mobile broadband," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 44–51, 2009.

[6] S. Nagata, Y. Ofuji, K. Higuchi, and M. Sawahashi, "Optimum resource block bandwidth for frequency domain channel-dependent scheduling in evolved UTRA downlink OFDM radio access," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '06)*, vol. 1, pp. 206–210, May 2006.

[7] A. Pokhariyal, T. E. Kolding, and P. E. Mogensen, "Performance of downlink frequency domain packet scheduling for the utran long term evolution," in *Proceedings of the IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, September 2006.

[8] A. Pokhariyal, G. Monghal, K. I. Pedersen et al., "Frequency domain packet scheduling under fractional load for the UTRAN LTE downlink," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '07)*, pp. 699–703, Dublin, Ireland, April 2007.

[9] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, "Dynamic packet scheduling performance in UTRA long term evolution downlink," in *Proceedings of the 3rd International Symposium on Wireless Pervasive Computing (ISWPC '08)*, pp. 308–313, May 2008.

[10] R. Kwan, C. Leung, and J. Zhang, "Multiuser scheduling on the downlink of an LTE cellular system," *Research Letters in Communications*, vol. 2008, no. 3, Article ID 323048, 4 pages, 2008.

[11] Y. Lin and G. Yue, "Channel-adapted and buffer-aware packet scheduling in LTE wireless communication system," in *Proceedings of the IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, October 2008.

[12] D. L. Villa, C. U. Castellanos, I. Z. Kovács, F. Frederiksen, and K. I. Pedersen, "Performance of downlink UTRAN LTE under control channel constraints," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '08)*, pp. 2512–2516, May 2008.

[13] R. Kwan, C. Leung, and J. Zhang, "Proportional fair multiuser scheduling in LTE," *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 461–464, 2009.

[14] A. Larmo, M. Lindström, M. Meyer, G. Pelletier, J. Torsner, and H. Wiemann, "The LTE link-layer design," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 52–59, 2009.

[15] A. Haghighat, Z. Lin, and G. Zhang, "Haar compression for efficient CQI feedback signaling in 3GPP LTE systems," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '08)*, pp. 819–823, March-April 2008.

[16] N. Kolehmainen, J. Puttonen, P. Kela, T. Ristaniemi, T. Henttonen, and M. Moisio, "Channel quality indication reporting schemes for UTRAN long term evolution downlink," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '08)*, pp. 2522–2526, May 2008.

[17] S. Lu, Y. Cai, L. Zhang et al., "Channel-aware frequency domain packet scheduling for MBMS in LTE," in *Proceedings of the IEEE 69th Vehicular Technology Conference (VTC '09)*, April 2009.

[18] G. Monghal, K. I. Pedersen, I. Z. Kovács, and P. E. Mogensen, "QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution," in *Proceedings of the IEEEVehicular Technology Conference (VTC '08)*, pp. 2532–2536, 2008.

[19] M. Wernersson, S. Wänstedt, and P. Synnergren, "Effects of QOS scheduling strategies on performance of mixed services over LTE," in *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, September 2007.

[20] B. Sadiq, R. Madan, and A. Sampath, "Downlink scheduling for multiclass traffic in LTE," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 510617, 2009.

[21] N. Wei, A. Pokhariyal, T. B. Sørensen, T. E. Kolding, and P. E. Mogensen, "Performance of MIMO with frequency domain packet scheduling in UTRAN LTE downlink," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '07)*, pp. 1177–1181, April 2007.

[22] S.-B. Lee, S. Choudhury, A. Khoshnevis, S. Xu, and S. Lu, "Downlink MIMO with frequency-domain packet scheduling for 3GPP LTE," in *Proceedings of the 28th IEEE Conference on Computer Communications*, pp. 2611–2615, 2009.

[23] R. Schoenen, R. Halfmann, and B. H. Walke, "MAC performance of a 3GPP-LTE multihop cellular network," in *Proceedings of the IEEE International Conference on Communications*, pp. 4819–4824, Beijing, China, May 2008.

[24] M. Al-Rawi, R. Jantti, J. Torsner, and M. Sagfors, "Opportunistic uplink scheduling for 3G LTE systems," in *Proceedings of the IEEE International Conference on Innovations in Information Technology*, November 2007.

[25] M. Al-Rawi, R. Jäntti, J. Torsner, and M. Sagfors, "On the performance of heuristic opportunistic scheduling in the uplink of 3g lte networks," in *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, September 2008.

[26] F. D. Calabrese, C. Rosa, M. Anas, P. H. Michaelsen, K. I. Pedersen, and P. E. Mogensen, "Adaptive transmission bandwidth based packet scheduling for LTE uplink," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '08)*, September 2008.

[27] F. D. Calabrese, P. H. Michaelsen, C. Rosa et al., "Search-tree based uplink channel aware packet scheduling for UTRAN LTE," in *Proceedings of the IEEE 67th Vehicular Technology Conference (VTC '08)*, pp. 1949–1953, May 2008.

[28] L. A. M. R. de Temiño, G. Berardinelli, S. Frattasi, and P. Mogensen, "Channel-aware scheduling algorithms for SC-FDMA in LTE uplink," in *Proceedings of the IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, September 2008.

[29] F. D. Calabrese, C. Rosa, K. I. Pedersen, and P. E. Mogensen, "Performance of proportional fair frequency and time domain scheduling in LTE uplink," in *Proceedings of the European Wireless Conference (EW '09)*, pp. 271–275, May 2009.

[30] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, "Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '09)*, pp. 2611–2615, April 2009.

[31] Z. Li, C. Yin, and G. Yue, "Delay-bounded power-efficient packet scheduling for uplink systems of LTE," in *Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '09)*, September 2009.

[32] R. Sandanalaksmi, K. Manivanan, S. Manikandan, R. Barathi, and D. Devanathan, "Fair channel aware packet scheduling algorithm for fast UL HARQ in UTRAN LTE," in *Proceedings of the International Conference on Control Automation, Communication and Energy Conservation, (INCACEC '09)*, pp. 1–5, June 2009.

[33] E. Yaacoub and Z. Dawy, "Centralized and distributed LTE uplink scheduling in a distributed base station scenario," in *Proceedings of the International Conference on Advances in Computational Tools for Engineering Applications, (ACTEA '09)*, pp. 11–15, July 2009.

[34] E. Yaacoub, H. Al-Asadi, and Z. Dawy, "Low complexity scheduling algorithms for the LTE uplink," in *Proceedings of the IEEE Symposium on Computers and Communications (ISCC '09)*, pp. 266–270, July 2009.

[35] E. Yaacoub and Z. Dawy, "A game theoretical formulation for proportional fairness in LTE uplink scheduling," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '09)*, April 2009.

[36] M. Anas, C. Rosa, F. D. Calabrese, K. I. Pedersen, and P. E. Mogensen, "Combined admission control and scheduling for QoS differentiation in LTE uplink," in *Proceedings of the 68th Semi-Annual IEEE Vehicular Technology (VTC '08)*, September 2008.

[37] S. Choi, K. Jun, Y. Shin, S. Kang, and B. Choi, "MAC scheduling scheme for VoIP traffic service in 3G LTE," in *Proceedings of the IEEE 66th Vehicular Technology Conference (VTC '07)*, pp. 1441–1445, October 2007.

[38] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and performance of semi-persistent scheduling for VoIP in LTE system," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '07)*, pp. 2861–2864, September 2007.

[39] H. Wang and D. Jiang, "Performance comparison of control-less scheduling policies for VoIP in LTE UL," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '08)*, pp. 2497–2501, April 2008.

[40] Y.-S. Kim, "An efficient scheduling scheme to enhance the capacity of VoIP services in evolved UTRA uplink," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, Article ID 732418, 2008.

[41] S. Saha and R. Quazi, "Priority-coupling—a semi-persistent MAC scheduling scheme for VoIP traffic on 3G LTE," in *Proceedings of the 10th International Conference on Telecommunications*, pp. 325–329, June 2009.

[42] F. Khan, *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance*, Cambridge University Press, Cambridge, UK, 2009.

[43] Ericsson, "Muting—further discussion and results," Tech. Rep. TSG-RAN WG1 R1-050763, 3rd Generation Partnership Project (3GPP), August-September 2005.

[44] Ericsson, "Inter-cell interference handling for E-UTRA," Tech. Rep. TSG-RAN WG1 R1-050764, 3rd Generation Partnership Project (3GPP), August-September 2005.

[45] Ericsson, "Downlink inter-cell interference coordination/avoidance evaluation of frequency reuse," Tech. Rep. TSG-RAN WG1 R1-061374, 3rd Generation Partnership Project (3GPP), May 2006.

[46] Alcatel, "OFDM air interface with QOS at cell edge," Tech. Rep. TSG-RAN WG1 R1-050272, 3rd Generation Partnership Project (3GPP), April 2005.

[47] "Multi-cell simulation results for interference co-ordination in new OFDM DL," Tech. Rep. TSG-RAN WG1 R1-050694, 3rd Generation Partnership Project (3GPP), August-September 2005.

[48] A. Simonsson, "Frequency reuse and intercell interference co-ordination in E-UTRA," in *Proceedings of the IEEE 65th Vehicular Technology Conference (VTC '07)*, pp. 3091–3095, April 2007.

[49] Siemens, "Interference mitigation—considerations and results on frequency reuse," Tech. Rep. TSGRAN WG1 R1-050738, 3rd Generation Partnership Project (3GPP), August-September 2005.

[50] X. Fan, S. Chen, and X. Zhang, "An inter-cell interference coordination technique based on users' ratio and multi-level frequency allocations," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '07)*, pp. 799–802, September 2007.

[51] TS36.423, "X2 application protocol (X2AP)," Tech. Rep. V8.3.0, 3GPP, 2008.

[52] X. Mao, A. Maaref, and K. H. Teo, "Adaptive soft frequency reuse for inter-cell interference coordination in SC-FDMA based 3GPP LTE uplinks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '08)*, pp. 4782–4787, December 2008.

[53] B. Krasniqi, M. Wrulich, and C. F. Mecklenbräuker, "Network-load dependent partial frequency reuse for LTE," in *Proceedings of the 9th International Symposium on Communications and Information Technology (ISCIT '09)*, pp. 672–676, September 2009.

[54] Y. Xiang, J. Luo, and C. Hartmann, "Inter-cell interference mitigation through flexible resource reuse in OFDMA based communication networks," in *Proceedings of the European Wireless*, April 2007.

[55] J. Ellenbeck, H. Al-Shatri, and C. Hartmann, "Performance of decentralized interference coordination in the LTE uplink," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '09)*, September 2009.

[56] M. C. Necker, "Coordinated fractional frequency reuse," in *Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM '07)*, pp. 296–305, October 2007.

[57] Y.-J. Chiang, Z. Tao, J. Zhang, and C.-C. J. Kuo, "A graph-based approach to multi-cell OFDMA downlink resource allocation," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '08)*, pp. 3712–3717, December 2008.

[58] R. Y. Chang, Z. Tao, J. Zhang, and C.-C. J. Kuo, "A graph approach to dynamic Fractional Frequency Reuse (FFR) in multi-cell OFDMA networks," in *Proceedings of the IEEE*

International Conference on Communications (ICC '09), June 2009.

[59] Nokia and Nokia Siemens Networks, "Downlink interference coordination," Tech. Rep. TSG-RAN WG1 R1-072974, 3rd Generation Partnership Project (3GPP), June 2007.

[60] A. Rácz, N. Reider, and G. Fodor, "On the impact of inter-cell interference in LTE," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '08)*, pp. 5436–5441, November-December 2008.

[61] S. Kumar, G. Monghal, J. Nin, I. Ordas, K. I. Pedersen, and P. E. Mogensen, "Autonomous inter cell interference avoidance under fractional load for downlink long term evolution," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '09)*, April 2009.

[62] Qualcomm Europe, "Subrame specific channel quality reporting," Tech. Rep. TSG-RAN WG1 R1-094998, 3rd Generation Partnership Project (3GPP), November 2009.

[63] "Feasibility study for further advancements for E-UTRA (LTEAdvanced) (Release 9)," Tech. Rep. TR 36.912 V9.2.0, 3rd Generation Partnership Project (3GPP), 2010.

[64] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMAX and LTE-advanced mobile systems," *IEEE Communications Magazine*, vol. 47, no. 10, pp. 100–105, 2009.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2010, Article ID 819729, 16 pages
doi:10.1155/2010/819729

Research Article

Scheduling for Improving System Capacity in Multiservice 3GPP LTE

Francisco Rafael Marques Lima,[1] Stefan Wänstedt,[2] Francisco Rodrigo Porto Cavalcanti,[1] and Walter Cruz Freitas Junior[1]

[1] *GTEL-Wireless Telecom Research Group, Department of Teleinformatics Engineering, Federal University of Ceará, Campus do Pici, 60455-760 Fortaleza, Brazil*
[2] *Ericsson AB, Laboratoriegränd 11, 97128 Luleå, Sweden*

Correspondence should be addressed to Francisco Rafael Marques Lima, rafaelm@gtel.ufc.br

Received 1 February 2010; Revised 3 May 2010; Accepted 26 June 2010

Academic Editor: Raymond Kwan

We study the impact of scheduling algorithms on the provision of multiple services in the long term evolution (LTE) system. In order to measure how well the services are provided by the system, we use the definition of joint system capacity. In this context, we claim that scheduling strategies should consider the current satisfaction level of each service and the offered load to the system by each service. We propose a downlink-scheduling strategy according to these ideas named capacity-driven resource allocation (CRA). The CRA scheduler dynamically controls the resource sharing among flows of different services such as delay-sensitive and rate demanding ones. Moreover, CRA scheduler exploits the channel-quality knowledge to utilize the system resources efficiently. Simulation results in a multicell scenario show that the CRA scheduler is robust regarding channel quality knowledge and that it provides significant gains in joint system capacity in single and mixed service scenarios.

1. Introduction

The cellular networks have allowed us to communicate with people who are at the most remote places in the world through mobile phone calls. Furthermore, we are used to searching for information and entertainment by utilizing fixed broadband access in our homes. With the introduction of third generation (3G) networks, besides phone calls the mobile phones are now capable of accessing data services such as Web browsing and e-mail. However, the increased demand for new multimedia services, lower costs and improved quality of service (QoS) provision continuously stimulate the evolution of mobile communications. Consequently, 3rd generation partnership project (3GPP) and other standardization bodies have been working with the specification of the next steps in mobile communications: the long term evolution (LTE) and LTE-Advanced.

LTE will bring advantages for subscribers with new applications such as interactive TV and user-generated videos, and for operators with backward compatibility with legacy networks and simpler architecture. Among the main features of LTE we can highlight the utilization of orthogonal frequency division multiple access (OFDMA) as the radio access technology in the downlink and a pure packet-based all-internet protocol (IP) architecture.

OFDMA is a multiple access scheme based on orthogonal frequency division multiplexing (OFDM) digital modulation scheme where multiple user equipments (UEs) get assigned subcarriers or subsets of them in order to be served simultaneously. One of the advantages of an OFDMA-based system is the opportunity to benefit from frequency and multiuser diversities. Due to the frequency diversity, it is unlikely that all frequency resources in a link have the same channel quality. The multiuser diversity comes from the fact that UEs located at different positions within a cell experience almost independent channel qualities [1].

All-IP is a broad concept which means that the core network will be completely packet-switched and based on IP [2]. The main advantages of an All-IP architecture are reduced costs and efficient support to mass-market usage of

any IP-based service. On the other hand, a packet-switched network imposes some challenges on the provision of QoS guarantees for delay-sensitive services such as voice that was traditionally provided over circuit-switched networks and now must share the system resources with other services. In this multiservice scenario, system operators expect to achieve high system capacity by fulfilling the heterogeneous QoS requirements of the multiple flows in the system. Although in LTE a UE can bear multiple service flows, without loss of generality, here, only one service flow is considered per UE. Consequently, flow and UE are interchangeable throughout the text.

In order to exploit the advantages of OFDMA multiple access scheme and guarantee the QoS of different services with distinct traffic patterns and requirements, scheduling is of utmost importance in LTE. Scheduling algorithms are responsible for selecting which UEs will have access to the system resources and with which configuration. Therefore, in this paper, we deal with downlink scheduling algorithms for capacity maximization in multiservice scenarios. Our main contributions in this paper are:

(i) review a reasonable definition of system capacity suitable for multiservice scenarios and discuss how different scheduling strategies impact on it;

(ii) propose a scheduling strategy with the objective of improving the joint system capacity of the LTE system. Specifically, our proposal takes into account many requirements and limitations imposed by the LTE architecture; and

(iii) present a performance evaluation by using a detailed computational simulator in order to analyze the possible benefits of our proposed scheduler when applied to the LTE system. In the performance evaluation, we study some relevant aspects in multiservice scenarios.

The remainder of this paper is organized as follows. In Section 2, we present a brief overview about scheduling algorithms in the literature and contextualize our proposal. Section 3 is devoted to the problem formulation and presentation of the assumed system modeling. Specifically, in the problem formulation we define the joint system capacity and discuss how scheduling algorithms impact on it. After that, we present the proposed scheduling algorithm in Section 4. Then, in Section 5, we present a performance evaluation of our proposal by using a computational simulator that models the main aspects and restrictions of the LTE system. Finally, we provide the main conclusions and perspectives of this work.

2. Related Works

Studies in scheduling algorithms for wireless networks have acquired emphasis with the introduction of packet-switched single-carrier networks, such as high speed downlink packet access (HSDPA) and code-division multiple access 2000 evolution-data only (CDMA2000 EV-DO), where the system resources are no longer dedicated to the flows but shared among them. One of the first approaches followed by the community research was to generalize the concepts of queuing theory employed in wireline schedulers for the wireless setting [3–6]. As an example, in [7] the authors provide important bounds on queue backlog for different scheduling algorithms in OFDMA networks. Although the theoretical results in these works are conditioned to strong requirements regarding information availability and involved stochastic processes, the achieved results have paved the path towards the design of wireless schedulers. Important insights such as the relevance of the use of channel quality information, fairness in wireless environment and QoS-related aspects (e.g., delay) were obtained.

Scheduling algorithms in general are designed to deal with (RT) and/or non-real time (NRT) services. RT services are characterized by the short time response between the communicating parts. These services have strict requirements regarding packet delay and jitter. As an example of this kind of service we can mention voice over IP (VoIP). On the other hand, NRT services do not have tight requirements concerning packet delay although high packet delays are unacceptable. In fact, when transmitting NRT services the major constraint is the information integrity, that is, information loss is not tolerable. Therefore, applications of this type must have error-correction or recovery mechanisms. Web browsing is an example of an NRT service.

Contributions on scheduling algorithms can be categorized according to the multiple access method (single- and multicarrier networks) and provided services by the network (single- and multiservice scenarios). The interested reader can refer to [8–11] for single-carrier schedulers developed for NRT and RT services, respectively. In [12–15] the reader can find multicarrier schedulers for single-service scenarios. Our focus on this paper is on multicarrier schedulers for multiservice scenarios.

One of the first works that studied the QoS provision in multiservice scenarios was [16] that focused on single-carrier system. The proposed scheduler is based on proportional fair (PF) [8] with an additional weight that depends on packet delay for RT services and on a token bucket control algorithm for NRT services. Another work that also studies PF-like schedulers for multiservice scenarios is [17] where the flows are differentiated by service-dependent weights that are either fixed or dependent on packet delay.

Many works have been published focusing on the LTE system and multiservice scenarios and a complete survey is out of the scope of this section. The objective is to show some guidelines and how our work innovates compared to the state of art. In [18], the authors propose a downlink scheduling algorithm based on PF that takes into account in its formulation the per-flow amount of data awaiting transmission. In the scheduling process, flows belonging to a high priority service (delay-sensitive service) have an explicit priority over the others. In [19, 20], the authors consider a mixed service scenario; however, the main concern is with delay-sensitive services such as VoIP and video. In [21], the authors propose a scheduling algorithm that has an inter and intraservice part. In the former, the scheduler defines which service will have the flows scheduled. In the latter, the scheduler selects

the flows that will be scheduled from the service defined in the interservice part. The results shown in that article were spectral efficiency, fairness and average throughput. Although the scheduler presented in [21] can be applied in multiservice scenario, the authors neither provide results with mixed service scenarios nor consider delay-sensitive services.

The paper [22] highlights the importance of using channel- and buffer size/delay-aware schedulers in achieving high performance in multicarrier systems. In this paper, we do follow this approach by using information about channel-quality and packet delay in our proposed solution. Also in [22], the authors conclude that strict prioritization of a specific service (as it is done in some of the commented previous articles in this section) is not suitable for multiservice scenarios.

In summary, the main objectives of the multiservice schedulers designed for LTE described in this section are either to provide better spectral efficiency while keeping fairness among flows or protecting the QoS of high priority services such as delay-sensitive ones. In this paper, we propose a scheduling algorithm to improve the joint system capacity of the LTE system. This approach has not been followed by the previous articles on LTE to the best of our knowledge. In the next section, we formally define the joint system capacity.

3. Problem Formulation and System Modeling

In this section, we formally define the joint system capacity and relate it with other existing system-capacity definitions. Moreover, we present the main aspects about LTE that are relevant to this work.

3.1. Capacity Definitions. There are many ways to measure the capacity of wireless systems. A well-known definition of capacity is the one provided by Shannon which consists in the maximum achievable set of rates in multiple access channels with an arbitrarily small probability of error [23]. As this metric represents a bound in performance, in practice, the sum of the transmitted data rates (downlink) or aggregated data rate is used. Usually, this metric is also normalized by the system bandwidth and expressed in b/s/Hz.

However, with the increased availability of new services in wireless networks, the user perceived quality or QoS should also be included in the capacity measures. In this sense, the system capacity could be defined as the maximum aggregated data rate subject to the constraint that the average experienced quality of all flows in the system should be fulfilled according to a given target. As average experienced quality, we can mention the average delay of all transmitted packets or the average packet throughput, for example.

As in the wireless systems the perceived QoS can significantly vary among different flows, we believe that fairness related aspects should be taken into account when defining the capacity measure. This can be accomplished by the joint system capacity that is shown in Section 3.2.

3.2. Joint System Capacity. The joint system capacity used in this paper was first defined in [24] and used as performance metric in many papers including [17, 25–27].

Consider a multiservice system where the offered load is measured as the number of connected flows in the system. Let us identify the service types by indices that compose the service set Ψ. In a wireless system, flows start and finish their data sessions in a dynamic process. Considering that the system is in stable state where the statistics can be considered stationary, the mean number of connected flows in the system or offered load from service s is ρ_s. The total offered load to the system by all services is given by

$$\rho^{\text{total}} = \sum_{s \in \Psi} \rho_s. \tag{1}$$

The fraction of the total load offered by service s is given by

$$f_s = \frac{\rho_s}{\rho^{\text{total}}}. \tag{2}$$

We define the service mix, $\mathbf{f}$, as a vector composed of the elements f_s. In order to measure the quality provided to the flows of a specific service $s \in \Psi$ when a scheduling strategy named SCHED is used, we consider the user satisfaction ratio $q_s^{\text{SCHED}}(\rho^{\text{total}}, \mathbf{f})$. The user satisfaction ratio for a given service is defined as the fraction of flows from this service whose data sessions ended with the QoS requirements fulfilled. The user satisfaction ratio is a nonincreasing function of the total offered load. Furthermore, the user satisfaction ratio depends on the service mix since the load imposed by the flows of each service to the system is different due to distinct QoS demands.

We consider that the individual capacity for a service is the maximum total offered load in which the majority of the ended data sessions of this service achieve the QoS requirements. More specifically, the individual capacity of service $s \in \Psi$ (measured in number of connected flows) with the scheduling strategy SCHED is defined as

$$i_s^{\text{SCHED}}(\mathbf{f}) = \max\left(\rho^{\text{total}} \mid q_s^{\text{SCHED}}\left(\rho^{\text{total}}, \mathbf{f}\right) \geq Q_s^{\text{thres}}\right), \tag{3}$$

where Q_s^{thres} consists in the user satisfaction ratio threshold for service s, that is, the minimum acceptable user satisfaction ratio for service s defined by the system operator.

In a mixed service scenario, we have to take into account the QoS provided to the flows of all services. The joint system capacity is the maximum total offered load in which all provided services fulfill the user satisfaction ratio threshold. Therefore, the joint system capacity when a scheduling algorithm SCHED is used, $c^{\text{SCHED}}(\mathbf{f})$, is defined as

$$c^{\text{SCHED}}(\mathbf{f}) = \min\left(i_s^{\text{SCHED}}(\mathbf{f}), \forall s \in \Psi\right). \tag{4}$$

The joint system capacity is able to capture relevant aspects of multiservice wireless systems such as user and service specific quality requirements. The scheduling algorithm proposed in this paper in Section 4 aims at improving the joint system capacity of LTE.

3.3. Impact of Scheduling Strategies on the Joint System Capacity. Once we have defined the joint system capacity, we will discuss the effects of some scheduling strategies on the system capacity. Consider, for example, a scheduling algorithm that gives explicit priority to a given service such as the approaches of some of the papers described in Section 2. In this case, the flows of the service with higher priority tend to be scheduled more often than the flows of other services. As a consequence, the individual capacity of the prioritized service will be higher than the other services. However, as defined in (4), the joint system capacity is limited by the service with lower individual capacity.

In Figure 1, we illustrate this issue in an example with a two-service case. In this figure, "Prio" is a scheduling strategy that tends to allocate most of the system resources to the flows of the second service that has high priority. On the other hand, "Balanced" is an example of another scheduling algorithm that is capable of balancing the QoS experienced by the flows of the different services in order to improve the joint system capacity. As can be seen in this figure, the joint system capacity with the first scheduler is i_1^{Prio} and for the second scheduler is i_1^{Balanced} with $i_1^{\text{Balanced}} > i_1^{\text{Prio}}$. In order to improve the joint system capacity, the scheduler "Balanced" had to degrade the individual capacity of service 2 in order to improve the individual capacity of service 1 (direction of the arrows in the figure). Although the degradation of the QoS provided to the flows of a specific service can seem odd at first, this is supported by the system operator's point-of-view. When offering wireless services the system operator is interested in fulfilling each per-service minimum user satisfaction ratio threshold (Q_s^{thres}) and, therefore, quality overprovision for a specific service will not bring additional benefit for the system operator.

In general, a condition to the joint system capacity maximization is that all provided services achieve the same individual system capacity [24]. As a conclusion, scheduling strategies with explicit priority for flows of a specific service are not able to maximize the joint system capacity.

In a stationary environment where aspects such as traffic mix proportions and channel conditions are kept statistically unchanged, a suitable set of weights of the PF-based schedulers such as in [17] could be found in order to provide QoS balancing among services. However, in real networks, these aspects are rather unpredictable and time-variant. Consequently, a fixed set of weights would not lead to an improved joint system capacity in this scenario.

In order to improve the joint system capacity, the scheduling strategy should perform a controlled resource sharing among flows of different services. Aspects such as the load of each service in the system (traffic mix proportion, $\mathbf{f}$) and the satisfaction level of each flow should be addressed. These are the main ideas of our proposed scheduling algorithm described in Section 4.

3.4. System Modeling. In this section, we point out the relevant aspects of the LTE system that impact on our work. For a complete description of LTE the interested reader can refer to the 3GPP's standards and the articles [28–30], for example. In Sections 3.4.1 and 3.4.2 we present the physical

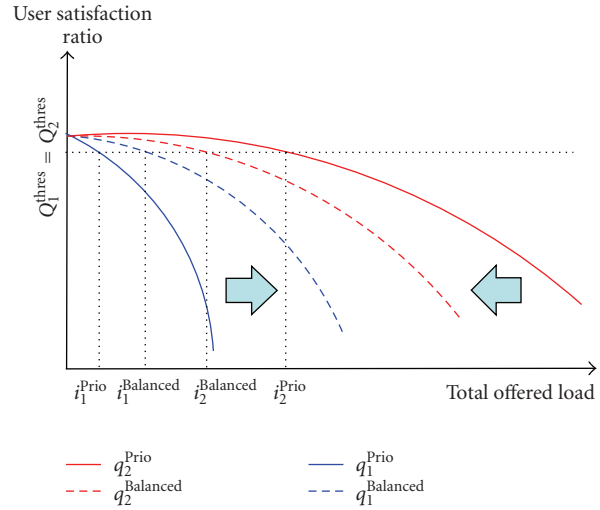


Figure 1: Illustration of the effects of scheduling strategies in the joint system capacity.

[31] and medium access control (MAC) [32] layers and in Section 3.4.3 we show the services types and user satisfaction model.

3.4.1. Physical Layer. The time domain structure of LTE is composed of radio frames of 10 ms. Each radio frame has 10 equally-sized subframes of length 1 ms. Subframes, in turn, consist of two slots of length 0.5 ms. The scheduling takes place in a subframe basis.

The default subcarrier spacing is 15 kHz and all subcarriers are grouped in sets of 12 subcarriers. A resource block in LTE is defined as a two-dimensional grid with 12 subcarriers in frequency and 0.5 ms in time that corresponds to 6 or 7 OFDM symbols depending on cyclic prefix length. The Resource Unit (RU) in the system is composed of two resource blocks concatenated in the time domain, that is, 12 subcarriers and 1 ms.

The resources are utilized by physical channels and signals. Physical channels are utilized for transmission of data and/or control information from the MAC layer. The physical signals are used to support physical-layer functionality and do not carry any information from the MAC layer [31].

Among the physical channels, we emphasize the function of physical downlink shared channel (PDSCH) and physical downlink control channel (PDCCH). The former is utilized for transmission of data traffic while the latter is used for downlink layer 1/layer 2 control signaling. Specifically, PDCCH is used to carry uplink scheduling grants and downlink scheduling assignments, such as PDSCH resource indication, transport format, hybrid automatic repeat request (HARQ) information and transport block size. Depending on the time-variant PDCCH capacity, different number of UEs can be scheduled in a given transmission time interval (TTI). In this work, we consider that there is a fixed limit in the number of scheduled UEs in a TTI. Although this is an important aspect of LTE system, this issue has not been

considered in the majority of the works about scheduling in the literature.

In this study, we consider that the allocated power per RU is fixed and is equal to the ratio between the available power and the number of RUs.

3.4.2. Medium Access Control. When a connection (or bearer) is established between the UE and the LTE core network a QoS class identifier (QCI) is specified. This defines whether the bearer is guaranteed bit-rate or not, target delay and loss requirements, for example. The enhanced node B (eNB) translates the QCI attributes into requirements for the air interface. The scheduling should allocate resources according to these requirements.

The HARQ comprises a number of processes where each process uses a simple stop-and-wait protocol. HARQ for downlink, that is the focus in this study, is asynchronous and adaptive. By asynchronous we mean that the scheduler has the freedom to choose the subframe for retransmission dynamically. In adaptive HARQ, the scheduler can use a different resource configuration for retransmission compared to the previous (re)transmission. In case the data is a retransmission of a previously stored data, the received data is soft combined with the data stored in the soft buffer.

3.4.3. Service Types and Model for User Satisfaction. The concept of user satisfaction is very important when interpreting the system performance. There are many parameters to consider such as the service type, technical parameters (e.g., delay and throughput), and even economical issues (such as the price to use the wireless service) [33, 34]. However, in this study we consider only technical aspects concerning the perceived quality by the end user.

In this work, we consider two classes of services that have been used as reference in the research community: RT and NRT services that were described in Section 2. Note that we can directly map the QCIs attributes to the service classes used in this study [35].

In the area of RT services, quite extensive models have been obtained that relate the frame erasure rate (FER) with perceived quality [36]. Therefore, we consider that a RT flow, j, is satisfied when

$$\gamma_j[k] = \frac{n_j^{\text{lost}}[k]}{n_j^{\text{lost}}[k] + n_j^{\text{succ}}[k]} \leq \gamma_j^{\text{req}}, \tag{5}$$

where $\gamma_j[k]$ is the accumulated FER for flow j at TTI k, and γ_j^{req} is the FER requirement of flow j. The variables $n_j^{\text{succ}}[k]$ and $n_j^{\text{lost}}[k]$ are the number of successfully transmitted and lost packets from flow j at TTI k since the session initialization, respectively. If the FER of a flow is higher than the requirement this flow is considered unsatisfied.

The satisfaction model for NRT flows is based on the average data rate and is suitable for services with bursty traffic [37]. A flow j that belongs to an NRT service is satisfied when

$$\bar{r}_j[k] = \frac{l_j[k]}{\left(t_j[k] \cdot a\right)} \geq \bar{r}_j^{\text{req}}, \tag{6}$$

where $\bar{r}_j[k]$ is the average data rate of flow j at TTI k computed from the session initialization, and $\bar{r}_j^{\text{req}}$ is the average data rate requirement of flow j. The variable $l_j[k]$ is the number of correctly transmitted bits from flow j at TTI k since the session initialization, $t_j[k]$ is the total active time of flow j at TTI k since the session initialization and a is the duration of one TTI. By active time we mean the total time that a flow has data to transmit. If the average data rate of a flow is lower than the requirement this flow is considered unsatisfied.

4. Capacity-Driven Resource Allocation (CRA)

Based on the previous discussion about joint system capacity and the main aspects/restrictions of the LTE architecture, we propose the scheduling algorithm CRA. The main objective of CRA is to improve the joint system capacity of the LTE system in multiservice scenarios.

4.1. CRA Overview. We have followed a common approach when designing schedulers for multicarrier systems that is to split the scheduling functionality into two parts: Resource Allocation and Resource Assignment. The Resource Allocation part is responsible for defining which flows will be scheduled and determining their required data rate at the current TTI, while the Resource Assignment part defines which resources will be assigned to the selected flows in the Resource Assignment part.

In Figure 2, we illustrate the main building blocks of the CRA scheduler. In the Resource Allocation part, the CRA scheduler firstly builds priority lists for each existing service. These services can be either NRT or RT. In the priority list, the flows that belong to a specific service are ordered according to a priority based on the satisfaction level. The prioritization intends to give transmission opportunities to the flows that can be easily satisfied. Besides, for each flow CRA also calculates the required data rate that this flow needs to transmit in the current TTI.

The last step of Resource Allocation part consists in the selection, according to the load imposed by each service, of the flows that will receive RUs in the Resource Assignment part. As the flows are arranged in priority lists, the task is to define how many flows of each service will be chosen to get resources in the Resource Assignment part. In this last part, the selected flows get assigned RUs in a channel opportunistic fashion.

4.2. CRA: Resource Allocation Part. As described previously, for each flow we determine the required data rate to transmit at the current TTI. For an NRT flow, this data rate is calculated as follows:

$$\Delta r_j[k] = \bar{r}_j^{\text{req}} \cdot \left(t_j[k] + 1\right) - \bar{r}_j[k-1] \cdot t_j[k-1]. \tag{7}$$

This required data rate represents the data rate that should be allocated to an unsatisfied flow in order to this flow stay satisfied even if it does not have transmission opportunities in the next TTI (see Appendix A for demonstration). Note that if a flow is already satisfied the required rate would

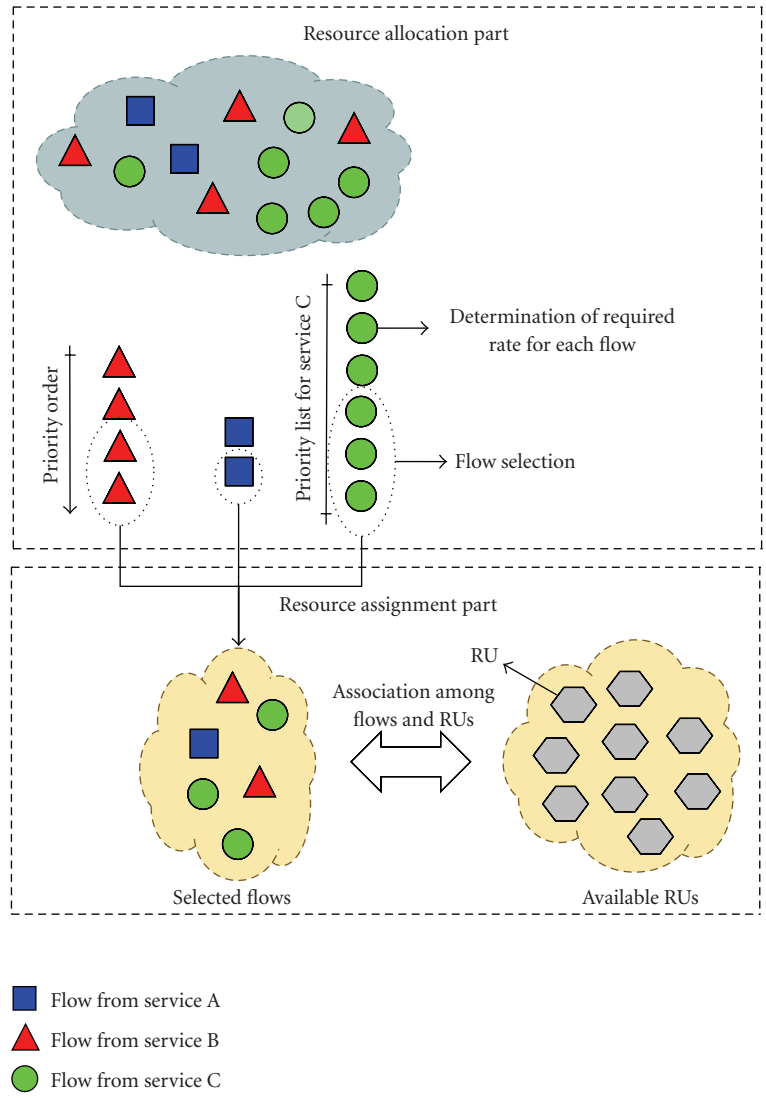


Figure 2: Building blocks of CRA. Illustration of the Resource Allocation and Resource Assignment parts.

be negative. Therefore the modulus (or absolute value) of this required rate can be seen as the "distance" that the average data rate of a flow ($\bar{r}_j[k]$) is from the average data rate requirement ($\bar{r}_j^{\text{req}}$).

For RT flows we define the required data rate as follows:

$$\Delta r_j[k] = \frac{b_j^{\text{oldest}}[k]}{a}, \tag{8}$$

where $b_j^{\text{oldest}}[k]$ represents the number of bits of the oldest packet in the transmit buffer of the eNB corresponding to the flow j at TTI k, that is, the packet that is waiting for transmission for the longer period of time. The choice of this required rate for RT flows is based on the fact that, in general, the upper protocol layers from this service split the data to transmit in small packets with short delay requirements. In this way, when these flows get transmit opportunities the complete packet should be transmitted in order to avoid packet discard.

Once the data rate demanded by each flow is determined, the next step is to build priority lists for each service. The priority lists are built according to the service classes. The main idea is to prioritize the flows that can be easily satisfied.

In the priority list for any service, the flows with retransmissions have the highest priority. Concerning NRT services, the flows that are currently unsatisfied have precedence over the ones that are satisfied. The prioritization is the opposite for RT flows. The reason for this strategy is the fact that users of NRT services tolerate temporary QoS fluctuations during the data session if in average the QoS is fulfilled. On the other hand, due to the quick response characteristic of RT services, a temporary oscillation in the experienced QoS compromises the whole session.

Besides the prioritization based on the satisfaction status, we assign a priority to each flow, p_j, in order to sort the flows within the set of satisfied and unsatisfied flows. An illustration of the process to build the priority list is shown in Figure 3. The priority for NRT flows is given by

$$p_j = \frac{\overline{\alpha_j}[k]}{\left\|\Delta r_j[k]\right\|}, \tag{9}$$

where $\overline{\alpha_j}[k]$ is the ratio between the transmit data rate of a flow j at TTI k in case it gets assigned all available RUs and the number of available RUs, and the operator $\|\cdot\|$ returns the absolute value. Therefore, within the group of unsatisfied flows, the flow that is in good channel condition and requires lower data rate to become satisfied than the other unsatisfied flows is more prioritized. This is a reasonable strategy in order to increase the number of satisfied flows in the system. In the group of satisfied NRT flows, the ones that are in good channel conditions and that are near to the unsatisfaction have precedence over the other satisfied flows. In Appendix B, we show that this prioritization is an optimum strategy to solve the problem of maximizing the number of satisfied flows when the flow's requirements are represented in terms of number of RUs.

Before defining the priority of an RT flow we define the concept of "distance" to the requirement for RT flows as in the following

$$w_j[k] = \begin{cases} \left\lfloor \dfrac{\left(n_j^{\text{succ}}[k] + n_j^{\text{lost}}[k]\right)\cdot \gamma_j^{\text{req}} - n_j^{\text{lost}}[k]}{1-\gamma_j^{\text{req}}} \right\rfloor, & \\ \qquad\qquad \text{if } \gamma_j[k] \le \gamma_j^{\text{req}} & \\ \left\lceil \dfrac{n_j^{\text{lost}}[k] - \left(n_j^{\text{succ}}[k] + n_j^{\text{lost}}[k]\right)\cdot \gamma_j^{\text{req}}}{\gamma_j^{\text{req}}} \right\rceil, & \\ \qquad\qquad \text{otherwise,} & \end{cases} \tag{10}$$

where the operators $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ return the first integer greater than or equal to and the first integer lower than or equal to a real number, respectively.

The variable $w_j[k]$ represents how many consecutive packets an unsatisfied flow j at TTI k has to successfully transmit to become satisfied (with $\gamma_j[k] \le \gamma_j^{\text{req}}$). For a satisfied flow j, $w_j[k]$ means the maximum number of packets that this flow can lose successively and still be satisfied. In other words, $w_j[k]$ defines how close (or far) the FER of a given flow is from the required FER. See Appendix B for the demonstration of (10).

In this way, the priority for an RT flow is given by

$$p_j = \frac{1}{\left(d_j^{\text{req}} - d_j^{\text{oldest}}\right)\cdot\left(w_j + 1\right)}, \tag{11}$$

where d_j^{req} and d_j^{oldest} are the packet delay requirement and current packet delay of the oldest packet of flow j, respectively. In Figure 4, we plot p_j in function of d_j^{oldest} and $w_j[k]$ for RT flows considering d_j^{req} equal to 80 ms. From this figure we can see that the flows that have packets with delays close to their deadlines and shorter "distance" between the current and required FER than the other flows are more prioritized. The rule is the same for satisfied and unsatisfied flows. By prioritizing flows with packet delays close to the deadline we take advantage of the fact that RT applications tolerate a certain packet delay without compromising the end user perceived quality. Consequently, more RT connected flows can be multiplexed in order to increase capacity.

The last part of Resource Allocation is the definition of which flows of each priority list will be chosen to get resources in the Resource Assignment part. Consider that set Ω_s is composed of the active flows (flows that have data to transmit) from service $s \in \Psi$. The number of flows that will be selected is constrained by the conditions below

$$y_s \approx \mu \cdot \frac{|\Omega_s|}{\sum_{p\in\Psi}\left|\Omega_p\right|}, \qquad \sum_{s\in\Psi} y_s \le \mu, \tag{12}$$

where y_s represents the number of selected flows from service s to be scheduled, $|\cdot|$ represents the cardinality of a set and μ is the maximum possible number of scheduled terminals in a TTI.

The first part in (12) states that the number of selected flows must be almost proportional to the number of active flows from each existing service in the cell. The second part has the objective of guaranteeing that the number of scheduled flows is equal to or smaller than the maximum number of terminals that can be scheduled in a TTI (as presented in Section 3.4.1). The main objective of these constraints is to provide a better resource distribution among the different services even when the offered load per service is unbalanced.

4.3. CRA: Resource Assignment Part. The main idea in the Resource Assignment is to distribute the RUs in a fair and opportunistic way among the selected flows in the Resource Allocation part. The Resource Assignment part is executed in phases. In each phase, all the flows get assigned one RU. However, the flow that will choose its RU first is the one that has the RU in better channel conditions among all other flows. The process continues until all flows get assigned one RU in the current phase. The flows compete for resources until receiving the number of RUs to transmit with the required data rate, $\Delta r_j[k]$, defined in the Resource Allocation part. In this case, they are taken out of the process. If all the selected flows have RUs enough to fulfill the required data rate, the remaining RUs are equally divided among all selected flows.

5. Performance Results

In this section, we present a case study where we apply the proposed scheduling strategy in the LTE system in a multiservice scenario composed of VoIP (RT service) and Web (NRT service). We firstly present the simulation setup

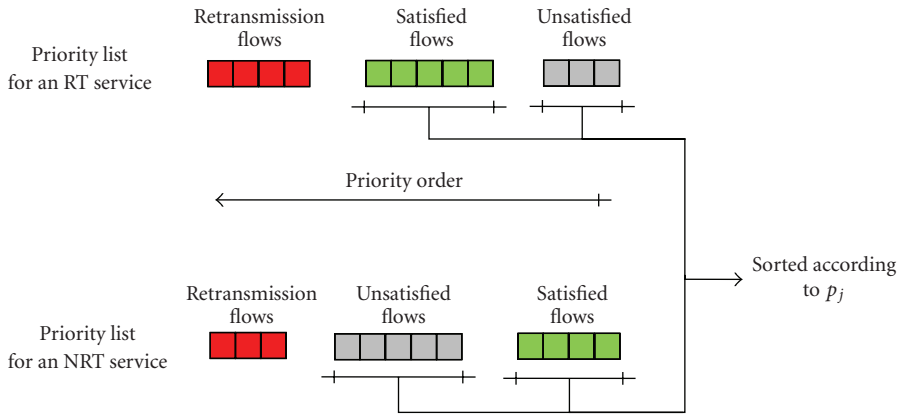


Figure 3: Priority lists for RT and NRT services. The prioritization takes into account the retransmission status, satisfaction level and a per-flow priority.

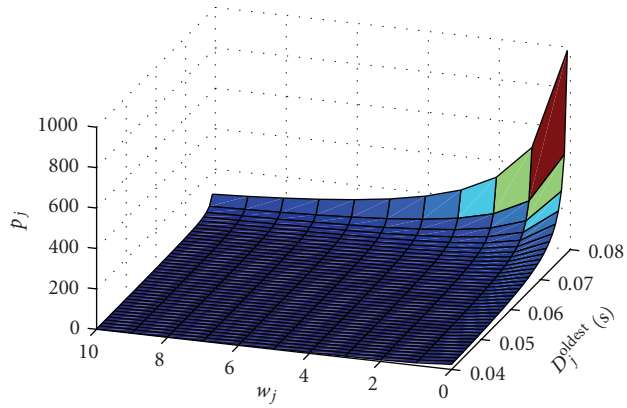


Figure 4: Illustration of p_j for RT flows considering D_j^{req} equal to 80 ms.

where we provide the simulation parameters and define reference schedulers used for comparison. The results are divided into two parts: sensitivity analysis and performance evaluation in single and mixed service scenarios.

5.1. Simulation Setup. The results presented in this section are drawn from a dynamic system-level simulator that models the LTE system according to 3GPP specifications detailed in Section 3.4. The simulator model includes multiple cells, intercell interference and propagation phenomena such as path loss, shadowing and fast fading. Moreover, the main aspects related to both radio interface layers and upper layers (Transport control protocol (TCP), user datagram protocol (UDP), IP and applications) were taken into account in the simulation models.

The Web traffic model is characterized as request-response traffic: a client, that utilizes a mobile station and is located in the radio network, requests one or more web pages within a session, that is, hypertext transfer protocol (HTTP) requests. The server generates and returns the web pages. Once a web page is received at the client, this user reads the web page for some seconds and then requests another web page. We consider that the Web pages have a fixed length and that the reading time follows an exponential distribution. The Web service can be mapped to QCI1, for example.

The VoIP packets are generated by a speech coder that mimics the adaptive multirate (AMR) codec. This coder produces voice frames every 20 ms during speech periods and small packets, named silence insertion description (SID) packets, to simulate background noise during silence periods. In this study we consider that there is a conversation between two clients, one out of the LTE network utilizing a computer (client A) and another client utilizing a UE (client B) in the radio access network. As the downlink is focused, the performance is measured in the client located in the radio network utilizing the UE. The model for conversation has three states: Client A talking, client B talking and mutual silence. The model switches between this three states with a time period drawn from exponential distributions. The VoIP service can be mapped to QCI8, for example. The main simulation parameters are shown in Table 1.

The reference schedulers used in the simulation are: delay scheduler (DS) and maximum rate (MR). All these reference schedulers give transmission opportunities to the flow with higher priority. The selected flow gets assigned its RUs in better channel condition until the data rate necessary to transmit all backlogged data is achieved. If the selected flow does not utilize all available RUs, the next more prioritized flow is selected to get resources and so on. Note that there is the limitation in the maximum number of scheduled flows as described in Section 3.4.1.

The difference between the reference schedulers is the prioritization. DS assigns the best RUs of the flow (VoIP or Web in this case) whose headline radio link control (RLC) SDU has the current greatest delay. Therefore, DS scheduler is a channel- and QoS-aware scheduler. The MR scheduler chooses the flow that can transmit more information bits when using the available bandwidth (better channel condition). In this way, as reference schedulers we have a strategy that takes into account channel and QoS aspects (DS) and another that is only channel-aware (MR).

Table 1: Main simulation parameters considered in this work. The parameters are classified in general parameters of LTE, propagation, deployment, and service-specific ones.

Parameter	Value	Unit
General		
Bandwidth	3	MHz
Carrier frequency	2	GHz
Number of RUs	15	—
Total cell power	20	W
Transport network packet delay (including Internet and Core Network (CN))	14	ms
PDCCH capacity (number of scheduling grants per TTI)	5 (static modeling)	—
Number of HARQ processes	8	—
Maximum number of HARQ retransmissions	10	—
VoIP/Web user satisfaction ratio thresholds	95/90	%
Propagation		
Path gain at 1 meter distance	−29.03	dB
Path gain per dB distance	−3.52	dB
Shadowing standard deviation	8	dB
Antenna type	SCM 3GPP [38]	—
Deployment		
Number of eNB/cells per eNB	3/3	—
Number of antennas in the UEs/cell	2/1	—
Cell radius	500	m
Frequency reuse	full	—
UE speed	3	km/h
Voip Flows		
RLC service data unit SDU discard period	80	ms
Mean talk period time	5	s
Voice activity factor	0.5	—
Frame size	264	bits
Frame period	20	ms
Maximum end-to-end VoIP frame delay	140	ms
SID frame size	39	bits
Required FER	1	%
Web Flows		
Web page size (fixed)	10,000	bytes
Mean reading time	1.5	s
Average data rate requirement	128	kbps

In the next sections, we will evaluate how these two strategies impacts on the joint system capacity in a multiservice scenario.

5.2. Sensitivity Analysis. CRA is a channel-aware scheduler, that is, it relies on channel quality measurements. Consequently, before analyzing simulation results in mixed service scenarios we devote this section to the sensitivity analysis of our proposal regarding channel state reporting. In Figure 5, we show the user satisfaction ratio versus the offered load (measured in number of flows in a cell) in the Web-only scenario when the channel state reporting period is increased. The CRA scheduler utilizes channel quality measurements for Web flows in two parts: in the priority calculation of Web flows and in the resource assignment.

In the priority calculation, an average channel quality measurement is considered. Therefore, it is expected that the dependence of this part on channel measurements is not critical. In the Resource Assignment, a per-RU channel quality measurement is utilized in order to assign the best resources to the UEs. Consequently, this part must be more affected by higher channel reporting periods. However, as it can be seen in Figure 5, the degradation in capacity of CRA considering a user satisfaction ratio threshold of 90% is of only 2 UEs, which represents a capacity loss of approximately 2% when the channel reporting period is changed from 10 ms to 25 ms. DS and MR also suffer a capacity loss of approximately 2 UEs considering the same user satisfaction ratio threshold. This similar performance among the schedulers points to a degradation in the link

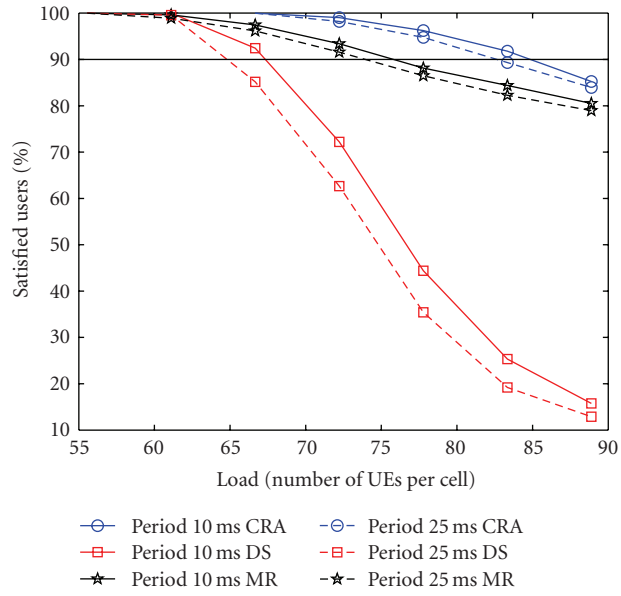


Figure 5: User satisfaction ratio in the Web-only scenario with variable channel state reporting periods for CRA, DS and MR schedulers.

adaptation as the main reason for the capacity loss. Link adaptation, which is common for any scheduler, also utilizes channel quality measurements.

In Figure 6, we show the user satisfaction ratio with different channel state reporting periods in a VoIP-only scenario. The channel quality measurements are only utilized in the Resource Assignment part of CRA when VoIP flows are concerned, that is a common part for any service type. The degradation observed in the VoIP service is similar to the one observed in the Web-only scenario in Figure 5. Changing the channel reporting period from 10 ms to 25 ms caused similar capacity decreases of approximately 2% for CRA and 3% for DS considering the user satisfaction ratio threshold of 95%. The capacity loss for MR scheduler cannot be measured in the user satisfaction ratio threshold of 95% because of its poor performance in the simulated load range. However, the degradation in user satisfaction ratio is similar to the visualized in the other schedulers. The performance loss was mainly caused by a degradation of link adaptation, as in the previous scenario.

5.3. Performance Evaluation. In Figure 7, we show the user satisfaction ratio for the simulated schedulers in three mixed service scenarios: 25% of VoIP and 75% of Web flows (v25w75), 50% of VoIP and 50% of Web flows (v50w50) and 75% of VoIP and 25% of Web flows (v75w25). Note that these proportions are related to the number of connected flows in the system and not active flows. In Figure 7(a), we show the user satisfaction ratio for VoIP service and in Figure 7(b) the user satisfaction ratio for Web service. In the axis of abscissas, we present the system load measured in number of flows in a cell (the sum of VoIP and Web flows).

In general, we can observe that the user satisfaction ratio for the simulated schedulers and services is improved when the percentage of VoIP flows in the system is increased. The reason for this behavior is the fact that a VoIP flow demands lower data rates and consequently resources of LTE system. In order to measure the individual service capacity, as defined in (3), the user satisfaction ratio thresholds of 95% for VoIP and 90% for Web should be considered as depicted in the figures.

In all simulated mixed service scenarios with the MR scheduler, the user satisfaction ratio for Web service is better than the VoIP one. This good performance for Web service is due to the burst nature of Web traffic and a more flexible QoS requirement based on average data rate. Because of the burst traffic pattern of Web, during the inactive periods of the flows in better channel conditions the MR scheduler can select the other flows. This works as a statistical time multiplexing mechanism that is not present in low-rate and regular VoIP traffic. Furthermore, when MR schedules VoIP flows the scheduling process is limited by the maximum number of scheduled UEs instead of the number of available RUs. This leads to a low resource usage.

When DS scheduler is concerned, we can observe in Figure 7 that the Web service experiences a lower individual capacity than VoIP in the simulated mixed service scenarios. Consequently, the former limits the joint system capacity of DS scheduler. The packet delay is one important measurement when scheduling RT services because it directly affects the FER that determines the user satisfaction for RT flows. Moreover, another reason to prioritize flows with high headline packet delays is that these flows usually have more than one buffered packet to transmit. The transport block size in LTE utilizing one RU can, depending on the modulation order and code rate, be greater than one RLC SDU that is mapped one to one with VoIP frames. As a result, scheduling flows with high headline packet delays increases the efficiency by reducing the protocol layer overheads and padding rate per sent VoIP packet [39]. This explains the good performance of DS for VoIP service. However, scheduling Web traffic based on packet delay usually grants flows with transmission opportunities that have too much buffered data due to poor channel conditions.

Despite the differences among the reference schedulers they all have in common one aspect: they do not consider satisfaction status and load per service in their formulation. Although the joint system capacity in the simulated mixed service scenarios with the CRA scheduler is limited by the Web service, the user satisfaction ratios provided to the VoIP and Web services are improved. In Figure 7, we can observe that the CRA scheduler achieves higher individual capacities for VoIP and Web services in the user satisfaction ratio thresholds of 95% and 90%, respectively. The scheduling in CRA is accomplished in a more intelligent way by considering the number of active flows from each service and the current QoS conditions of each flow. The Resource Assignment part in the CRA scheduler was designed to give equal opportunities to each selected flow get its resources in better channel quality. In the following, we show that the CRA scheduler provides improved joint system capacity also in the single-service scenarios.

In Figure 8, we show the system capacity region that is built from the user satisfaction ratio curves for several

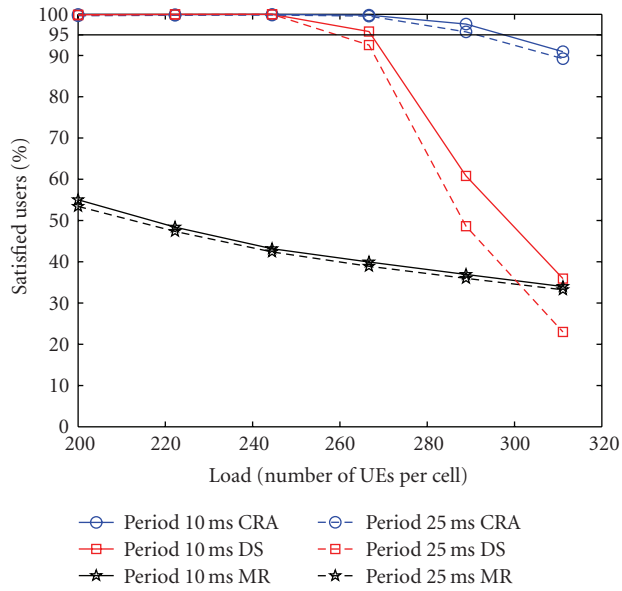
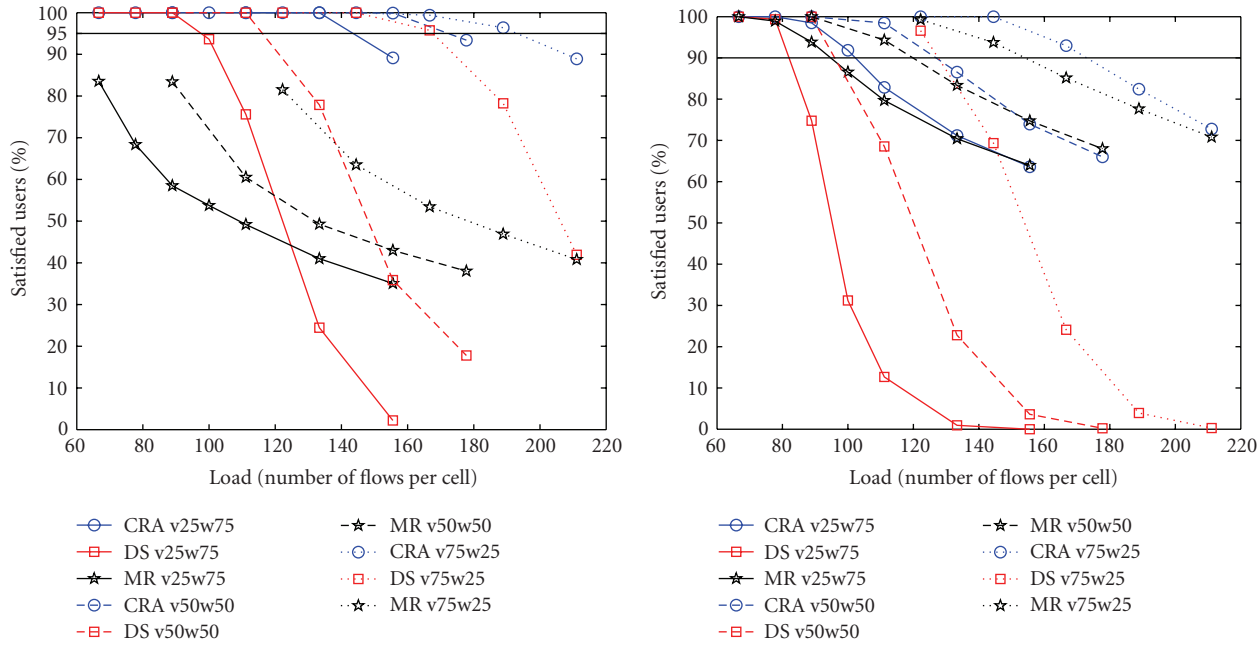


Figure 6: User satisfaction ratio in the VoIP-only scenario with variable channel state reporting periods for CRA, DS and MR schedulers.



(a) User satisfaction ratio for VoIP service

(b) User satisfaction ratio for Web service

Figure 7: User satisfaction ratio in the mixed service scenarios 25% VoIP and 75% Web, 50% VoIP and 50% Web and 75% VoIP and 25% Web for CRA, DS and MR schedulers.

traffic mixes. This is an important result since it allows us to assess the performance of schedulers when the network is submitted to different traffic mixes. The MR scheduler is not included in this figure because its user satisfaction ratio for VoIP service was lower than the VoIP satisfaction threshold for the simulated offered load.

By increasing the user satisfaction ratio for each service, the CRA scheduler also provides greater capacity region. The gain of the CRA scheduler over DS in joint capacity can be quantified by the larger area (below the curves) in the system capacity region. The gain of the CRA scheduler over DS is of approximately 37%. This gain may not be completed realized in real deployments since there are many other aspects in real networks that are not feasible to model even in a detailed simulator such as the one used in this study. However, we expect that even in real deployments our proposal is able to overperform the reference schedulers concerning the joint system capacity. Therefore, we believe that the main ideas of

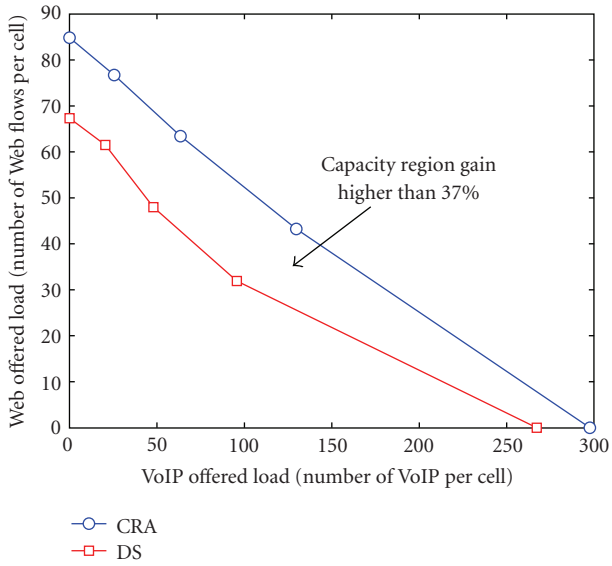


Figure 8: Capacity region.

our proposed scheduler should be considered in the design of scheduling algorithms for the LTE system in order to improve the joint system capacity in multiservice scenarios.

6. Conclusions

The provision of sustainable quality of service (QoS) to different flows with heterogeneous requisites is an important issue in multiservice wireless networks such as long term evolution (LTE). In this scenario, the system capacity should not only take into account the flows of a specific service. In fact, the system capacity should measure how well all the services are provided by the system. Therefore, in order to improve the system capacity, scheduling algorithms are of utmost importance due to their intrinsic task of distributing resources to the flows of different services in a short transmission time interval (TTI).

Some approaches in scheduling design found in the literature are not capable of improving the system capacity in a general setting. The main reason for that is the disregard of two aspects: the offered load to the system by each service and satisfaction level of the connected flows. In this context we proposed a downlink scheduling algorithm named capacity-driven resource allocation (CRA) whose main objective is to improve the system capacity in multiservice wireless networks. In addition, CRA was designed according to the main restrictions and characteristics of the LTE architecture.

By the performance results of a sensitivity analysis regarding the periodicity of channel state report, we concluded that the performance of the CRA scheduler is as dependent of this metric as reference schedulers are. Morevoer, in mixed service scenarios the simulation results have shown that the CRA scheduler is able to provide an overall gain in joint capacity higher than 37% over reference schedulers.

Although the presented performance evaluation considers a scenario with two services, the ideas in our proposed scheduler are general enough to deal with many concurrent services in the same packet-switched network. In this multiservice scenario, CRA is capable of providing a better QoS balancing independently of time varying aspects such as channel conditions and service mix proportions.

Appendices

A. Demonstration of (7)

Without loss of generality consider an NRT flow j that is currently unsatisfied at TTI k, that is, $\bar{r}_j[k] < \bar{r}_j^{\text{req}}$. Therefore, we would like to know which data rate should be allocated at the current TTI k in order to the flow j become satisfied for the next λ TTIs even if no resource is assigned to it, that is, $\bar{r}_j[k+\lambda] = \bar{r}_j^{\text{req}}$. Expanding $\bar{r}_j[k+\lambda]$ we have

$$\begin{aligned}\bar{r}_j[k+\lambda] &= \frac{l_j[k+\lambda]}{a \cdot t_j[k+\lambda]} = \frac{l_j[k-1]+\mu}{a \cdot \left(t_j[k]+\lambda\right)} \\ &= \frac{a \cdot \left(\bar{r}_j[k-1]\right) \cdot \left(t_j[k-1]\right)+\mu}{a \cdot \left(t_j[k]+\lambda\right)},\end{aligned} \tag{A.1}$$

where μ is the amount of data (bits) that should be transmitted at TTI k.

Therefore, we have to find μ by solving the following equation

$$\frac{a \cdot \left(\bar{r}_j[k-1]\right) \cdot \left(t_j[k-1]\right)+\mu}{a \cdot \left(t_j[k]+\lambda\right)} = \bar{r}_j^{\text{req}}. \tag{A.2}$$

The solution of this equation is

$$\mu = a \cdot \left(t_j[k]+\lambda\right) \cdot \bar{r}_j^{\text{req}} - a \cdot \bar{r}_j[k-1] \cdot t_j[k-1]. \tag{A.3}$$

Finally, the current data rate that should be allocated to the flow j at TTI k in order to this flow become satisfied for the next λ TTIs even if no resource is assigned to it, $\Delta r_j[k]$, is given by

$$\Delta r_j[k] = \frac{\mu}{a} = \left(t_j[k]+\lambda\right) \cdot \bar{r}_j^{\text{req}} - \bar{r}_j[k-1] \cdot t_j[k-1]. \tag{A.4}$$

Note that in (7), we considered $\lambda = 1$. The choice of the parameter λ depends on the satisfaction level that we intend to provide to the scheduled flows. When the required rate ($\Delta r_j[k]$) is calculated using high values of λ the scheduled flows will stay satisfied for several TTIs. On the other hand, the required rate increases with the parameter λ. Therefore, the scheduled flows will get more resources in order to fulfill their required rate. In this way, few flows could be scheduled simultaneously. Therefore, we have chosen $\lambda = 1$ in order to allow for better resource distribution among flows.

B. Maximization of Satisfied Flows

In this appendix, we show the intuition behind the flow prioritization in (9). The objective of this flow prioritization is to decrease the number of unsatisfied NRT flows (or increase the number of satisfied flows) at the current TTI. The task to be performed in the Resource Allocation part is to define the flows that should get system resources at the Resource Assignment part of the CRA algorithm. In the Resource Assignment part, the scheduled flows are associated with the available resources.

The flows' requirements are represented by the required data rate $\Delta r_j[k]$. As different flows experience different channel states in each individual RU, the task of defining the scheduled flows in order to achieve the stated objective is a hard problem to solve. Indeed, our problem would become easier to solve if the flows' requirements were directly represented in number of required RUs instead of data rate.

The representation of flow's requirement in number of RUs can be performed by the following relation

$$\overline{m}_j = \left\lceil \frac{\Delta r_j[k]}{\overline{\alpha}_j} \right\rceil, \tag{B.1}$$

where $\overline{m}_j$ is the required number of resources demanded by flow j at the current TTI considering the average transmit data rate among all RUs and $\lceil \cdot \rceil$ returns the first integer greater than or equal to a real number. Obviously, some information about individual channel quality of each RU is lost when we consider this representation. However, as point out in [40] that used a similar approach to solve a subproblem, this procedure is justified by the low performance degradation and reduced computational complexity when an opportunistic resource assignment is performed.

According to this, we can formulate our problem as

$$\begin{aligned} \max_{\mathbf{m}} \quad & \sum_{j=1}^{J} u\left(m_j - \overline{m}_j\right) \\ \text{subject to} \quad & \sum_{j=1}^{J} m_j \le N, \\ & m_j \in \mathbb{N}, \end{aligned} \tag{B.2}$$

where m_j is the number of resources allocated to flow j, $\mathbf{m}$ is a $J \times 1$ column vector composed of m_j, N is the number of available RUs, J is the number of active flows at the current TTI and, finally, $u(\cdot)$ is the step function that assumes the value 0 when its argument is negative and 1 otherwise. In summary, this is an optimization problem to define the number of RUs that should be assigned to the flows so as to maximize the number of satisfied flows (with $m_j \ge \overline{m}_j$) constrained to the limited number of RU. Problem (B.2) is a combinatorial optimization problem with potentially multiple optimum solutions.

Algorithm 1 is able to find one of the optimum solutions. Basically, the resources are allocated to the flows with lower required number of resources in order to become satisfied. In order to prove the optimality of this algorithm consider a solution given by this algorithm, $\mathbf{m}^* = [m_1^* \; m_2^* \; \cdots \; m_J^*]$, that leads to L satisfied flows with $L < J$. Suppose also that there is another solution, $\mathbf{m}' = [m_1' m_2' \; \cdots \; m_J']$, with H satisfied flows where $L < H < J$. For the sake of this proof, consider that $\mathbf{o} = [o_1 \; o_2 \; \cdots \; o_J]$ is the vector with the index of the flows sorted in the ascending order of $\overline{m}_j$ according to the line 2 of Algorithm 1, that is, $\overline{m}_{o_1} \le \overline{m}_{o_2} \le \cdots \le \overline{m}_{o_J}$. Consider also that $\mathbf{x} = [x_1 \; x_2 \; \cdots \; x_H]$ is a vector of length H with the indices (disposed in any order) of the satisfied flows given by solution $\mathbf{m}'$.

In order to make the solutions $\mathbf{m}^*$ and $\mathbf{m}'$ lead to L and H satisfied flows, respectively, the following constraints should be fulfilled

$$\begin{aligned} m_{o_1}^* &\ge \overline{m}_{o_1} & m_{x_1}' &\ge \overline{m}_{x_1} \\ &\vdots & &\vdots \\ m_{o_L}^* &\ge \overline{m}_{o_L} & m_{x_H}^* &\ge \overline{m}_{x_H} \\ \sum_{j=o_1}^{o_L} m_j^* &\le N & \sum_{j=x_1}^{x_H} m_j' &\le N, \end{aligned} \tag{B.3}$$

By adding these constraints we have that

$$N \ge \sum_{j=o_1}^{o_L} m_j^* \ge \sum_{j=o_1}^{o_L} \overline{m}_j, \tag{B.4}$$

$$N \ge \sum_{j=x_1}^{x_H} m_j' \ge \sum_{j=x_1}^{x_H} \overline{m}_j. \tag{B.5}$$

Particularly, the following equation derived from (B.5) should also hold

$$N \ge \sum_{j=x_1}^{x_{L+1}} m_j' \ge \sum_{j=x_1}^{x_{L+1}} \overline{m}_j. \tag{B.6}$$

As in Algorithm 1, the flows with lower required number of RUs are selected firstly we have that

$$\begin{aligned} \overline{m}_{x_1} &\ge \overline{m}_{o_1}, \\ \overline{m}_{x_1} + \overline{m}_{x_2} &\ge \overline{m}_{o_1} + \overline{m}_{o_2}, \\ &\cdots \qquad\qquad \cdots, \\ \overline{m}_{x_1} + \overline{m}_{x_2} + \cdots + \overline{m}_{x_L} &\ge \overline{m}_{o_1} + \overline{m}_{o_2} + \cdots + \overline{m}_{o_L}, \\ \overline{m}_{x_1} + \overline{m}_{x_2} + \cdots + \overline{m}_{x_L} + \overline{m}_{x_{L+1}} &\ge \overline{m}_{o_1} + \overline{m}_{o_2} \\ &\quad + \cdots + \overline{m}_{o_L} + \overline{m}_{o_{L+1}}. \end{aligned} \tag{B.7}$$

However, as the solution found by the Algorithm 1 was able to satisfy only L flows we have that

$$\overline{m}_{o_1} + \overline{m}_{o_2} + \cdots + \overline{m}_{o_L} + \overline{m}_{o_{L+1}} > N, \tag{B.8}$$

and consequently by the last constraint in (B.7) we have that

$$\overline{m}_{x_1} + \overline{m}_{x_2} + \cdots + \overline{m}_{x_L} + \overline{m}_{x_{L+1}} > N. \tag{B.9}$$

(1) $m_j \leftarrow 0\ \forall j$
(2) $\mathbf{o} \leftarrow \text{sort}_{\forall j}(\overline{m}_j)$ {Sort in the ascending order}
(3) $\theta \leftarrow N$
(4) $i \leftarrow 1$
(5) **while** $(\theta > 0)$ AND $(i \leq J)$ **do**
(6) $\quad j^* \leftarrow o_i$
(7) $\quad$ **If** $(\theta - \overline{m}_{j^*}) > 0$ **then**
(8) $\quad\quad m_{j^*} \leftarrow \overline{m}_{j^*}$
(9) $\quad\quad \theta \leftarrow \theta - \overline{m}_{j^*}$
(10) $\quad$ **else**
(11) $\quad\quad m_{j^*} \leftarrow \theta$
(12) $\quad\quad \theta \leftarrow 0$
(13) $\quad$ **end if**
(14) $\quad i \leftarrow i + 1$
(15) **end while**
(16) **If** $\theta > 0$ **then**
(17) $\quad i \leftarrow 1$
(18) $\quad$ **while** $\theta > 0$ **do**
(19) $\quad\quad j^* \leftarrow o_i$
(20) $\quad\quad m_{j^*} \leftarrow m_{j^*} + 1$
(21) $\quad\quad \theta \leftarrow \theta - 1$
(22) $\quad\quad i \leftarrow i + 1$
(23) $\quad\quad$ **if** $i > J$ **then**
(24) $\quad\quad\quad i \leftarrow 1$
(25) $\quad\quad$ **end if**
(26) $\quad$ **end while**
(27) **end if**
(28) Output of the algorithm: **m**

Algorithm 1: Algorithm to solve the problem (B.2).

As a consequence, (B.9) contradicts (B.5), (B.6) and (B.3). Therefore, the solution given by Algorithm 1 provides an optimum solution of problem (B.2).

As selecting the flows with higher priority p_j (given by (9)) is similar to the steps of Algorithm 1 where the flows with lower $\overline{m}_j$ are chosen first, we can conclude that the employed prioritization is a reasonable strategy to increase the number of satisfied flows.

B. Demonstration of (10)

Consider an RT flow j that is currently unsatisfied at TTI k, that is, $\gamma_j[k] > \gamma_j^{\text{req}}$. Therefore, we would like to know how many packets, ν, this flow has to successfully transmit in a row so as to become satisfied, that is, $\gamma_j[k'] = \gamma_j^{\text{req}}$ where $k' > k$. In this way, the FER at TTI k' is given by

$$\gamma_j[k'] = \frac{n_j^{\text{lost}}[k']}{n_j^{\text{lost}}[k'] + n_j^{\text{succ}}[k']} = \frac{n_j^{\text{lost}}[k]}{n_j^{\text{lost}}[k] + \left(n_j^{\text{succ}}[k] + \nu\right)}. \tag{C.1}$$

Therefore, we have to solve the following equation

$$\frac{n_j^{\text{lost}}[k]}{n_j^{\text{lost}}[k] + \left(n_j^{\text{succ}}[k] + \nu\right)} = \gamma_j^{\text{req}}. \tag{C.2}$$

The solution of this equation is

$$\nu = \frac{n_j^{\text{lost}}[k] - \left(n_j^{\text{succ}}[k] + n_j^{\text{lost}}[k]\right) \cdot \gamma_j^{\text{req}}}{\gamma_j^{\text{req}}}. \tag{C.3}$$

If the RT flow j is currently satisfied, that is, $\gamma_j[k] \leq \gamma_j^{\text{req}}$, we need to know the maximum number of packets, ϵ, that this flow can lose successively and still be satisfied, that is, $\gamma_j[k'] = \gamma_j^{\text{req}}$ where $k' > k$.

The FER at TTI k' is given by

$$\gamma_j[k'] = \frac{n_j^{\text{lost}}[k']}{n_j^{\text{lost}}[k'] + n_j^{\text{succ}}[k']} = \frac{\left(n_j^{\text{lost}}[k] + \epsilon\right)}{\left(n_j^{\text{lost}}[k] + \epsilon\right) + n_j^{\text{succ}}[k]}. \tag{C.4}$$

The equation to be solved is as follows

$$\frac{\left(n_j^{\text{lost}}[k] + \epsilon\right)}{\left(n_j^{\text{lost}}[k] + \epsilon\right) + n_j^{\text{succ}}[k]} = \gamma_j^{\text{req}}. \tag{C.5}$$

This equation is solved by setting ϵ as follows

$$\epsilon = \frac{\left(n_j^{\text{succ}}[k] + n_j^{\text{lost}}[k]\right) \cdot \gamma_j^{\text{req}} - n_j^{\text{lost}}[k]}{1 - \gamma_j^{\text{req}}}. \tag{C.6}$$

Acknowledgment

This work was supported by the Research and Development Center, Ericsson Telecomunicações S.A., Brazil, under EDB/UFC.22 Technical Cooperation Contract.

References

[1] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proceedings of the IEEE International Conference on Communications (ICC '95)*, vol. 1, pp. 331–335, Seattle, Wash, USA, June 1995.

[2] 3GPP, "All-IP Network (AIPN) feasibility study," Tech. Rep. TR22.978 V8.0.0—Release 8, 3rd Generation Partnership Project, December 2008.

[3] V. Bharghavan, S. Lu, and T. Nandagopal, "Fair queuing in wireless networks: issues and approaches," *IEEE Personal Communications*, vol. 6, no. 1, pp. 44–53, 1999.

[4] Y. Cao and V. O. K. Li, "Efficient algorithms for broadband space-time coded wireless communication," *Proceedings of the IEEE*, vol. 89, no. 1, pp. 76–87.

[5] H. Fattah and C. Leung, "An overview of scheduling algorithms in wireless multimedia networks," *IEEE Wireless Communications*, vol. 9, no. 5, pp. 76–83, 2002.

[6] S. Shakkottai and T. S. Rappaport, "Research challenges in wireless networks: a technical overview," in *Proceedings of the 5th International Symposium on Wireless Personal Multimedia Communications*, vol. 1, pp. 12–18, October 2002.

[7] G. Wunder and C. Zhou, "Queueing analysis for the OFDMA downlink: throughput regions, delay and exponential backlog bounds," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 871–881, 2009.

[8] F. P. Kelly, A. K. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[9] S. Wänstedt, F. Rui, M. Ericsson, and M. Nordberg, "Providing reliableand efficient VoIP over cellular networks," in *Proceedings of the Future Telecommunications Conference*, October 2005.

[10] B. Wang, K. I. Pedersen, T. E. Kolding, and P. E. Mogensen, "Performance of VoIP on HSDPA," in *Proceedings of the 61st IEEE Vehicular Technology Conference (VTC '05)*, vol. 4, pp. 2335–2339, Stockholm, Sweden, June 2005.

[11] P. Hosein, "Scheduling of VoIP traffic over a time-shared wireless packet data channel," in *Proceedings of the 7th IEEE International Conference on Personal Wireless Communications (ICPWC '05)*, pp. 38–41, January 2005.

[12] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, "Dynamic packet scheduling performance in UTRA long term evolution downlink," in *Proceedings of the 3rd IEEE International Symposium of Wireless Pervasive Computing (ISWPC '08)*, pp. 308–313, May 2008.

[13] A. Pokhariyal, K. I. Pedersen, G. Monghal, et al., "HARQ aware frequency domain packet scheduler with different degrees of fairness for the UTRAN long term evolution," in *Proceedings of the 65th IEEE Vehicular Technology Conference (VTC '07)*, pp. 2761–2765, April 2007.

[14] G. Mongha, K. I. Pedersen, I. Z. Kovacs, and P. E. Mogensen, "QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '08)*, pp. 2532–2536, May 2008.

[15] R. Kwan, C. Leung, and J. Zhang, "Multiuser scheduling on the downlink of an lte cellular system," *Research Letters in Communication*, vol. 2008, Article ID 323048, 4 pages, 2008.

[16] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–153, 2001.

[17] A. R. Braga, E. B. Rodrigues, and F. R. P. Cavalcanti, "Packet scheduling for VoIP over HSDPA in mixed traffic scenarios," in *Proceedings of the 17th IEEE International Symposium Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, pp. 1–5, Helsink, Finland, September 2006.

[18] B. Chen, H. Hu, B. Wang, and H. Wang, "A novel multi-service scheduling scheme for E-UTRA," in *Proceedings of the 3rd IEEE/IFIP International Conference in Central Asia on Internet (ICI '07)*, pp. 1–5, September 2007.

[19] S. Choi, K. Jun, Y. Shin, S. Kang, and B. Choi, "MAC scheduling scheme for VoIP traffic service in 3G LTE," in *Proceedings of the 66th IEEE Vehicular Technology Conference (VTC '07)*, pp. 1441–1445, Baltimore, Md, USA, October 2007.

[20] S. Wang, Y. Gao, X. Gu, H. Tian, and P. Zhang, "Packet scheduling for multimedia traffics in downlink multi-user OFDM systems," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '06)*, pp. 1–4, September 2006.

[21] M. Gidlund and J.-C. Laneri, "Scheduling algorithms for 3GPP long-term evolution systems: from a quality of service perspective," in *Proceedings of the 10th IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA '08)*, pp. 114–117, August 2008.

[22] B. Sadiq, R. Madan, and A. Sampath, "Downlink scheduling for multiclass traffic in LTE," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 510617, 18 pages, 2009.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1991.

[24] A. Furuskär, *Radio resource sharing and bearer service allocation for multi-bearer service, multi-access wireless networks*, Ph.D. thesis, Royal Institute of Technology (KTH), Radio Communication Systems, April 2003.

[25] M. Ericson and S. Wänstedt, "Mixed traffic HSDPA scheduling—impact on VoIP capacity," in *Proceedings of the 65th IEEE Vehicular Technology Conference (VTC '07)*, pp. 1282–1286, Dublin, Ireland, April 2007.

[26] A. Furuskär and J. Zander, "Multiservice allocation for multiaccess wireless systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 174–183, 2005.

[27] E. B. Rodrigues, F. R. P. Cavalcanti, and S. Wänstedt, "QoS-driven adaptive congestion control for voice over IP in multiservice wireless cellular networks," *IEEE Communications Magazine*, vol. 46, no. 1, pp. 100–107, 2008.

[28] D. M. Sacristán, J. F. Monserrat, J. Cabrejas-Penuelas, D. Calabuig, S. Garrigas, and N. Cardona, "On the way towards fourth-generation mobile: 3GPP LTE and LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 354089, 10 pages, 2009.

[29] S. Parkvall, E. Dahlman, A. Furuskär et al., "LTE-advanced—evolving LTE towards IMT-advanced," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '08)*, pp. 1–5, September 2008.

[30] D. Astély, E. Dahlman, A. Furuskär, Y. Jading, M. Lindström, and S. Parkvall, "LTE: the evolution of mobile broadband," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 44–51, 2009.

[31] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); long term evolution (LTE) physical layer; general description," Tech. Rep. TS 36.201 V8.2.0—Release 8, 3rd Generation Partnership Project, December 2008.

[32] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification," Tech. Rep. TS36.321 V8.4.0—Release 8, 3rd Generation Partnership Project, December 2008.

[33] N. Enderlé and X. Lagrange, "User satisfaction models and scheduling algorithms for packet-switched services in UMTS," in *Proceedings of the 57th IEEE Semiannual Vehicular Technology Conference (VTC '03)*, vol. 3, pp. 1704–1709, Jeju, South Korea, April 2003.

[34] L. Badia, M. Boaretto, and M. Zorzi, "A users' satisfaction driven scheduling strategy for wireless multimedia QoS," in *Proceedings of the Quality of Future Internet Services (QoFIS '03)*, vol. 2811, pp. 203–213, Stockholm, Sweden, October 2003.

[35] 3GPP, "Technical Specification Group Services and System Aspects; policy and charging control architecture," Tech. Rep. TS 23.203V9.4.0—Release 9, 3rd Generation Partnership Project, March 2010.

[36] 3GPP, "TSG-SA Codec Working Group: Mandatory speech codec; AMR speech codec; interface to Iu and Uu," Tech. Rep. TS26.102, 3rd Generation Partnership Project, 1999.

[37] 3GPP, "Physical layer aspects for evolved universal terrestrial radio access (UTRA)," Tech. Rep. TR 25.814 V7.1.0—Release 7, 3rd Generation Partnership Project, September 2006.

[38] 3GPP, "Spatial channel model for multiple input multiple output (MIMO) simulations," Tech. Rep. TR 25.996 V8.0.0, 3rd Generation Partnership Project, December 2008.

[39] F. Persson, "Voice over IP realized for the 3GPP long term evolution," in *Proceedings of the 66th IEEE Vehicular Technology Conference (VTC '07)*, pp. 1436–1440, October 2007.

[40] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1150–1158, 2003.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2010, Article ID 316749, 9 pages
doi:10.1155/2010/316749

Research Article

Hierarchical Modulation with Vector Rotation for E-MBMS Transmission in LTE Systems

Hui Zhao, Xiaoping Zhou, Yunchuan Yang, and Wenbo Wang

Key Laboratory of Universal Wireless Communication, Beijing University of Posts and Telecommunications, Ministry of Education, Xitucheng Road no.10, Haidian District, Beijing 100029, China

Correspondence should be addressed to Hui Zhao, hzhao@bupt.edu.cn

Received 1 April 2010; Revised 4 July 2010; Accepted 6 August 2010

Academic Editor: Raymond Kwan

Enhanced Multimedia Broadcast and Multicast Service (E-MBMS) is considered of key importance for the proliferation of Long-Term Evolution (LTE) network in mobile market. Hierarchical modulation (HM), which involves a "base-layer" (BL) and an "enhancement-layer" (EL) bit streams, is a simple technique for achieving tradeoff between service quality and radio coverage. Therefore, it is appealing for MBMS. Generally, HM suffers from the severe performance degradation of the less protected EL stream. In this paper, HM with vector rotation operation introduced to EL stream is proposed, in order to improve EL's performance. With the proper interleaving in frequency domain, this operation can exploit the inherent diversity gain from the multipath channel. In this way, HM with vector rotation can effectively enhance multimedia broadcasting on quality video and coverage. The simulation results with scalable video coding (SVC) as source show the significant benefits in comparison with the conventional HM and alternative schemes.

1. Introduction

Video broadcast and multicast are expected to constitute a significant portion of the load on future mobile communication systems. MBMS and E-MBMS of 3GPP provide the means of delivering mobile TV to the mass market, which has already been standardized in 3GPP UTRAN (UMTS Terrestrial Radio Access Network) Release-6, 7, and 8 (R8 is just LTE). But it is still open in R10 (i.e., LTE-Advanced). In this paper, we will discuss the radio transmission of E-MBMS with point-to-multipoint connections. Regarding the feature of multimedia service, the source often adopts the multiple resolution code [1]. Hence, unequal error protection (UEP) transmission is a natural support for this kind of information [2]. Namely, the coarse resolution gets a better protection than the fine resolution. Therefore, the coarse resolution information will be recovered by all receivers to obtain the crude reconstruction while the fine resolution information can only be recovered by some receivers due to propagation conditions. In short, the combination of multiple resolution source code and UEP transmission can balance tradeoff between video quality and robust coverage in media broadcast applications.

As a specific example of UEP transmission, HM is widely used in digital broadcasting systems, such as, DVB-T, MediaFLO, and UMB-BCMCS. HM consists of non-uniform signal constellations to provide unequal bit error protection for BL and EL streams. Many researches have been carried out on this topic. Some literatures discussed the application of HM in multiple input multiple output (MIMO) systems [3, 4] or in LTE system based on Orthogonal Frequency Division Multiplexing (OFDM) [5, 6], while some others reported the joint optimization between HM and separate/joint source-channel coding in [7, 8]. Providing much attention to HM itself, [9] has proposed an enhanced HM scheme along with three optimization criteria, in which the signal constellation of EL stream is rotated by a certain angle, named as "HM with CR" in this paper. The authors of [10] presented a concatenated encoding scheme with HM to well protect the EL stream, which greatly complicates the transceiver structure.

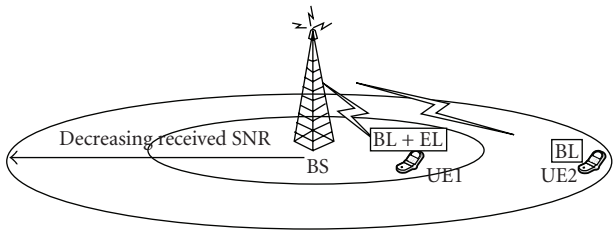


Figure 1: A broadcast model with HM.

This paper also aims to improve the protection of the EL stream through a simple approach. As we all know, to efficiently decrease the bit-error probability (BER) further, we can resort to exploit the diversity gain under high received signal to noise ratio (SNR) condition or the coding gain in low SNR region. And the user at the center of cell who can recover the EL signal generally has a higher SNR than that at the edge, as illustrated in Figure 1. Hence, this paper applies the idea of signal space diversity (SSD) [11] to provide diversity gain to the EL stream, which is termed as "HM with vector rotation (VR)". Rough predictions on video quality and coverage are given using SVC video as media example. The simulation results in LTE E-MBMS system show that our scheme is better than the existing HM and "HM with CR" due to the diversity gain at the cost of a slight increase in detection complexity.

The rest of the paper is organized as follows: Section 2 introduces the system model of an OFDM system with HM, reviews the definition of "HM with CR", and presents the existing problem. In Section 3, we propose "HM with VR". After the scheme description of transceiver, we give the analysis of diversity gain of EL stream and receiver complexity, comparing with the classical HM. Subsequently, the simulation in Sections 4 and 5 verifies the benefits of our proposed scheme in the video quality and the radio coverage, respectively. Finally, conclusions are drawn in Section 6.

2. System Model and Problem Description

Compared to the conventional modulation methods (such as, QPSK, 16-QAM, and 64-QAM, etc.), HM allows the multiplexing transmission of multilayer streams in superposition with different transmission qualities, which stem from the concept of superposition coding in [12]. Without loss of generality, only the two-layer transmission is considered in this paper, that is, one BL and one EL.

2.1. System Model with HM. In Figure 2, we show a transceiver diagram that incorporates HM into a UTRA LTE-based OFDM transmission. Each stream is encoded and mapped to the constellation symbols according to the importance. That is to say, BL stream is mapped to the most significant bits (MSB) and EL to the less significant bits (LSB). After the serial/parallel (S/P) operation, the symbols are then converted to the time domain using Inverse Fast Fourier Transform (IFFT). With the added cyclic prex (CP), the OFDM symbols are sent over the frequency-selective multipath fading channel. The receiver removes the CP performs FFT and P/S operation. Then the received signal at the kth subcarrier in frequency-domain can be written as

$$y_k = h_k x_k + n_k, \tag{1}$$

where h_k is the channel fading coefficient in frequency domain, x_k is the complex hierarchical symbol, n_k is assumed to be zero-mean circularly symmetric complex Gaussian noise with distribution $n_k \sim N_c(0, \sigma_n^2)$. We only discuss this system model in frequency domain, thus the subscript k can be omitted for short later.

According to the definition of HM, the HM symbol x can be rewritten as

$$x = \alpha x^{\mathrm{bl}} + \beta x^{\mathrm{el}}, \tag{2}$$

where x^{bl} and x^{el} are the symbols, respectively, mapped to MSB and LSB; α and β are the power parameters for BL and EL and satisfy $\alpha > \beta$; hence, the power ratio between the two layers is defined as

$$\eta = \frac{\alpha^2}{(\alpha^2 + \beta^2)}. \tag{3}$$

This ratio will determine how much more the BL stream is protected against errors than is the EL stream.

The receiver is assumed to have the perfect channel state information. After channel equalization and soft demodulation, the two streams with soft information are sent to the separate decoders.

2.2. HM with CR. The regular HM suffers from interlayer interference. In order to recover the capacity loss due to this interference, [9] has presented an enhanced "HM with CR" from an information-theoretical perspective. Corresponding with the regular QPSK/16 QAM HM in Figure 3(a), the principle of "HM with CR" is shown in Figure 3(b). The symbol of EL is rotated in counter-clockwise by θ ($0 \leqslant \theta \leqslant \pi/4$). The resulted symbol is expressed as

$$x = \alpha x^{\mathrm{bl}} + \beta e^{j\theta} x^{\mathrm{el}}. \tag{4}$$

The optimal θ is chosen to maximize the minimum Euclidean distance (MED) of the resulted constellation points, which will lead to a better performance.

2.3. Problem Description. As shown in Figure 1, in a broadcasting system with HM, the users at the edge of cell (e.g., UE2) experience the low SNR channel due to the long distance to the base station (BS). For guaranteeing the recover of the BL information at the edge, the power allocated to BL should be large enough, while the power allocated to EL must be cut down with the constant transmit power constraint. However, if the EL stream can be well protected also, the users at the center (e.g., UE1) will enjoy the higher video quality. In other words, with the same video quality, the coverage area will be enlarged. As we all know, regarding to the BER performance, the diversity gain becomes critical when SNR is high and the coding gain when

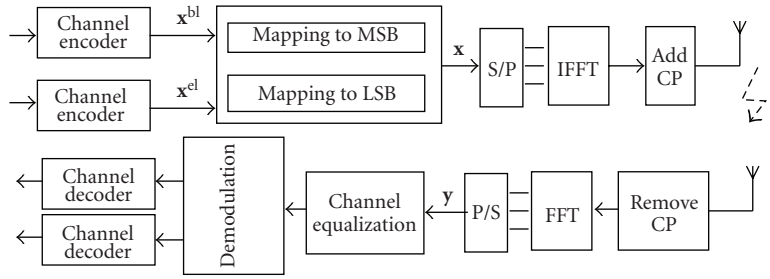


Figure 2: Transceiver diagram of OFDM system with HM.

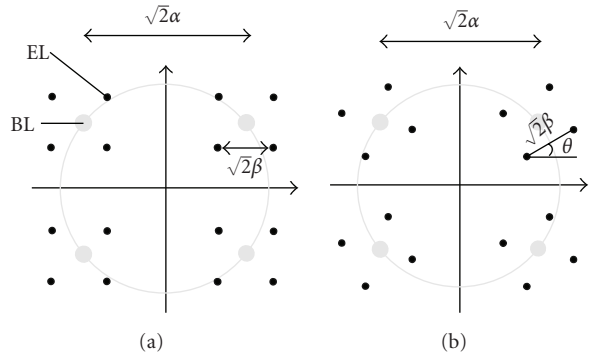


Figure 3: QPSK/16 QAM HM constellation for (a) regular HM (b) "HM with CR".

SNR is in medium or low region. If we neglect the influence of interlayer interference, it is conjectured that the SNR of UE1's EL signal is higher than that of UE2's BL signal in the most cases. Accordingly, we will exploit the diversity gain for EL while BL has earned the coding gain due to the large allocated power.

3. Proposed HM with Vector Rotation

SSD, also named modulation diversity, is a technique to provide the diversity gain for a single antenna system in the fast fading channel. To simply explain, SSD is just used to multiply an N-by-N precoding matrix to a group of N modulated symbols (i.e., vector rotation), then let the precoded symbols experience the fully interleaved fading channel. The proper choice of precoding matrix can earn full diversity of N for the symbols. In this paper, we apply SSD in HM for improving the EL performance.

3.1. Scheme Description. Figure 4 illustrates the transceiver diagram of the proposed "HM with VR". The difference between Figures 4 and 2 lies in the following. First, the N continuous EL data symbols mapped to LSB are precoded by a rotation matrix $\mathbf{R}$; second, a pseudorandom symbol interleaving in frequency domain is implemented and the interleaving duration is limited in the range of a codeword; finally, the two streams are separately detected. Here are the details.

Analogous to (2) and (4), the N hierarchical symbols x_i $(i = 1, 2, \ldots, N)$ after the operation of "HM with VR" compose a hierarchical vector as

$$\mathbf{x} = [x_1, x_2, \ldots, x_N]^T = \alpha \mathbf{x}^{bl} + \beta \mathbf{R} \mathbf{x}^{el}, \tag{5}$$

where $\mathbf{x}^{bl} = [x_1^{bl}, x_2^{bl}, \ldots, x_N^{bl}]^T$ and $\mathbf{x}^{el} = [x_1^{el}, x_2^{el}, \ldots, x_N^{el}]^T$ are the BL and EL vectors, respectively; $\mathbf{R}$ is an N-by-N unitary matrix, referred to [11]. Then in the N continuous symbol durations, the equivalent system model in the frequency domain is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{6}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$ is the received vector; $\mathbf{H} = \mathrm{diag}(h_1 \quad h_2 \quad \cdots \quad h_N)$ is the equivalent channel matrix, whose elements can be regarded as mutually independent due to the assumption of ideal symbol interleaving in frequency domain; $\mathbf{n} = [n_1, n_2, \ldots, n_N]^T$ is the noise vector with zero mean and correlation matrix $\sigma_n^2 \mathbf{I}_N$; here $\mathbf{I}_N$ denotes an N-by-N identity matrix.

Regarding demodulation, the bits of both BL and EL streams are jointly judged in HM. But as to "HM with VR", we apply the separate detection to each stream and the interference cancellation (IC) for the sake of complexity simplification. First, the BL data is detected. Here the interference from EL is regarded as a part of the general noise $\mathbf{n}'$:

$$\hat{\mathbf{x}}^{bl} = \alpha^{-1} \mathbf{H}^{-1} \mathbf{y} = \mathbf{x}^{bl} + \mathbf{n}', \tag{7}$$

where $\mathbf{H}^{-1}$ is the inverse of $\mathbf{H}$.

Then the interference on EL from BL is regenerated and cancelled from the received vector. The resulting received vector of the single EL is

$$\mathbf{y}^{el} = \mathbf{y} - \alpha \mathbf{H} \cdot \mathcal{Z}\left(\hat{\mathbf{x}}^{bl}\right) = \beta \mathbf{H} \mathbf{R} \mathbf{x}^{el} + \mathbf{n}'', \tag{8}$$

where $\mathcal{Z}(\hat{\mathbf{x}}^{bl})$ means demodulating each element of $\hat{\mathbf{x}}^{bl}$ to the nearest constellation symbol; let $\Delta \mathbf{x}^{bl} = \mathcal{Z}(\hat{\mathbf{x}}^{bl}) - \mathbf{x}^{bl}$, which denotes the demodulation error, then $\mathbf{n}'' = \alpha \mathbf{H} \cdot \Delta \mathbf{x}^{bl} + \mathbf{n}$ is zero mean and correlation matrix $\sigma_{n''}^2 \mathbf{I}_N = \alpha^2 \,\mathrm{diag}(|h_1 \cdot \Delta x_1^{bl}|^2, |h_2 \cdot \Delta x_2^{bl}|^2, \ldots |h_N \cdot \Delta x_N^{bl}|^2) + \sigma_n^2 \mathbf{I}_N$.

Finally, implement N-symbol maximum likelihood (ML) or other near-ML low-complexity detection to $\mathbf{y}^{el}$ to get the EL data estimation $\hat{\mathbf{x}}^{el}$.

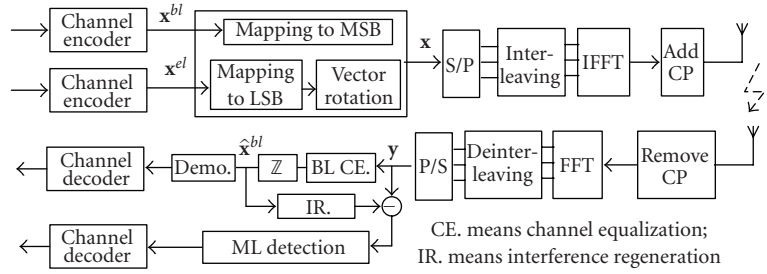


Figure 4: Transceiver diagram of the proposed "HM with VR".

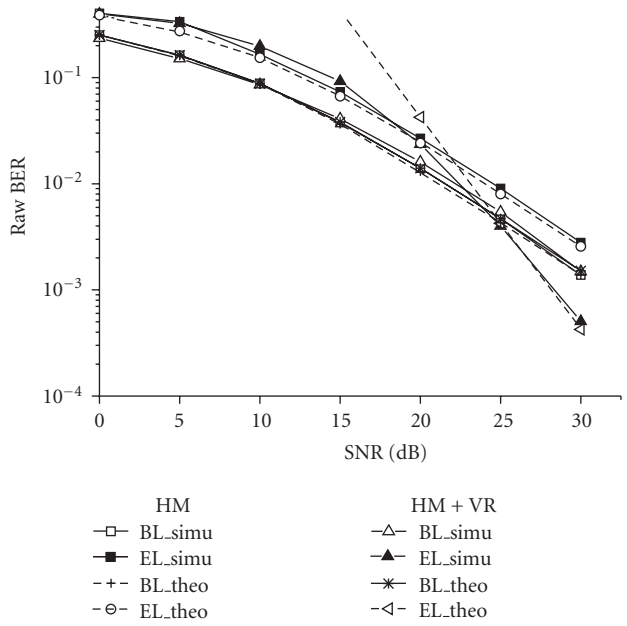


Figure 5: Validation of Raw BER formula.

3.2. Analysis on Error Probability. It has been demonstrated in [11] that the symbol diversity order of SSD is equal to the dimension of square rotation matrix N when rotation matrix is designed reasonably and the N rotated symbols are detected jointly. Although the larger N is leading to the more diversity gain, the value of N cannot be too large in this paper, because the performance of BL will deteriorate due to the interference from EL when N is large. Furthermore, with the consideration of detection complexity, we take $N = 2$ in this paper. Here, we give the performance comparison between "HM with VR" and the classical HM by the derivation of bit error probability (BER) for both BL and EL streams.

In [13], the author has deduced the BER expressions of the generalized hierarchical-QAM constellations over fading channel. Hence, the BER formulas of each stream of HM and BL stream of "HM with VR" can be obtained directly with reference to [13]. As to "HM with CR", the BER expression of EL stream is hard to derive due to the complex decision region. Hence, we only give the simulation comparison in Figure 5.

Before deducing the BER of EL stream of "HM with VR", we give its symbol pairwise error probability (SPEP) firstly to analyze the diversity order. Here we use moment generation function- (MGF-) based approach, different with the canonical probability density function- (PDF-) based approach used in [11].

(1) SPEP. Equation (8) can be rewritten as (the value of $\mathbf{R}$ is from [11])

$$\mathbf{y}^{\mathrm{el}} = \beta\mathbf{H}\mathbf{R}\mathbf{x}^{\mathrm{el}} + \mathbf{n}'' = \underbrace{\beta\begin{bmatrix} h_1 & 0 \\ 0 & h_2 \end{bmatrix}\begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}}_{\widetilde{\mathbf{H}}}\mathbf{x}^{\mathrm{el}} + \mathbf{n}'', \tag{9}$$

where $\phi = 0.5\arctan(2)$; $\widetilde{\mathbf{H}}$ is the equivalent channel matrix; here, we observe the diversity order at high SNR so that the error propagation from BL can be neglected and $\mathbf{n}''$ has the same distribution as $\mathbf{n}$.

Let the difference vector $\Delta\mathbf{x}^{\mathrm{el}} = \mathbf{x}^{\mathrm{el}} - \hat{\mathbf{x}}^{\mathrm{el}} = [\delta_1 \quad \delta_2]^T$; the conditional SPEP for $\mathbf{x}^{\mathrm{el}}$ can be expressed by the Gaussian Q-function as [14]

$$P\left(\mathbf{x}^{\mathrm{el}} \neq \hat{\mathbf{x}}^{\mathrm{el}} \mid \mathbf{H}\right) = Q\left(\sqrt{\frac{\left|\mathbf{H}\mathbf{R}\cdot\Delta\mathbf{x}^{\mathrm{el}}\right|^2}{2\sigma_n^2/\beta^2}}\right) = Q\left(\sqrt{\frac{\mu}{2\sigma_n^2/\beta^2}}\right), \tag{10}$$

where $\mu \triangleq \left|\mathbf{H}\mathbf{R}\Delta\mathbf{x}^{\mathrm{el}}\right|^2 = (\mathbf{R}\cdot\Delta\mathbf{x}^{\mathrm{el}})^H\mathbf{H}^H\mathbf{H}\mathbf{R}\cdot\Delta\mathbf{x}^{\mathrm{el}}$. According to the definition, one has

$$\mu = |h_1|^2\omega_1^2 + |h_2|^2\omega_2^2, \tag{11}$$

where $\omega_1 = |\delta_1\cos\varphi + \delta_2\sin\varphi|$ and $\omega_2 = |\delta_1\sin\varphi - \delta_2\cos\varphi|$ are real.

We now apply the Craig's formula [15] to derive the conditional SPEP in (10):

$$\begin{aligned} P\left(\mathbf{x}^{\mathrm{el}} \neq \hat{\mathbf{x}}^{\mathrm{el}} \mid \mathbf{H}\right) &= Q\left(\sqrt{\frac{\mu}{2\sigma_n^2/\beta^2}}\right) \\ &= \frac{1}{\pi}\int_0^{\pi/2}\exp\left(\frac{-\mu}{4\sin^2\theta\cdot\sigma_n^2/\beta^2}\right)d\theta \\ &= \frac{1}{\pi}\int_0^{\pi/2}\exp\left(-\frac{|h_1|^2\omega_1^2 + |h_2|^2\omega_2^2}{4\sin^2\theta\cdot\sigma_n^2/\beta^2}\right)d\theta. \end{aligned} \tag{12}$$

Since h_1 and h_2 are i.i.d $\sim \mathrm{N}_c(0,1)$, we apply MGF-based approach [14] to obtain the unconditional SPEP as follows:

$$P\left(\mathbf{x}^{\mathrm{el}} \neq \hat{\mathbf{x}}^{\mathrm{el}}\right) = \frac{1}{\pi}\int_0^{\pi/2}\left[\left(1+\frac{\beta^2\omega_1^2}{4\sigma_n^2\sin^2\theta}\right)\left(1+\frac{\beta^2\omega_2^2}{4\sigma_n^2\sin^2\theta}\right)\right]^{-1}d\theta. \tag{13}$$

At high SNR, the upper bound of (13) is given by

$$\begin{aligned} P\left(\mathbf{x}^{\mathrm{el}} \neq \hat{\mathbf{x}}^{\mathrm{el}}\right) &\leq \frac{1}{\pi}\int_0^{\pi/2}\left[\left(\frac{\beta^2\omega_1^2}{4\sigma_n^2\sin^2\theta}\right)\left(\frac{\beta^2\omega_2^2}{4\sigma_n^2\sin^2\theta}\right)\right]^{-1}d\theta \\ &= \frac{1}{\pi}\left[\frac{\beta^2\omega_1^2}{4\sigma_n^2}\cdot\frac{\beta^2\omega_2^2}{4\sigma_n^2}\right]^{-1}\int_0^{\pi/2}\sin^4\theta d\theta \\ &= \frac{12}{\beta^4\omega_1^2\omega_2^2}\left(\frac{2}{\sigma_n^2}\right)^{-2}. \end{aligned} \tag{14}$$

Because the transmission power is 2 when $N = 2$, $(2/\sigma_n^2)$ in (14) is just SNR. From (14), we could easily find that the exponent of SNR is -2, that is, diversity order is two. But for HM, the hierarchical modulated symbols only experience one subcarrier channel so as to only earn the symbol diversity of 1.

Further, we can find the union bound on the average SPEP, denoted by $\mathrm{PEP}^{\mathrm{el}}$, by traversing all the possible vectors, the size of which is L

$$\mathrm{PEP}^{\mathrm{el}} = \frac{2}{L}\sum_{i=1}^{L-1}\sum_{j=i+1}^{L} P\left(\mathbf{x}_i^{\mathrm{el}} \neq \mathbf{x}_j^{\mathrm{el}}\right). \tag{15}$$

(2) BER. The two symbols of EL are detected jointly, so the error probability of the ith symbol x_i^{el} is related to the other (denoted as $x_{\bar{i}}^{\mathrm{el}}$, $i, \bar{i} \in [1,2]$), which can be expressed as

$$\begin{aligned} P\left(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}}\right) &= \underbrace{P\left(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}} \mid x_{\bar{i}}^{\mathrm{el}} = \hat{x}_{\bar{i}}^{\mathrm{el}}\right)}_{A}\left[1 - P\left(x_{\bar{i}}^{\mathrm{el}} \neq \hat{x}_{\bar{i}}^{\mathrm{el}}\right)\right] \\ &\quad + \underbrace{P\left(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}} \mid x_{\bar{i}}^{\mathrm{el}} \neq \hat{x}_{\bar{i}}^{\mathrm{el}}\right)}_{B} P\left(x_{\bar{i}}^{\mathrm{el}} \neq \hat{x}_{\bar{i}}^{\mathrm{el}}\right). \end{aligned} \tag{16}$$

Because both x_i^{el} and $x_{\bar{i}}^{\mathrm{el}}$ are equiprobably chosen from the same constellation, their symbol error probability (SER) is the same, that is, $P(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}}) = P(x_{\bar{i}}^{\mathrm{el}} \neq \hat{x}_{\bar{i}}^{\mathrm{el}}) = \mathrm{SER}^{\mathrm{el}}$. Then, (16) becomes

$$\mathrm{SER}^{\mathrm{el}} = A \cdot \left(1 - \mathrm{SER}^{\mathrm{el}}\right) + B \cdot \mathrm{SER}^{\mathrm{el}}. \tag{17}$$

Here, A can be derived because the ML detection of x_i^{el} is equivalent to the maximal-ratio combining in a single-input multiple-output system since $x_i^{\mathrm{el}} = \hat{x}_{\bar{i}}^{\mathrm{el}}$. In the appendix, we give the expression of A by the same MGF-based approach. In the process of the ML detection, no operation of hard decision is executed and $\hat{x}_{\bar{i}}^{\mathrm{el}}$ is just a soft symbol. Hence, if $x_{\bar{i}}^{\mathrm{el}} \neq \hat{x}_{\bar{i}}^{\mathrm{el}}$, the probability of decision error cannot be deduced easily so as to hard to derive the expression of B. Here, we utilize the known SPEP to derive the SER. The SPEP also can be expressed as

$$\begin{aligned} \mathrm{PEP}^{\mathrm{el}} = \sum_{x_i^{\mathrm{el}}, x_{\bar{i}}^{\mathrm{el}} \in C} & P\left(x_i^{\mathrm{el}}\right)P\left(x_{\bar{i}}^{\mathrm{el}}\right) \\ & \cdot \left\{2P\left(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}} \mid x_{\bar{i}}^{\mathrm{el}} = \hat{x}_{\bar{i}}^{\mathrm{el}}\right)\left[1 - P\left(x_{\bar{i}}^{\mathrm{el}} \neq \hat{x}_{\bar{i}}^{\mathrm{el}}\right)\right]\right. \\ & \left. + P\left(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}} \mid x_{\bar{i}}^{\mathrm{el}} \neq \hat{x}_{\bar{i}}^{\mathrm{el}}\right)P\left(x_{\bar{i}}^{\mathrm{el}} \neq \hat{x}_{\bar{i}}^{\mathrm{el}}\right)\right\}. \end{aligned} \tag{18}$$

Utilizing the definition of A and B in (16), (18) can be simplified as

$$\mathrm{PEP}^{\mathrm{el}} = 2A \cdot \left(1 - \mathrm{SER}^{\mathrm{el}}\right) + B \cdot \mathrm{SER}^{\mathrm{el}}. \tag{19}$$

Then, from (17) and (19), we can get the following closed-form expression of SER:

$$\mathrm{SER}^{\mathrm{el}} = \frac{\mathrm{PEP}^{\mathrm{el}} - A}{1 - A}. \tag{20}$$

Moreover, the raw BER is scaled to the SER for the square M_c-QAM modulation. Therefore, we yield the BER of EL stream as

$$\mathrm{BER}^{\mathrm{el}} = \frac{\mathrm{SER}^{\mathrm{el}}}{\log_2(M_c)}. \tag{21}$$

In the following, we validate the above deduction by contrasting with the simulation results. As shown in Figure 5, both theoretical and simulated curves of the raw BER for EL of "HM + VR" are overlapped nearly at the medium and high SNR region. This is because our ideal assumption of no error spread in IC is hold when SNR is high enough and what we can calculate is just the upper bound of BER in fact. On the other hand, the argument about diversity gain is also demonstrated by observing the slope of these curves.

3.3. Analysis on Detection Complexity. We discuss the detection complexity for "HM with VR". Here, we apply the QRM [16] detection to (9), which is a popular near-ML low-complexity algorithm and explained as follows: first, do QR decomposition to $\widetilde{\mathbf{H}}$, as shown in (22); second, left-multiply $\mathbf{Q}^H$ to $\mathbf{y}^{\mathrm{el}}$ of (9); then detect x_i^{el} (i: from 2 to 1) in turn because of the up-triangular form of $\mathbf{R}$.

$$\widetilde{\mathbf{H}} = \beta\begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix}\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} = \mathbf{QR}. \tag{22}$$

Here, due to $N = 2$, QR decomposition can be executed by a simple and ingenious way. Just let

$$\mathbf{Q} = c\begin{pmatrix} h_1\cos\varphi & h_2^*\sin\varphi \\ -h_2\sin\varphi & h_1^*\cos\varphi \end{pmatrix}, \tag{23}$$

where $c = \beta^{-1}(|h_1|^2 + |h_2|^2)^{-1}$. Then

$$\mathbf{R} = \mathbf{Q}^H \tilde{\mathbf{H}} = c \begin{pmatrix} \underbrace{|h_1|^2\cos^2\varphi + |h_2|^2\sin^2\varphi}_{r_{1,1}} & \underbrace{(|h_1|^2 - |h_2|^2)\sin\varphi\cos\varphi}_{r_{1,2}} \\ 0 & \underbrace{h_1 h_2}_{r_{2,2}} \end{pmatrix}. \tag{24}$$

The QR decomposition is completed. This process only needs 13 real multiplications. And the complexity of this preprocessing of detection can nearly be ignored because the channel always keeps constant in a long symbol duration.

In QRM, the heavy complexity lies in the metric computation for each possible symbol candidate. The parameter M, which means the number of surviving candidates, can make a tradeoff between complexity and performance. Here, $M = 1 \sim 4$ for detecting EL data for QPSK/16 QAM "HM with VR". Table 1 shows the complexity contrast of metric computation between HM/"HM with CR" and "HM with VR" in detail, where $\mathbf{Q}(i,:)$ denotes the ith row of $\mathbf{Q}$. When M is small, the times of complex multiplication for "HM with VR" is close to that for HM/"HM with CR". At this time, the performance loss is also acceptable for QPSK/16 QAM "HM with VR".

4. Performance on Video Quality

For a fair evaluation of the different transmission schemes in LTE E-MBMS system, the error probability in transmission link is not a proper metric at all. In this paper, we measure the quality of the video with the metric of average peak-signal-to-noise ratio (PSNR) of the video frames, defined as the ratio of the squared maximum pixel value to the mean square error of the reconstructed lossy image at the receiver [17]. In the following, we describe how to estimate the PSNR for the given source rate, and received SNR in transmission link.

4.1. Simulation Method. The derivation of PSNR involves the following three steps.

Step 1 (Get the packet error probability (PER)). For a given information bit rate pair of BL and EL streams $(R^{\mathrm{bl}}, R^{\mathrm{el}})$, we can get their PER pair at the different SNR by Monte Carlo simulation, denoted as $(\mathrm{PER}^{\mathrm{bl}}, \mathrm{PER}^{\mathrm{el}})$. It is noted that whether an EL packet is correctly detected or not is on the premise of accurately detecting the BL packet yet.

Step 2 (Estimate the PSNR). We assume that the relation between the bit rate R and the video distortion D (in terms of PSNR) of the transmitted video is known, that is, $D = \mathbb{Q}(R)$, where $\mathbb{Q}(R)$ denotes the operational rate-distortion function of the source coder. The scalable video stream is assumed to truncate at any rate. Hence, the distortion when only BL is correctly detected will be $D^{\mathrm{bl}} = \mathbb{Q}(R^{\mathrm{bl}})$ and the distortion when both BL and EL are correctly detected will be $D^{\mathrm{be}} = \mathbb{Q}(R^{\mathrm{bl}} + R^{\mathrm{el}})$. Therefore, if the probability of correctly detecting a packet for each layer is known, the expected PSNR of the received layered video can be found by (referred to [18])

$$\begin{aligned} D_{\exp}\left(\mathrm{SNR}, R^{\mathrm{bl}}, R^{\mathrm{el}}\right) &= \mathbb{F}_{\mathrm{be}}\left(\mathrm{SNR}, R^{\mathrm{bl}}, R^{\mathrm{el}}\right) \cdot \mathbb{Q}\left(R^{\mathrm{bl}} + R^{\mathrm{el}}\right) \\ &\quad + \left[\mathbb{F}_{\mathrm{b}}\left(\mathrm{SNR}, R^{\mathrm{bl}}\right) - \mathbb{F}_{\mathrm{be}}\left(\mathrm{SNR}, R^{\mathrm{bl}}, R^{\mathrm{el}}\right)\right] \cdot \mathbb{Q}\left(R^{\mathrm{bl}}\right) \\ &= \left[1 - \mathrm{PER}^{\mathrm{el}}\right] \cdot \mathbb{Q}\left(R^{\mathrm{bl}} + R^{\mathrm{el}}\right) + \left(1 - \mathrm{PER}^{\mathrm{bl}}\right) \\ &\quad \cdot \mathrm{PER}^{\mathrm{el}} \cdot \mathbb{Q}\left(R^{\mathrm{bl}}\right), \end{aligned} \tag{25}$$

where $\mathbb{F}_{\mathrm{b}}$ is the probability of correctly detecting a packet of BL, which is a function of $(\mathrm{SNR}, R^{\mathrm{bl}})$; $\mathbb{F}_{\mathrm{be}}$ the probability of detecting both BL and EL, which is a function of $(\mathrm{SNR}, R^{\mathrm{bl}}, R^{\mathrm{el}})$. Then $[\mathbb{F}_{\mathrm{b}} - \mathbb{F}_{\mathrm{be}}]$ represents probability of detecting only BL but not EL.

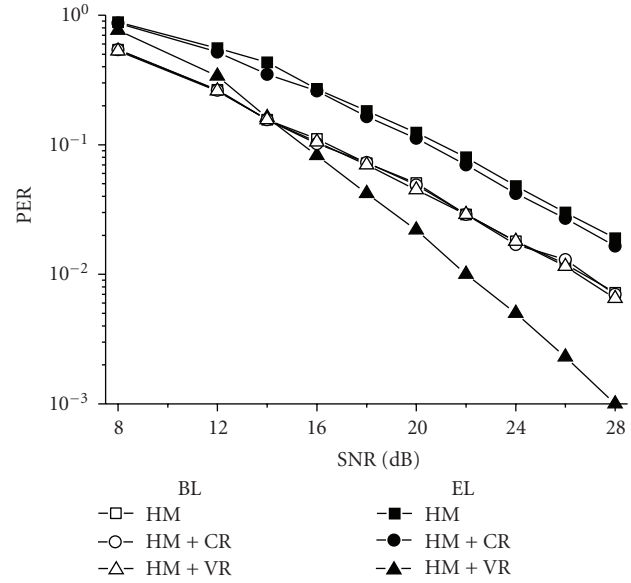


Figure 6: Comparison in link performance.

4.2. Simulation Results. All the parameters used in our simulations are based on UTRA LTE 3GPP documents [19]. Table 2 shows the respective parameters. Due to the effect of macrodiversity, the MBSFN channel is a 18-path model, in which frequency-selectivity is enough. The power ratio for HM is constant in the simulation of this section, that is, $\eta = 0.7$. Typical data rates for MBMS video streams in CIF resolution is 256 kbps. So we set both R^{bl} and R^{el} as 256 kbits/s. The used rate-distortion function is referred to [20].

Figure 6 gives the PER of each layer for the three schemes. Their BL performance is similar. However, owning to the higher diversity gain, the EL of "HM with VR" outperforms the other two ELs obviously and is even better than BL at high SNR. That is because the BL of the users at the center

Table 1: Complexity comparison between HM and "HM with VR".

	Operation for detecting symbols of two continuous intervals	Complex multiplication	Complex addition	Module	Times
HM/HM with CR	$\lvert y_i/h_i - \hat{x}_i \rvert \times 2$	1	1	1	$2M_c$
HM with VR	$\lvert y_i^{\mathrm{bl}}/h_i - \hat{x}_i^{\mathrm{bl}} \rvert \times 2$	1	1	1	$M_c/4$
	$\lvert \mathbf{Q}(2,:)\mathbf{y}^{\mathrm{el}} - r_{2,2}\hat{x}_2^{\mathrm{el}} \rvert$	3	2	1	$M_c/4$
	$\lvert \mathbf{Q}(1,:)\mathbf{y}^{\mathrm{el}} - r_{1,2}\hat{x}_2^{\mathrm{el}} - r_{1,1}\hat{x}_1^{\mathrm{el}} \rvert$	3	2	1	$M \cdot M_c/4$

Table 2: LTE E-MBMS system parameters in simulation.

Parameters	Values
Carrier frequency	2.0 GHz
Bandwidth	10 MHz
Sampling frequency	15.36 MHz
Subcarrier space	15 KHz
Used resource block	2
Number of antennas	1×1
Simulation unit	1 subframe
OFDM symbols	12 symbols/subframe
FFT size	1024
CP size	256
Channel model	MBSFN Propagation Channel [21]
Velocity	2.7 km/h
Modulation	HM, HM + CR, HM + VR (QPSK/16 QAM)
Code rate	0.5

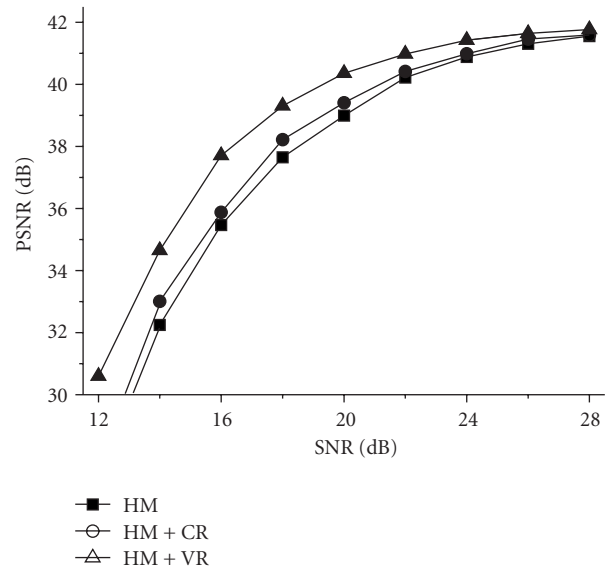


Figure 7: Comparison in video quality.

is protected enough. At this time, it is not necessary to stress that BL should get more protection than EL, although this condition is significant when SNR is relatively low. On the other hand, the EL of "HM with CR" only has a slight advantage over that of HM due to the limited increase of MED. Figure 7 shows the video quality comparison. Let PSNR = 32 dB as the quality benchmark, our proposed "HM with VR" can save 1.5 dB in received SNR comparing with the regular HM and 1.1 dB comparing with "HM with CR". The saving SNR means the lower transmit power requirement or the larger coverage.

5. Performance on Coverage

5.1. Simulation Method. To illuminate the benefit on coverage more clearly, the power ratio is adjusted to evaluate the supported distance between the BS and the UE at the cell edge with the PSNR benchmark.

Step 1 (Determine the required SNR). Firstly, adjust the power ratio η from 0.6 to 1. For each certain η, we can obtain a group of curves of expected PSNR versus received SNR similar to that in Figure 7. According to the benchmark, we can record the required SNR values for the three schemes, respectively, from the curves.

Step 2 (Compute the distance). Rooting in the path loss formula, the relation between the distance d (i.e., the cell radius, in meter) and average received SNR (in dB) is

$$\mathrm{SNR} = \mathrm{SNR}_t - 10\alpha \log_{10} d, \tag{26}$$

where SNR_t is the transmit SNR; α is the path loss constant and is set as 3.76. We consider a coverage range up to about 300 meters because the distance between two BS is 500 meters. SNR_t is chosen so that the received SNR is 5 dB at the cell edge. Therefore, according the recorded SNR values in Step 1, we can derive the distance for any HM scheme with any power ratio.

5.2. Simulation Results. With the different power allocation to each layer, Figure 8 shows the corresponding coverage distances for the three schemes. Firstly, the coverage is increasing with η because the allocated power to BL is to guarantee the reception of users at the edge. However, if the power of BL is too large, EL stream cannot be decoded correctly so that PSNR is deteriorated. For guaranteeing the PSNR benchmark, the coverage distance will be shortened. Hence, the distance is optimal for these schemes when η is about 0.85. We discount the distance to coverage area. Then, our proposed "HM with VR" can extend the coverage by 10% and 7%, respectively, relative to HM and "HM with CR".

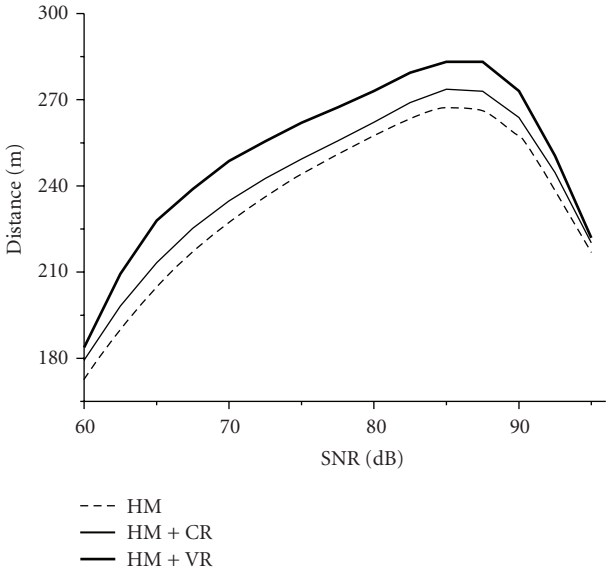


Figure 8: Comparison in coverage.

6. Conclusion

In this paper, we focus on the application of HM for E-MBMS in LTE system. After introducing the regular HM and an enhanced HM with CR, we proposed an HM with VR and its corresponding detection algorithm to improve the performance of less-protected enhancement-layer data stream, in which the idea stems from the concept of signal space diversity in single antenna system. With the suggestion of the dimension of rotated vector $N = 2$, we proved that the EL can earn the full diversity by both SPEP derivation and BER simulation. At the same time, the analysis of detection complexity when $N = 2$ indicates that "HM with VR" will not incur the heavy complexity. We evaluate the proposed scheme in a more integrative metric, not the common error probability, but the video distortion measured by PSNR. The simulation results showed that the proposed "HM with VR" can save 1.5 dB compared with HM, 1.1 dB with "HM with CR" on the required SNR to reach the video quality benchmark. Besides, the simulation results about coverage showed that the extended coverage area ratios are 10% and 7%, respectively.

Appendix

In the appendix, we deduce the expression of $P(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}} \mid x_{\bar{i}}^{\mathrm{el}} = \hat{x}_{\bar{i}}^{\mathrm{el}})$ in (16).

Due to $N = 2$ and $x_{\bar{i}}^{\mathrm{el}} = \hat{x}_{\bar{i}}^{\mathrm{el}}$, the ML detection of x_i^{el} is equivalent to the maximal-ratio combining (MRC) in a single-input multiple-output system. Thus, the post-SNR of x_i^{el} after MRC is

$$\gamma = \sum_{j=1}^{2} \frac{\left|\left[\widetilde{\mathbf{H}}\right]_{i,j}\right|^2 |\delta_i|^2}{\sigma_n^2} = \frac{\beta\left(|h_1|^2 + |h_2|^2\right)|\delta_i|^2}{\sigma_n^2}, \quad (A.1)$$

where $[\widetilde{\mathbf{H}}]_{i,j}$ denotes the element at ith row and jth column. Then, the conditional SER of EL is given by

$$P\left(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}} \mid x_{\bar{i}}^{\mathrm{el}} = \hat{x}_{\bar{i}}^{\mathrm{el}}, \widetilde{\mathbf{H}}\right) = Q\left(\sqrt{\bar{\gamma}}\right) = Q\left(\sqrt{\frac{\beta\left(|h_1|^2 + |h_2|^2\right)|\delta_i|^2}{\sigma_n^2}}\right). \quad (A.2)$$

By the same MGF-based approach, we can yield the analytic expression of $P(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}} \mid x_{\bar{i}}^{\mathrm{el}} = \hat{x}_{\bar{i}}^{\mathrm{el}})$ after taking the expectation of channel element:

$$P\left(x_i^{\mathrm{el}} \neq \hat{x}_i^{\mathrm{el}} \mid x_{\bar{i}}^{\mathrm{el}} = \hat{x}_{\bar{i}}^{\mathrm{el}}\right) = \frac{1}{\pi}\int_0^{\pi/2}\left(\frac{\beta^2|\delta_i|^2}{2\sigma_n^2\sin^2\theta}\right)^{-2} d\theta = \frac{3\sigma_n^4}{4\beta^4|\delta_i|^4}. \quad (A.3)$$

Acknowledgment

This work was supported by National Hi-Tech R&D 863 Program of China under Grant no. 2009AA01Z244.

References

[1] M. Effros, "Universal multiresolution source codes," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2113–2129, 2001.

[2] S. Borade, B. Nakiboğlu, and L. Zheng, "Some fundamental limits of unequal error protection," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '08)*, pp. 2222–2226, July 2008.

[3] G. Li, Z. Zhang, and Z. Lu, "Adaptive hierarchical modulation over correlated MIMO fading channels," in *Proceedings of the 9th International Conference on Signal Processing (ICSP '08)*, pp. 1920–1925, October 2008.

[4] S. A. Ramprashad and C.-E. W. Sundberg, "Hierarchical QAM BICM MIMO systems with iterative decoding and applications to media broadcast," in *Proceedings of the 9th IEEE International Symposium on Wireless, Mobile and Multimedia Networks (WoWMoM '08)*, pp. 1–9, June 2008.

[5] N. Souto, A. Correia, R. Dinis, J. C. Silva, and L. Abreu, "Multiresolution MBMS transmissions for MIMO UTRA LTE systems," in *Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '08)*, pp. 1–6, April 2008.

[6] A. Correia, N. Souto, A. Soares, R. Dinis, and J. Silva, "Multiresolution with hierarchical modulations for long term evolution of UMTS," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 240140, pp. 1–12, 2009.

[7] D. Pradas, A. Bouabdallah, J. Lacan, M. A. Vázquez Castro, and M. Bousquet, "Cross-layer optimization of unequal protected layered video over hierarchical modulation," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '09)*, pp. 1–6, 2009.

[8] S. S. Arslan, P. C. Cosman, and L. B. Milstein, "Progressive source transmissions using joint source-channel coding and hierarchical modulation in packetized networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '09)*, pp. 1–6, 2009.

[9] S. Wang, S. Kwon, and B. K. Yi, "On enhancing hierarchical modulation," in *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '08)*, pp. 1–6, April 2008.

[10] S. M. S. Sadough and P. Duhamel, "On the interaction between channel coding and hierarchical modulation," in *Proceedings of the IEEE International Conference on Communications (ICC '09)*, pp. 1–5, June 2009.

[11] J. Boutros and E. Viterbo, "Signal space diversity: a power-and bandwidth-efficient diversity technique for the rayleigh fading channel," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1453–1467, 1998.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd edition, 2006.

[13] P. K. Vitthaladevuni and M.-S. Alouini, "A recursive algorithm for the exact BER computation of generalized hierarchical QAM constellations," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 297–307, 2003.

[14] M. K. Simon and M. S. Alouini, *Digital Communication over Fading Channels*, John Wiley & Sons, New York, NY, USA, 2005.

[15] J. W. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," in *Proceedings of the IEEE Military Communications Conference (MILCOM '91)*, pp. 2551–2555, November 1991.

[16] K. Higuchi, H. Kawai, N. Maeda, and M. Sawahashi, "Adaptive selection of surviving symbol replica candidates based on maximum reliability in QRM-MLD for OFCDM MIMO multiplexing," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '04)*, pp. 2480–2486, December 2004.

[17] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, 2008.

[18] Ç. Bilen, E. Erkip, and Y. Wang, "Layered video multicast using diversity embedded space time codes," in *Proceedings of the 32nd International Conference on Sarnoff Symposium (SARNOFF '09)*, pp. 1–5, April 2009.

[19] 3GPP TS 36.212 v8.7.0 "E-UTRA Physical channels and modulation", 2009.

[20] V. Vukadinović and J. Huschke, "Statistical multiplexing gains of H.264/AVC video in E-MBMS," in *Proceedings of the 3rd International Symposium on Wireless Pervasive Computing (ISWPC '08)*, pp. 468–474, May 2008.

[21] R4-094198, "LTE MBSFN Channel Model," Ericsson, 3GPP TSG-RAN4 #53 Jeju, Korea, 2009.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2010, Article ID 402831, 10 pages
doi:10.1155/2010/402831

Research Article

Distributed Graph Coloring for Self-Organization in LTE Networks

Furqan Ahmed,[1] Olav Tirkkonen,[1,2] Matti Peltomäki,[3] Juha-Matti Koljonen,[3] Chia-Hao Yu,[1] and Mikko Alava[3]

[1] *Department of Communications and Networking, Aalto University, P.O. Box 13000, 00076 Aalto, Finland*
[2] *Nokia Group, Nokia Research Center, P.O. Box 407, 00045 Helsinki, Finland*
[3] *Department of Applied Physics, Aalto University, P.O. Box 14100, 00076 Aalto, Finland*

Correspondence should be addressed to Olav Tirkkonen, olav.tirkkonen@tkk.fi

Received 1 April 2010; Revised 1 July 2010; Accepted 27 August 2010

Academic Editor: Seppo Hämäläinen

Primary Component Carrier Selection and Physical Cell ID Assignment are two important self-configuration problems pertinent to LTE-Advanced. In this work, we investigate the possibility to solve these problems in a distributive manner using a graph coloring approach. Algorithms based on real-valued interference pricing of conflicts converge rapidly to a local optimum, whereas algorithms with binary interference pricing have a chance to find a global optimum. We apply both local search algorithms and complete algorithms such as Asynchronous Weak-Commitment Search. For system level performance evaluation, a picocellular scenario is considered, with indoor base stations in office houses placed in a Manhattan grid. We investigate a growing network, where neighbor cell lists are generated using practical measurement and reporting models. Distributed selection of conflict-free primary component carriers is shown to converge with 5 or more component carriers, while distributed assignment of confusion-free physical cell IDs is shown to converge with less than 15 IDs. The results reveal that the use of binary pricing of interference with an attempt to find a global optimum outperforms real-valued pricing.

1. Introduction

Self-organization is a wide ranging research and standardization trend in modern networking. In the scope of wireless networking, research on Self-Organized Networks (SONs) ranges from general principles of cognitive and ad hoc network to concrete problems in standardization and implementation of near future mobile networks [1, 2]. Here, we concentrate on a specific SON problem of current interest for the standardization of the next release of the Evolved Universal Terrestrial Radio Access Network (E-UTRAN), a.k.a. Long Term Evolution (LTE). This release, being standardized by the 3rd Generation Partnership Project (3GPP), is known as LTE-Advanced (LTE-A), being standardized by the 3rd Generation Partnership Project (3GPP).

New features of next-generation wireless networks will have an impact on SON, resulting in new use cases and requirements. Of particular interest are Local Area deployments of femto-and picocells. One of the four evaluation scenarios for IMT-Advanced is an indoor scenario [3]. This lends relevance to studying not only automated networking functions, but also autonomous functions. In this paper, we concentrate on two autonomous self-configuration functions, which can be mapped to a graph coloring problem.

The first problem we address is Primary Component Carrier Selection. In 3GPP discussions, carrier aggregation is an essential feature of LTE-A [4]. This leads to the problem of component carrier selection—an individual Base Station (BS) may potentially not operate on all the aggregated carriers, but just on a subset of them. A viable alternative for robust operation is that each BS selects one carrier as a primary carrier, on which the BS has a full set of control channels with full coverage [5, 6]. The Primary Component Carrier Selection (PCCS) problem as such is a direct relative to the well-studied frequency assignment problem [7]. In [5, 6], an autonomous version PCCS was discussed. Another use case for self-configuration, discussed in [1, 8], is automated Physical Cell ID (PCI) assignment. In LTE, the physical cell

ID is needed to distinguish the signal of one BS from the signal of another. Accordingly, neighboring BSs should not have the same PCI. Moreover, to avoid confusion in Hand-Over (HO), a cell should not have two neighbors with the same PCI. Thus the PCI configuration problem becomes a graph coloring problem on the graph of two-hop neighbors. Graph coloring aspects of PCI assignment in LTE have been addressed in [9], where a centralized approach was discussed. A distributed solution based on reserving part of the ID-space to be used for newly switched on cells, was considered in [8]. Apart from [8], to the best of our knowledge, the problem has not been addressed in a distributed manner before.

We investigate simple distributed graph coloring algorithms for both applications addressed. Performance is analyzed for indoor environments based on office houses of the type discussed in [10], placed in a Manhattan grid. A realistic model on UE measurements and reporting is used. Neighbor relations between BSs are determined by handover measurements performed by the User Equipments (UEs). We investigate a dynamical network setting, where the network grows by adding BSs, one by one. When a new BS is added, it self-configures, based on measurements and discussions with the neighbors. We observe that when the measurement and reporting load of the UEs is small, it is beneficial to base distributed decisions on binary conflicts, not real interference couplings. Also, we find the minimum number of component carriers and physical cell IDs that are required for these distributed algorithms to converge.

This paper is organized as follows. In Section 2, we discuss the self-configuration problems addressed. In Section 3, we discuss simple distributed graph coloring algorithms, and their properties. In Section 4, system and network models are presented. Section 5 discusses results for Primary Component Carrier Selection, and Section 6 is on Physical Cell ID Assignment. Finally, conclusions are given in Section 7.

2. Self-Configuration Problems and Coloring

2.1. Autonomous Primary Component Carrier Selection. In carrier aggregation, an operator aggregates a number of component carriers for LTE-A operation. The component carriers may have any of the allowed LTE bandwidths (1.4, 3, 5, 10, 15, or 20 MHz), and they may be either contiguous or noncontiguous [4]. It is natural to take one of the component carriers as a primary one in each cell [5, 6]. This primary carrier serves mobility purposes, and has accordingly a full set of control channels with maximum coverage. The other carriers may be used to boost the data rate, when needed. From the perspective of guaranteeing control channel coverage, the selection of the primary component carrier becomes a classical frequency assignment problem (FAP) [7]. When no other issues than interference is taken into account, and each resource is indistinguishable (i.e., there are no specific channel-related reasons for a BS to favor one carrier more than another), FAP is equivalent to graph coloring. However, when the problem is addressed in an autonomous manner, such as in [6], the local decision makers (BSs) may have more local information at hand for making the decision. Thus the coloring problem may be addressed based on real-valued interference costs, not just on conflicts, as in the classical approaches. For example, in [6], each BS selects the primary component carrier according to real-valued interference information of the neighbors.

Here, we address Primary Component Carrier Selection as a distributed graph coloring problem. According to the discussion in [4], the relevant number of colors for this case is less than ten.

2.2. Autonomous Physical Cell ID Assignment. In LTE, the physical cell ID N_{ID} determines the structure of many channels used in the cell. The ID itself is given by the synchronization channel, and there are 504 different ones. An important use of PCI is to separate cells in handover (HO) measurements. To guarantee proper cell search and handover performance, the PCI assignments should be:

(i) conflict-free: the PCI should be unique in the cell area—no neighbors that the UEs may synchronize to and consider as HO candidates should have the same PCI;

(ii) confusion-free: a cell should not have two neighbors with the same PCI—this guarantees that outward handovers are treated in a proper manner.

Conflict freeness makes the PCI assignment problem a graph coloring problem on the graph of neighbors. Confusion freeness makes it a graph coloring problem on the graph of two-hop neighbors. As any two neighbors of a cell should not have the same PCI, this means that no cell should have a two-hop neighbor (a neighbor of a neighbor) with the same PCI [9]. The space of PCIs may be divided into smaller parts for multiple reasons. For example, parts of the PCI space may be reserved for different layers (macro/micro/femto layers), or part of the space may be restricted for newly switched on, or reconfiguring BSs [8]. In addition, the PCI explicitly determines the structure of the downlink (DL) and uplink (UL) reference signals [11], and this may lead to a much tighter problem for PCI assignment. Related to downlink, there are six different subcarrier groups that downlink reference signals may be mapped to. The subcarrier shift is determined by N_{ID} mod 6. In normal shared channel operation there may not be significant differences related to which subcarriers neighboring BSs have their reference signals on. However, if Collaborative Multipoint transmission (CoMP) with joint beamforming [4] is employed, the reference signal placement becomes an issue. In joint beamforming CoMP, a UE may receive a joint transmission from multiple BSs. In this case, it should be capable of reliable estimation of the channel from these BSs. This requires reference signals to be orthogonal. If orthogonality is achieved in the frequency domain, as in LTE, neighboring BSs should have different subcarrier shifts, and accordingly different N_{ID} mod 6.

The uplink reference signals are grouped into 30 sequence groups, so that sequences with most severe cross-correlations are grouped into the same group. It is desirable

that neighboring cells use different sequence groups, so that it can be guaranteed that reference signals in neighboring cells have good cross-correlation properties. In addition, group hopping is possible. The group used by a cell at any time is determined by $N_{\mathrm{ID}} \bmod 30$, and accordingly, neighboring cells should have differing $N_{\mathrm{ID}} \bmod 30$. Based on this, we observe that PCI assignment in LTE may require conflict freeness with 30 colors, and with CoMP, conflict freeness with 6 colors. For confusion freeness, a larger space of colors is possible. It is worthwhile to investigate, how many PCIs are needed to guarantee confusion-free autonomous PCI assignment.

3. Distributed Graph Coloring Algorithms

Much of the previous work on distributed graph coloring algorithms concentrates on finding colorings with $\Delta + 1$ or $O(\Delta)$ colors, where Δ is the largest number of neighbors of any node. For these cases, rapidly converging distributed algorithms exist, both deterministic and stochastic, and the convergence characteristics can be analyzed in closed form; see [12] and references therein. For more greedy cases, when the number of colors is smaller than Δ, generic constraint satisfaction algorithms may be used. In the survey paper [13], three algorithms proposed by Yokoo and Hirayama are discussed. Distributed Breakout (DBO) is an algorithm based on local reasoning that is capable of breaking out from a local minimum. Asynchronous Backtracking (ABT) and Asynchronous Weak-Commitment Search (AWC) are complete algorithms that are able to satisfy all constraints if possible. In [14], Distributed Stochastic Algorithms (DSAs) are considered, and shown to outperform DBO. DSAs are synchronous algorithms belonging to the wide class of local search algorithms; see [15] for a review.

We want to perform the allocation of the resources so that the BSs execute a routine asynchronously, that is, one at a time, and the order in which they do this is random but fixed. This models an operation where each BS updates its decision at regular intervals, according to a clock which is not synchronized with its neighbors. Also, the time it takes to communicate the change to the neighbors is assumed to be negligible compared to the update interval. Accordingly, we will select a few asynchronous local search algorithms, as well as AWC and ABT, for evaluation.

3.1. Local Search Algorithms. In this paper, four simple distributed local search algorithms will be used for graph coloring. The considered algorithms can be classified according to two characteristics.

The first classification is related to the type of interference pricing. When real-valued interference couplings are used, a real-valued price may be considered between neighbors using the same resource [16, 17]. In contrast, when binary conflicts are considered, the decisions are made based on the number of conflicting neighbors only, not based on the strength of the conflicts.

A second classification is according the number of alternatives tried by a node, when updating which resource to use. The first alternative is a random selection of resource, as in Monte Carlo algorithms; see [18]. The BS randomly selects one of the resources not used by itself at the moment. It starts to use the new resource, if the price (interference price, or number of conflicts), is less than or equal to the price with the earlier resource. In a multiple-try algorithm (e.g., [14]), the node calculates the price for all resources, and randomly selects one of the resources with the lowest price. The binary algorithms have an *Absorbing Local Optimum*, that is, the BS does not change its resource once it is in a conflict-free state [14, 19]. The motivation for this is that in a global optimum of a graph coloring problem, each node sees a local optimum. With this method, unnecessary reconfigurations of the BSs are avoided. In addition, when a node is not in a local optimum, we allow *plateau moves*. This means that if the price of the tested resource, or with multiple-try, the lowest price of all resources, is the same as the price of the resource being used, the node changes the resource. This allows the algorithm to break out from a local optimum and search for a global optimum, as discussed in [14, 15].

The resulting four algorithms are called

(i) Bin: binary pricing, random candidate resource selection,

(ii) Real: real pricing, random candidate resource selection,

(iii) BinMulti: binary pricing, all candidate resources considered, random selection among the best candidates,

(iv) RealMulti: real pricing, all candidate resources considered, random selection among the best candidates.

The real-pricing algorithms with random, and best candidate selection are known from [16, 17], respectively. The binary multiple-try algorithm is an asynchronous version of DSA-D of [14] with $p = 1$. The simplest algorithm, binary pricing with random candidate selection, is to our knowledge not known in the literature as a distributed algorithm. Similar principles are used when solving problems in statistical physics using the so-called zero-temperature Markov Chain Monte Carlo algorithm [18]. The four local search algorithms can be compactly described as follows. Let $G(\mathcal{V}, \mathcal{E})$ be a graph where the BSs are the vertices $v \in \mathcal{V}$, $\mathcal{E}$ is the set of edges, and $w(v, v')$ is the weight of an edge connecting v and v'. For real-valued pricing, $w \in \mathbb{R}$, whereas for binary pricing, $w \in \{0, 1\}$. The set of neighbors of v is $\mathcal{N}_v = \{v' \mid w(v, v') > 0\}$, and the set of colors is $\mathcal{C}$. If v uses color c_v, the local interference price experienced by v is

$$P_v(c_v) = \sum_{v' \in \mathcal{N}_v} \delta(c_v, c_{v'}) w(v, v'), \tag{1}$$

where δ is the Kronecker delta symbol. When v starts executing the routine, its color is c_v. The candidate new color is

$$\tilde{c}_v = \begin{cases} \mathrm{rand}(\mathcal{C} \neg \{c_v\}), & \text{for random selection,} \\ \mathrm{rand} \arg\min_{c \in \mathcal{C}} P_v(c), & \text{for multiple try,} \end{cases} \tag{2}$$

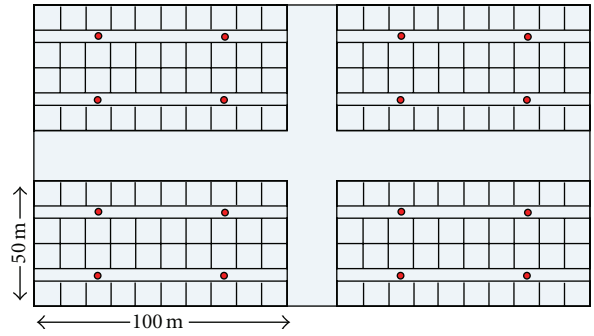


Figure 1: Layout of four-building scenario in the WINNER 2 office Manhattan path loss model. Red dots represent pico-cell BSs.

where rand($\mathcal{X}$) selects a random element from the set $\mathcal{X}$. The color after execution is

$$\begin{cases} \widetilde{c}_v, & \text{if } P_v(\widetilde{c}_v) \leq P_v\,(c_v),\ P_v(c_v) > 0, \\ c_v, & \text{otherwise.} \end{cases} \tag{3}$$

3.2. Complete Constraint Satisfaction Algorithms. For comparison, we consider two complete constraint satisfaction algorithms discussed in [13]. Both utilize global IDs to break the symmetry between agents, and operate in principle on a complete search tree. In Asynchronous Backtracking, priority values of agents follow global IDs and agents communicate their current values to neighbors asynchronously, using messages. Preceding agents in alphabetical order have higher priority. Agents try to find value assignments consistent with higher priority agents. If no such value exists, new constraints are generated and communicated to a higher priority agent, which then attempts to change its assignment. A value once selected is not changed unless lower priority agents force it. This makes a wrong value selection very expensive for large scale problems.

Asynchronous Weak-Commitment Search is based on a message passing principle similar to ABT. AWC improves on ABT by making use of a minimum conflict heuristic and dynamic reordering of agents (i.e., changing priorities) to minimize the number of constraints generated. This often enables recovery from bad value selection without an exhaustive search.

In ABT and AWC, a constraint received from a lower priority agent may involve the state of an agent that is not a neighbor of the receiving agent. To handle this, ABT and AWC require a protocol of establishing new communication links between agents, so that information pertinent to constraints is available at the agents. For more details on these algorithms, see [13].

For larger networks, a preprocessing step of graph partitioning may be used first to divide the graph into a set of smaller loosely connected ones, hence, enabling an efficient framework for obtaining the solution to underlying smaller Constraint Satisfaction Problems in a concurrent way. This approach is investigated in [20].

Table 1: Noise level parameters.

Transmit power P	20 dBm
Noise figure	9 dB
Signal bandwidth	20 MHz
BW efficiency	0.9

4. System Model

For performance analysis of the discussed algorithms for both Primary Component Carrier Selection and automated PCI configuration, in a picocellular network, an office building path loss model in a Manhattan grid has been constructed. The nodes correspond to base stations or cells, and the edges are weighted by real or binary-valued interferences. The path loss model and operation of BSs and UEs is discussed in following subsections.

4.1. Path Loss Model. The main propagation characteristics are according to the Winner path loss models in [10]. We consider both intra- and interbuilding interference by placing a number of multiple-floor buildings in a Manhattan grid; see Figure 1. The buildings model modern office buildings comprising of rooms and corridors. Propagation inside the buildings is modeled according to the Winner A1 model of [10], and propagation between the buildings is modeled as the Manhattan-grid path loss model B1 of [10]. Distance dependent path loss is calculated from the parameters A, B, C as

$$PL = A\log_{10}(d) + B + C\log_{10}\left(\frac{f_c}{5}\right) + X + FL, \tag{4}$$

where d is the distance between the transmitter and receiver, f_c is the carrier frequency, X is the wall and window loss, and FL is the floor loss.

As we are interested in the performance of a large system, we have chosen the buildings to have many floors. Also, we consider wrap-around boundary conditions in all directions (including the floor-dimension). The modeled system thus consists of a Manhattan grid of very tall buildings, and is essentially three-dimensional.

In addition to distance-dependent path loss, we consider shadow fading, according to [10]. It is assumed that the measurements UEs are performing are averaged over the channel coherence time. Accordingly, no fast fading is modeled. Parameters determining the thermal noise level can be found in Table 1. The parameters model a picocellular transmitter utilizing a full LTE bandwidth.

4.2. Model of UE Measurements and Reporting. To have a realistic model of the information that the BSs base their decisions on, we assume that the UE selects the strongest BS to be the serving BS, and reports some of the strongest interferers to the serving BS. If the measurement capabilities of the UEs are taken into account, not all UEs will be able to measure all BSs. We model this with a *synchronization threshold* H_{synch}, which is a threshold in SINR under which

Table 2: Neighbor relation parameters.

Synch threshold	$H_{\text{synch}} = -7$ dB
Reporting load	$L_{\text{rep}} = 1$
UEs per room	1
UEs per corridor	5
Cell selection	best C/I

a UE is not able to synchronize to a BS. This threshold is determined by the synchronization sequences used in the standard, and UE implementation. For LTE, a typical value would be -7 dB. The measurement and reporting load of the UE is taken into account by limiting the number of neighboring BSs that the UE should measure and report to the serving BS to a small number L_{rep}.

4.3. Model of BS Operation. We model the situation where the BSs collect information from their area for a sufficient time, before they start changing the network configuration. We consider the spatial coherence properties of shadow fading to be such that shadow fading is constant within a room. Thus we consider that sufficient statistics of the network is collected when there is one sample per independent shadow fading realization, meaning one UE per room. To consider a constant UE density, we assume 5 UEs per corridor. The parameters determining the neighbor relations are summarized in Table 2.

Once the BS has gathered sufficient information, it decides a cost for the interference caused by another BS to the UEs it serves. For each interferer, the BS has statistics of the interference produced, according to the received reports. In our model, the BS simply considers the worst interference caused by the interferer to any of its served UEs as the interference cost caused by that interferer. The interferences caused to different UEs are measured in terms of the Carrier-to-Interference ratio measured when synchronized to the interferer—the signal power from the interferer divided by the signal power from all other BSs (including the serving one).

4.4. Neighbor Relation and Interference Coupling. The neighbor relation between the cells is determined on a per-drop basis. Each drop represents a network configuration with fixed shadow fading, and a fixed neighbor relation. These relations are determined per drop as follows.

(i) 50 UEs are dropped per floor, evenly distributed in rooms and corridors.

(ii) UEs perform cell selection; serving cell is cell with best C/I.

(iii) Each UE selects primary HO candidate, BS with second best C/I.

(iv) If C/I of primary HO candidate is above the synchronization threshold H_{synch}, synchronization to primary HO candidate is considered successful. Otherwise the UE has no HO candidate.

(v) All BSs that are primary HO candidates of a UE served by the BS, are considered neighbors of the BS.

(vi) The interference coupling between a cell and its neighbor is the highest interference (relative to the carrier power) caused by this neighbor to a UE served by the cell. The interference coupling of non-neighboring cells is 0.

The resulting distribution of the number of neighbors per cell arising from the used neighbor definition can be found in Figure 2(a). For component carrier selection, and PCI conflict freeness, the number of neighbors is essential. For PCI confusion freeness, a BS should have no two-hop neighbors with the same ID, accordingly, the distribution of two-hop neighbors is relevant. This is depicted in Figure 2(b). The distributions are collected over 500 drops. It should be noted that in the simulated scenario, there are 96 BSs, and in one building, there are 24 BSs. The median number of neighbors is 3, and the median number of two-hop neighbors is 9.

4.5. Model of Network Growth. Performance is estimated in a growing network. First a randomly selected subset of 86 of the total 96 BSs in the network scenario are switched on. These are colored with a carrier/PCI according to the problem investigated in a conflict and, for PCI, confusion-free manner. The remaining 10 BSs in the network are then switched on, one-by-one. When a BS is switched on, it first selects a color, and then starts to execute a distributed carrier/PCI selection algorithm.

5. Graph Coloring for Primary Component Carrier Selection

In this section, we compare the performance of network algorithms employing different distributed graph coloring methods for Primary Component Carrier Selection. The idea is to share a small number of resources, which represents a component carrier that a LTE-A system may use, as efficiently as possible. Each cell selects a primary carrier. The output of the algorithm is measured by the distribution of SINR experienced at the nodes, once the primary carrier is distributed. The traffic model is static, and for simplicity it is assumed that there is no secondary usage of the resources. The aim is to analyze the characteristics of the different classes of graph coloring algorithms discussed above, especially related to their local versus global optimization characteristics. The performance metrics evaluated are the following.

(i) Probability of convergence and number of iterations: whether or not the network will be able to find a conflict-free state. In such a state, there are no conflicts above H_{synch} (real-valued or binary).

(ii) Number of cell reboots per added BS: when BSs running a routine of a distributed network algorithm are able to find a conflict-free primary component carrier configuration, the number of times any BS has changed its carrier in the process is measured. This should be a small number, preferably less than one.

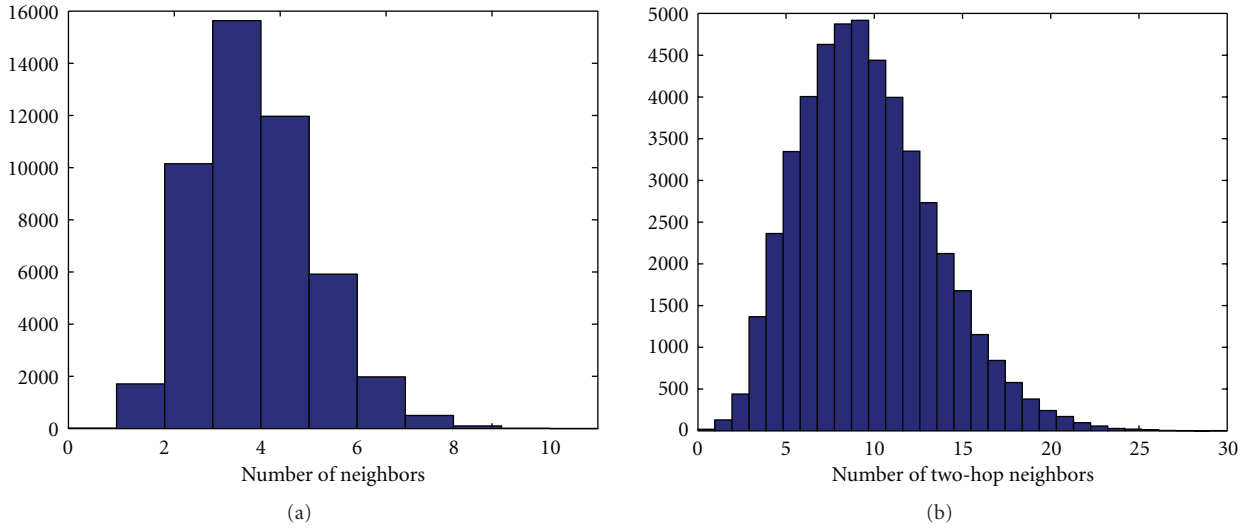


Figure 2: The distribution of the number of neighbors (a) and the number of two-hop neighbors (b).

(iii) The resulting Carrier-to-Interference (C/I) ratios of the selected primary component carriers. This is the most important measure in this case.

The algorithms used are the local search algorithms discussed in Section 3, two alternatives related to pricing, real-valued or binary, and two alternatives related to selecting a new resource to try: random or multiple try.

After switching on, a BS randomly selects a primary component carrier, starts serving UEs and collecting HO measurements. Once this is done, it starts executing a distributed graph coloring algorithm. The algorithm is run until it converges (to a local optimum for real-valued pricing, or to a global optimum for binary pricing), or until 1000 iterations have been done. An iteration is a cycle during which all BSs try to update the resource used once. After that, statistics of the experienced Carrier-to-Interference (C/I) ratios experienced by the users in the system are gathered. Based on these statistics, metrics for comparing system performance can be evaluated.

In Figure 3, the convergence properties are plotted. It can be seen that the binary algorithms converge with 5 component carriers whereas the real-valued algorithms require 7. This shows that the real-valued algorithms have a significant probability to get stuck in local minima in the modeled scenario. In Figure 4(a), the number of cell reboots per added BS is plotted. This plot is in good accordance with the plots in 3. The CDFs obtained using synchronization threshold $H_{\text{synch}} = -7$ dB are shown in Figure 4(b). From the figure it is visible that binary-pricing algorithms perform better than real-valued pricing ones, especially for users in the low C/I region.

In this case, the reporting load of the UEs was low, and the BSs have little information to base their decisions on. Thus, when the interference couplings are only statistically related to the typical user's C/I situation, using binary-valued pricing with plateau moves to attempt global optimization outperforms local optimization based on real-valued interference.

6. Graph Coloring for Automated Physical Cell ID Assignment

6.1. Performance of Distributed Algorithms. Performance is evaluated in the picocellular indoor office environment discussed in Section 4. The target is to find a number of PCIs that is sufficient to allow newly entering cells to configure their PCI in a confusion-and conflict-free manner, with a low level of cell-reboots required. The primary performance metrics are the same as in the previous section, except that instead of conflict freeness, confusion freeness is the target. Also, the C/I distribution is not considered. It is meaningless, as HO confusion is by definition a binary effect. The four distributed local search algorithms discussed in Section 3 are evaluated. In addition, the complete ABT and AWC constraint satisfaction algorithms are used.

For confusion freeness, we need to determine the prices used in the graph of two-hop neighbors. With binary pricing, the only issue is whether or not there is a confusion, that is, the binary and constraint satisfaction algorithms run directly on the two-hop neighbor graph. With real-valued pricing, a real interference distance is calculated for two-hop neighbors as the sum of the dB-scale real-valued interference price of both hops.

Additionally, the confusion couplings and their prices are assumed to be symmetrized. This can be understood as a by-product of the negotiations required to collect confusion information at the decision making node. For symmetrization purposes, or for the operation of the algorithm itself, a signaling channel between the neighboring cells is needed. In LTE systems, the X2 interface provides a natural candidate for this.

After the BS is switched on, it first scans for the synchronization channels of its neighbors, performing so-called

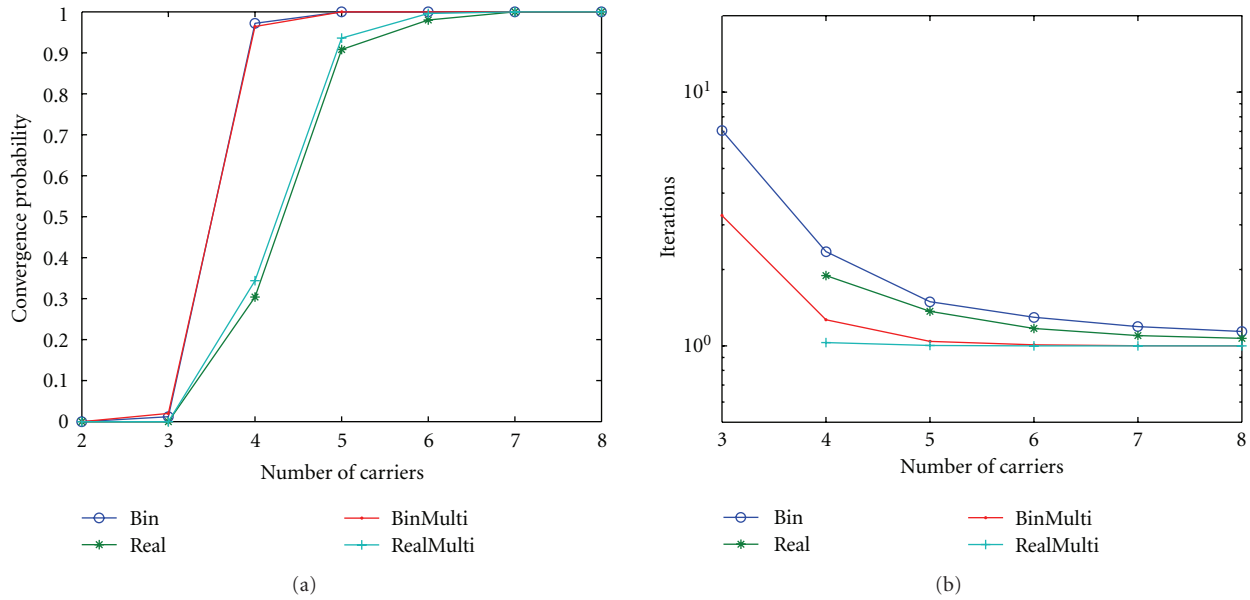


Figure 3: Convergence properties versus number of PCIs for different autonomous Primary Component Carrier Selection algorithms. (a) Convergence probability. (b) Number of iterations for convergence (only converged drops considered).

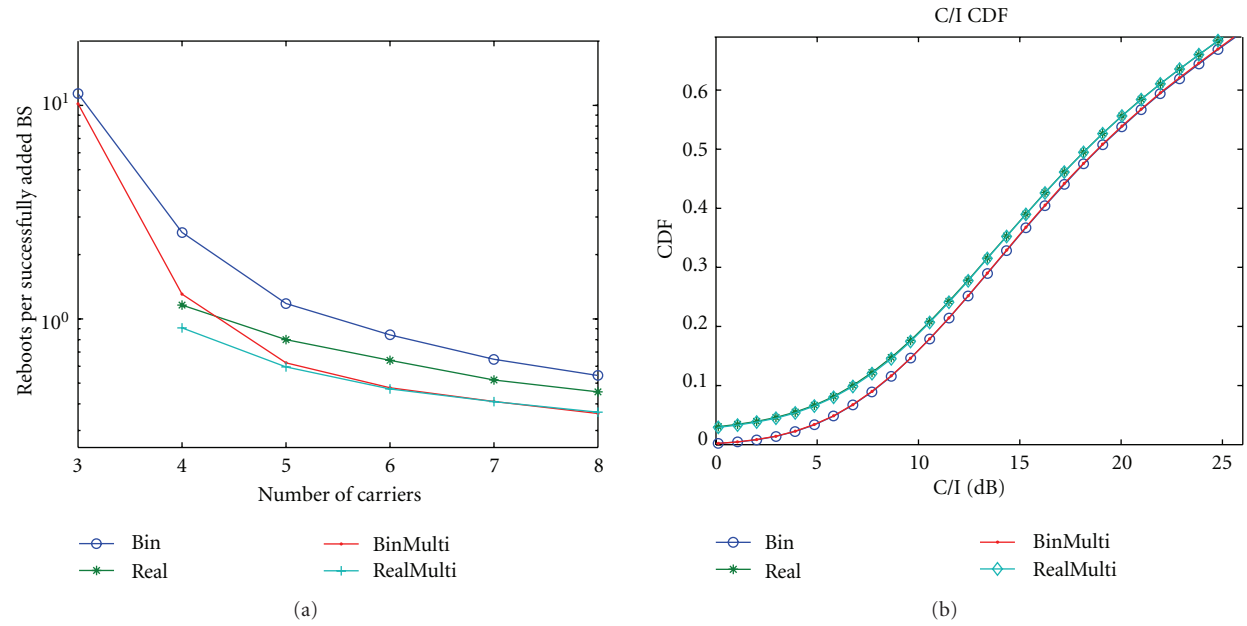


Figure 4: Performance of Primary Component Carrier Selection algorithms. (a) Number of cell reboots versus number of PCIs (only converged drops considered). (b) C/I CDFs for five component carriers.

over-the-air measurements. It finds neighbors with a C/I above H_{synch}. It selects a PCI randomly from the space of PCIs available, avoiding the ones that are used by the neighbors it is able to synchronize to. Note that this does not guarantee conflict freeness with first-hop neighbors, as the set of neighbors a BS can reliably synchronize to is smaller than the set of neighbors. After selecting an initial color, the BS starts serving UEs, collects HO candidate information from them, and performs an Automated Neighbor Relation (ANR) function to establish a connection with the neighboring BSs. If the BS selected a PCI causing a conflict with a neighbor, it is assumed that the neighbors of the conflicting pair are able to identify the conflict by ANR.

With an initial PCI chosen and neighbor relations established, conflict and confusion resolution is performed; a routine of a networking algorithm is executed at each node. If a confusion-free PCI configuration is identified, the results are stored, and a new BS is switched on. The algorithm runs until it converges, or until 1000 iterations have been done.

Related to conflict freeness, conclusions may be drawn from the analysis in the previous section. Thus, for example, a conflict-free assignment of PCI modulo 6, should be possible with the algorithms considered. This may be relevant for DL pilot orthogonalization, as discussed in Section 2. Considering confusion freeness, convergence probabilities for different number of PCIs are shown in Figure 5(a). As the maximum number of two-hop neighbors observed in the network is 26, it can be seen that it is possible to find a confusion-free PCI configuration with a significantly smaller number of PCIs. All algorithms except the random-try real-pricing algorithm converge within 1000 iterations for $N_{\text{PCI}} = 15$. The complete ABT algorithm performs worse than the best local search algorithms. The reason for this is that ABT relies on exhaustive search with a fixed ordering of the agents to find the solution, which is not possible with the limited number of iterations considered here, when the number of colors is small. Recall that the largest number of states in the system is an astronomical N_{PCI}^{96}.

AWC performs much better than ABT, and is clearly the best algorithm when it comes to convergence. AWC avoids exhaustive search by dynamic updating of priorities and minimum conflict heuristics. The binary-pricing algorithms are able to find a converged PCI for $N_{\text{PCI}} = 12$, as opposed to the multiple-try real-price algorithm $N_{\text{PCI}} = 15$, indicating a gain of ∼25%. From this, it is evident that when the target is to achieve confusion freeness, the property of the real-pricing algorithms to get stuck in a local optimum leads to nondesirable results.

In Figure 5(b), the average number of iterations for convergence is reported, where converged drops only are considered. It is notable that ABT and AWC do not fall down as rapidly as the local search algorithms with high N_{PCI}. The reason for this is the use of global IDs to prioritize in conflict situation, which forces the nodes to negotiate for a longer time to identify the node that should solve the conflict.

The number of cell reboots for converged drops are plotted in Figure 6. The multiple-try algorithms require a clearly lower number of reboots than the algorithms selecting a new PCI candidate randomly. This is natural, as the multiple-try algorithms always find a confusion-free PCI if available, and accordingly converge more rapidly. When comparing the number of reboots (for the converged drops) for the real- and binary-pricing algorithms, it can be seen that when the random-try real-pricing algorithm is able to find the global optimum, it does it with less reboots than the random-try binary algorithm. With the multiple-try algorithms, there is no difference in the number of reboots between the binary and real-pricing algorithms in the cases that all drops converge. Comparing to ABT and AWC shows that just as in the case of the number of iterations, these require more reboots than the best local search algorithms.

It should be noted that the fact that the number of reboots does not asymptotically vanish is due to the random initial selection of PCI for the switched on cell. There is always a nonzero probability that this initial selection is conflicting/confused with a neighbor/two-hop neighbor. However, the results in Figure 6 tell us that it is counterproductive to reserve a part of the PCI space for switching-on cells, as suggested in [8]. If that is done, there is always at least one cell reboot per added BS. Here, we see that with a sufficiently large space of PCIs, one get to significantly smaller number of reboots.

As a summary, these results point to a tradeoff when selecting a distributed confusion resolution algorithm. If one is extremely greedy related to the number of PCI's, one has to rely on the complete AWC algorithm, which requires an additional protocol to establish communication between cells that are not two-hop neighbors. If one is moderately greedy, one may do with a multiple-try local search algorithm based on binary confusion pricing.

6.2. Analysis of PCIs Required for Convergence. Here, we try to understand the relationships of the distributions of the number of neighbors and the number of two-hop neighbors in Figure 2, and how they affect the number of PCIs required for confusion freeness. These PDFs are crucial statistics for the behavior of different algorithms, and they are scenario-specific. In the scenario investigated here, for conflict freeness, we need at least 5 PCIs. To realize a confusion-free PCI configuration, we make a trivial observation from the PDF distributions that this can be achieved if we have 27 PCIs. With this number, which is close to the 30 different modulo 30 uplink sequence groups defined in LTE, we can easily configure the network in a distributed manner even in the worst scenario. However, reducing the number of required PCIs (or PCI groups) will allow certain flexibility on system design, as argued in Section 2. A more aggressive and yet straightforward observation is to take the mean value of the distribution as an estimation of the required PCIs. In the PDF of two-hop neighbors, we have the mean value around 10, while 11–15 PCIs are needed for convergence in our numerical results. By definition, two-hop neighbors are neighbors of the neighbors, implying that the distributions in Figure 2 are correlated. A node with more one-hop neighbors is likely to have more two-hop neighbors. We utilize this property for further discussion on the required number of PCIs.

For cells with different numbers of one-hop neighbors, it is more geometrically uniform if these cells are evenly distributed in the considered scenario. This means that for any cluster of cells, their statistics would be close to the statistics of the whole scenario. If this is the case, one can use first-order analysis to estimate the required number of PCIs. With uniform geometry, we can assume that each neighbor will on average introduce $\mathcal{X}$ new two-hop neighbors

$$\mathcal{X} = \overline{N_{1h}} \cdot p, \tag{5}$$

where $\overline{N_{1h}}$ is the mean number of one-hop neighbors and p is the probability of being a new two-hop neighbor. Using first-order moments of the distribution of one- and two-hop neighbors, one can train p by solving

$$\overline{N_{2h}} = \left(\overline{N_{1h}} - 1\right) + \left(\overline{N_{1h}} - 1\right) \cdot \mathcal{X}, \tag{6}$$

where $\overline{N_{2h}}$ is the mean number of two-hop neighbors. In (6), the first $\overline{N_{1h}} - 1$ comes from the fact that your first

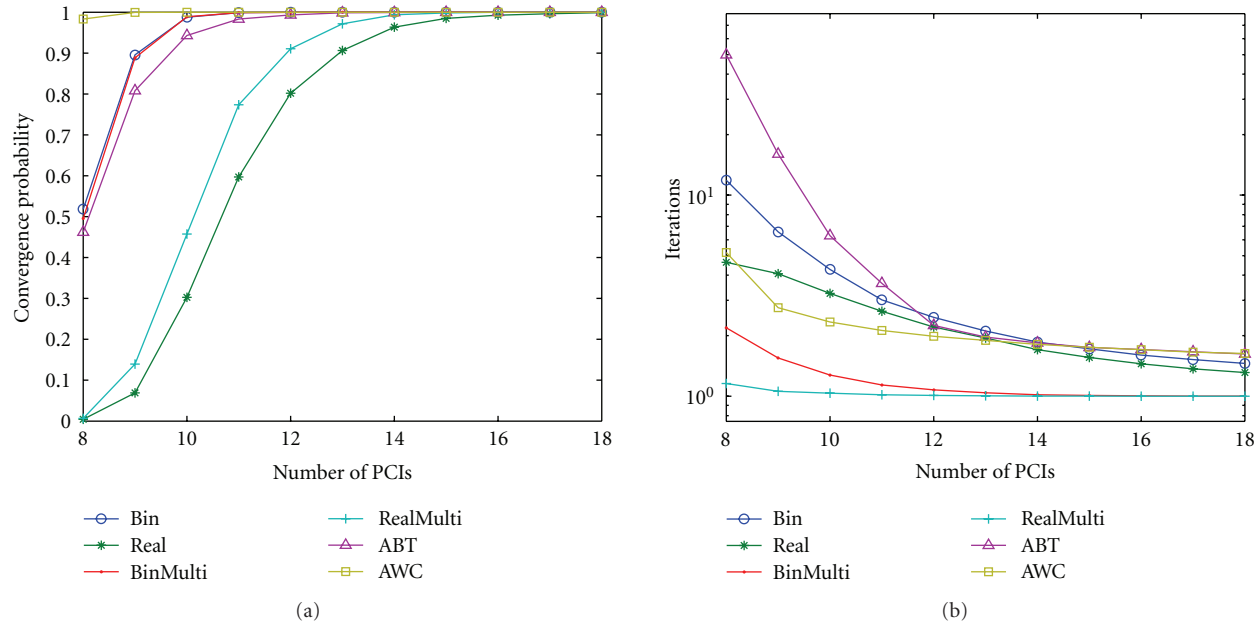


Figure 5: Convergence properties versus number of PCIs for different autonomous PCI configuration algorithms. (a) Convergence probability. (b) Average number of iterations (only converged drops considered).

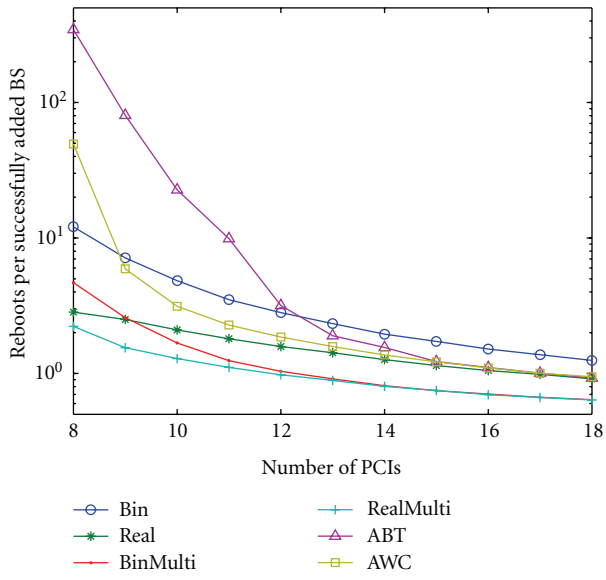


Figure 6: The number of cell reboots versus number of PCIs for different autonomous PCI configuration algorithms (only converged drops considered).

neighbor's neighbors are all new two-hop neighbors except for yourself. The second $\overline{N_{1h}} - 1$ considers the remaining one-hop neighbors.

After training p, the required PCIs can be estimated by calculating the additionally needed PCIs, out of those for avoiding conflict. We thus have the number of required PCIs as

$$N_{\text{PCI}} = N_{\max,1h} + (N_{\max,2h} - N_{\max,1h}) \cdot p, \quad (7)$$

where $N_{\max,1h}$ and $N_{\max,2h}$ are the maximum number of one- and two-hop neighbors, respectively.

On the other hand, if the statistics of clusters of cells is not similar with each other, cells with similar number of neighbors are geometrically closer to each other. In terms of PDF of two-hop neighbors, the higher end represents the statistics of a cluster with high number of one-hop neighbors. Thus, the tail at the higher end is weaker, as is the case in two-hop neighbor distribution of Figure 2. Besides, if we follow the same method as for uniform geometric case, one would see that p is higher than 1, indicating that the mean value of one-hop neighbors is not sufficient to support the mean value of two-hop neighbors. To accommodate this difference, we use similar procedure as uniform geometric case but use $N_{\max,1h}$ instead of $\overline{N_{1h}}$ in training p. This will enhance the fact that the higher end of two-hop PDF corresponds to cell clusters where each cell has higher number of neighbors. This results in $p = 0.3$ and therefore, the required number of PCIs is according to (7) is 12.

With this mechanism, it may be possible to develop a SON algorithm to determine the minimum number of PCIs required for confusion freeness just based on collecting statistics of the number of neighbors from the base stations.

7. Conclusion

We evaluated the use of distributed graph coloring algorithms for two self-configuration problems, pertinent to LTE-A; Primary Component Carrier Selection and Physical Cell ID Assignment.

Simple distributed algorithms based on local search were used for the self-configuration tasks. Real-valued coloring algorithms are based on a real interference price. Distributed

versions of such algorithms converge rapidly to a local optimum, where they are stuck. Discretizing the interference prices to binary conflicts, simple local search algorithms are able to move around on plateaus where the number of conflicts is constant, and then potentially find a globally optimum, conflict-free state.

For evaluating the distributed algorithms, an office Manhattan scenario was used. It was observed that despite the a priori high connectivity of the BSs, Autonomous Primary Component Carrier Selection works with a number of component carrier that is roughly half of the maximum number of neighbors, and Autonomous PCI Assignment works with a number of PCIs that is roughly half of the maximum number of two-hop neighbors. This is much less than the usually considered practical lower limit for distributed graph coloring, which is slightly larger than the order of the maximum number of neighbors [12].

Related to component carrier selection, the results lead to a conclusion that dividing a 100 MHz LTE-A system bandwidth to five component carriers of 20 MHz is a viable strategy, even if autonomous selection should be done.

For Physical Cell ID assignments, the results show that the space of PCIs can be significantly reduced without jeopardizing confusion freeness. For example, if a SON protocol for guaranteeing confusion-free PCI assignment is designed, it may well be designed to operate on a PCI modulo 30 basis, so that the reuse of UL reference signal sequence groups is automatically distributed to far-away cells. The results show that there is no gain from reporting other than binary confusion prices. Moreover, when N_{PCI} is large enough (larger than 14 in this example), it is counterproductive to have a space of temporary PCIs. Just selecting from the space of all PCIs randomly, the expected number of cell reboots is well below 1, which is the minimum when a temporary PCI is used. The price for this is that there is a small possibility that a cell which is not newly switched on has to change its PCI. Local search algorithms were contrasted to complete algorithms, that are distributed realizations of a search tree. It was observed that the best complete algorithm outperforms local search algorithms, when the number of PCIs is very small.

Acknowledgments

The work of F. Ahmed has been supported by the Academy of Finland (grant number 133652). The work of M. Peltomäki, J.-M. Koljonen and C.-H. Yu has been supported by Nokia.

References

[1] F. Lehser, Ed., "Next generation mobile networks— recommendation on SON and O&M requirements," Tech. Rep., NGMN Alliance, December 2008.

[2] M. Döttling and I. Viering, "Challenges in mobile network operation: towards self-optimizing networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 3609–3612, April 2009.

[3] ITU-R, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," Tech. Rep. M 2135, 2008.

[4] 3GPP, "Feasibility study for further advancements of E-UTRA (LTE-Advanced)," Tech. Rep. TR 36.912, 2009.

[5] 3GPP, "Primary component carrier selection, monitoring, and recovery," Tech. Rep. R1-091371, 2009.

[6] L. G. U. Garcia, K. I. Pedersen, and P. E. Mogensen, "Autonomous component carrier selection: interference management in local area environments for LTE-advanced," *IEEE Communications Magazine*, vol. 47, no. 9, pp. 110–116, 2009.

[7] A. Eisenblätter, M. Grötschel, and A. M. Koster, "Frequency planning and ramifications of coloring," *Discussiones Mathematicae Graph Theory*, vol. 22, no. 1, pp. 51–88, 2002.

[8] Nokia Siemens Networks and Nokia, "SON use case: cell Phy ID automated configuration," Tech. Rep. R3-080376, 3GPP, 2008.

[9] T. Bandh, G. Carle, and H. Sanneck, "Graph coloring based physical-cell-ID assignment for LTE networks," in *Proceedings of the ACM International Wireless Communications and Mobile Computing Conference (IWCMC '09)*, pp. 116–120, June 2009.

[10] P. Kyösti et al., "Winner II channel models," Tech. Rep. D1.1.2 V1.2, 2007, http://www.ist-winner.org.

[11] 3GPP, "Evolved universal terrestrial radio access; physical channels and modulation (release 8)," Tech. Rep. TS 36.211 v8.6.0, 2009.

[12] F. Kuhn and R. Wattenhofer, "On the complexity of distributed graph coloring," in *Proceedings of the 25th Annual ACM Symposium on Principles of Distributed Computing (PODC '06)*, pp. 7–15, July 2006.

[13] M. Yokoo and K. Hirayama, "Algorithms for distributed constraint s: a review," *Autonomous Agents and Multi-Agent Systems*, vol. 3, no. 2, pp. 185–207, 2000.

[14] W. Zhang, G. Wang, Z. Xing, and L. Wittenburg, "Distributed stochastic search and distributed breakout: properties, comparison and applications to constraint optimization problems in sensor networks," *Artificial Intelligence*, vol. 161, no. 1-2, pp. 55–87, 2005.

[15] P. Galinier and A. Hertz, "A survey of local search methods for graph coloring," *Computers and Operations Research*, vol. 33, no. 9, pp. 2547–2562, 2006.

[16] J. O. Neel and J. H. Reed, "Performance of distributed dynamic frequency selection schemes for interference reducing networks," in *Proceedings of the IEEE Military Communications Conference (MILCOM '06)*, October 2006.

[17] B. Babadi and V. Tarokh, "A distributed asynchronous algorithm for spectrum sharing in wireless ad hoc networks," in *Proceedings of the 42nd Annual Conference on Information Sciences and Systems (CISS '08)*, pp. 831–835, March 2008.

[18] M. Kardar, *Statistical Physics of Fields*, Cambridge University Press, Cambridge, Mass, USA, 2007.

[19] M. Alava, J. Ardelius, E. Aurell et al., "Circumspect descent prevails in solving random constraint satisfaction problems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 40, pp. 15253–15257, 2008.

[20] M. A. Salido and F. Barber, "Distributed CSPs by graph partitioning," *Applied Mathematics and Computation*, vol. 183, no. 1, pp. 491–498, 2006.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2010, Article ID 894846, 12 pages
doi:10.1155/2010/894846

Research Article

On the Coverage Extension and Capacity Enhancement of Inband Relay Deployments in LTE-Advanced Networks

Abdallah Bou Saleh,[1,2] Simone Redana,[1] Jyri Hämäläinen,[2] and Bernhard Raaf[1]

[1] *NSN Research - Radio Systems, Nokia Siemens Networks, St.-Martin-Strasse 76, 81541 Munich, Germany*
[2] *Department of Communications and Networking, Aalto University School of Science and Technology, P.O. Box 3000, Aalto, FIN-02015, Espoo, Finland*

Correspondence should be addressed to Abdallah Bou Saleh, abdallah.bou_saleh.ext@nsn.com

Received 1 April 2010; Accepted 30 May 2010

Academic Editor: Wenbo Bo Wang

Decode-and-forward relaying is a promising enhancement to existing radio access networks and is currently being standardized in 3GPP to be part of the LTE-Advanced release 10. Two inband operation modes of relay nodes are to be supported, namely Type 1 and Type 1b. Relay nodes promise to offer considerable gain for system capacity or coverage depending on the deployment prioritization. However, the performance of relays, as any other radio access point, significantly depends on the propagation characteristics of the deployment environment. Hence, in this paper, we investigate the performance of Type 1 and Type 1b inband relaying within the LTE-Advanced framework in different propagation scenarios in terms of both coverage extension capabilities and capacity enhancements. A comparison between Type 1 and Type 1b relay nodes is as well presented to study the effect of the relaying overhead on the system performance in inband relay node deployments. System level simulations show that Type 1 and Type 1b inband relay deployments offer low to very high gains depending on the deployment environment. As well, it is shown that the effect of the relaying overhead is minimal on coverage extension whereas it is more evident on system throughput.

1. Introduction

The Universal Mobile Telecommunications System (UTR AN) Long Term Evolution (LTE) is being designed to enhance third generation (3G) radio access technologies (RATs) and it represents a major step towards the International Mobile Telecommunications- (IMTs-) Advanced technologies of the International Telecommunication Union-Radiocommunications Sector (ITU-R). The IMT-Advanced RATs are expected to offer increased broadband capacity with high quality of service (QoS) for next generation multimedia services, such as, high-definition TV (HDTV) content, video chat, mobile TV, and real-time gaming. Requirements for the IMT-Advanced technologies are defined by ITU-R in a circular letter issued in March 2008 calling for candidate RATs [1].

In response to ITU-R circular letter, 3rd Generation Partnership Project (3GPP) made a formal submission in September 2009 proposing that LTE Rel.10 and beyond would be evaluated as a candidate IMT-Advanced technology [2]. The proposed RAT is referred to as LTE-Advanced and is expected to satisfy and overtake IMT-Advanced requirements [3]. In technology evolution, LTE-Advanced defines the framework for further significant advancements to LTE Rel.8 and Rel.9.

While LTE Rel.8 supports peak data rates exceeding 300 Mbps in the downlink (DL) and 75 Mbps in the uplink (UL), LTE-Advanced Rel.10 is expected to offer up to 1 Gbps in the DL and 500 Mbps in the UL in low mobility environments. Extended carrier bandwidths up to 100 MHz will be supported in LTE-Advanced while the maximum bandwidth in Rel.8 is limited to 20 MHz. Furthermore, increased spectral efficiency up to 30 bps/Hz in DL and 15 bps/Hz in UL, along with improved cell edge capacity, decreased user and control plane latencies, and a more homogeneous user experience over the cell area are urged [3].

To address these stringent requirements, different key technologies have been investigated in the 3GPP study item

on LTE-Advanced, of which bandwidth extension through spectrum/carrier aggregation, relay node (RN) deployments, improved multiple-input multiple-output (MIMO) schemes, coordinated multipoint transmission and reception (CoMP), and local area optimization features such as femtocell deployments [4, 5]. In this paper, we will focus on relay nodes as a major technology enhancement.

Decode-and-forward (DF) relay nodes are currently being specified in the 3GPP work item on LTE-Advanced networks to meet the growing demand and challenging requirements for coverage extension and capacity enhancement [5]. RNs are characterized by wireless backhaul and low power consumption that is due to their relatively small size. The connection between RN and the core network is carried out through evolved Node B (eNB), referred to as the donor eNB in 3GPP terminology. The wireless backhaul enables deployment flexibility and eliminates the high costs of a fixed backhaul. Furthermore, RNs do not have strict installation guidelines with respect to radiation, visual disturbance, and planning regulation. Therefore, installing RNs involves lower operational expenditure (OPEX) [6] and faster network upgrade when operators aim to improve the QoS [7]. The cost efficiency of RNs is further investigated in [6, 8, 9]. Due to the compact physical characteristics and low power consumption, RNs can be mounted on structures like lamp posts with power supply facilities. According to previous technical studies, RNs promise to increase the network capacity [10, 11] and to better distribute resources in the cell, or alternatively, extend the cell coverage area [11–14].

In 3GPP standardization, relaying is being considered mainly as a cost-efficient coverage improvement technology. Different RN categories are specified according to the applied spectrum usage approaches on the wireless backhaul link between RN and its donor eNB. Inband relaying utilizes the same spectrum on the relay backhaul as that used to serve the user equipments (UEs) in the RN cells whereas outband RNs use a different spectrum. With the relay (eNB-RN) and access (RN-UE) links being time-division multiplexed in the former approach, operating both links on a single carrier frequency will impose limitations on the resource utilization efficiency, thus, limiting the performance of inband RNs. On the other hand, the latter approach relaxes the resource limitation on the relay link but it may increase the deployment costs since a separate extra spectrum is needed [5]. Considering the scarcity of available spectrums and the high costs of licensed ones, another possible mean to boost the performance of relays is to introduce enough isolation between the antennas used for the access and relay links, for example, provided by means of well-separated and well-isolated antenna structures. This will allow simultaneous operation of both links and, thus, relaxes the relay link.

In 3GPP, RNs are classified according to their relaying strategy. Three types of relays are identified out of which Type 1 and Type 1b inband relays and Type 1a outband RNs are expected to be supported by LTE Rel.10 [5]. A Type 1 RN controls a cell of its own, that is, it has its own physical cell ID and includes functionalities such as radio resource management, scheduling, and hybrid automatic repeat request (HARQ) retransmission control. Type 1a and Type 1b RNs are characterized by the same set of features as Type 1 RN above, except that the former utilizes outband backhaul spectrum usage approach, whereas the latter operates inband but with adequate antenna isolation between access and relay links.

In this paper, we investigate the performance of Type 1 and Type 1b inband RNs in different propagation scenarios within the LTE-Advanced framework. Performance evaluation focuses on both coverage extension and network capacity aspects. In the former case, results are given in terms of an exchange ratio between the relay nodes and macro-cellular eNBs. That is, we examine how many small nodes like relays are needed to replace a conventional eNB, while keeping the same system performance in terms of 10%-ile throughput cumulative distribution function CDF level. We recall the exchange ratio from [14] where it has been explained in more details. Average cell throughput and throughput CDF plots assuming a fixed coverage area are also presented.

The goal of the work is two-folded. On one hand, the relative gain of certain heterogeneous deployments (inband RNs) in comparison with homogeneous (eNB-only) deployments is investigated in different propagation scenarios. The study gives insights for relay deployments of LTE-Advanced networks in different environments. As the performance of a network depends significantly on the propagation conditions, it is essential to validate the deployment of RNs in different environments and give guidelines to the deployment costs and prospective gains. On the other hand, it is important as well to investigate the effect of the relaying overhead required for the inband relay link.

We use system level simulations to investigate the impact of the deployment scenario on the RN performance. Moreover, comparison between Type 1 and Type 1b relays will be carried out and also some analytical considerations are used to characterize the impact of modeling. Simulation results have been produced using a tool that follows up-to-date 3GPP models.

The rest of the paper is organized as follows. In Section 2, the applied evaluation methodology is explained. In Section 3, the propagation scenarios as well as the system model and simulation assumptions are given. In Section 4, performance evaluation and analysis are carried out. The paper is then concluded in Section 5.

2. Evaluation Methodology

The performance of Type 1 and Type 1b relay deployments is evaluated in terms of network coverage and capacity. In the latter case, the average cell throughput is used as a performance measure, while in the former case, results are given in terms of the exchange ratio between RNs and macro-cellular eNBs. This exchange ratio reflects the coverage gains due to RN deployments.

In the following, we first explain the approach used to model Type 1 and Type 1b relaying on a system level

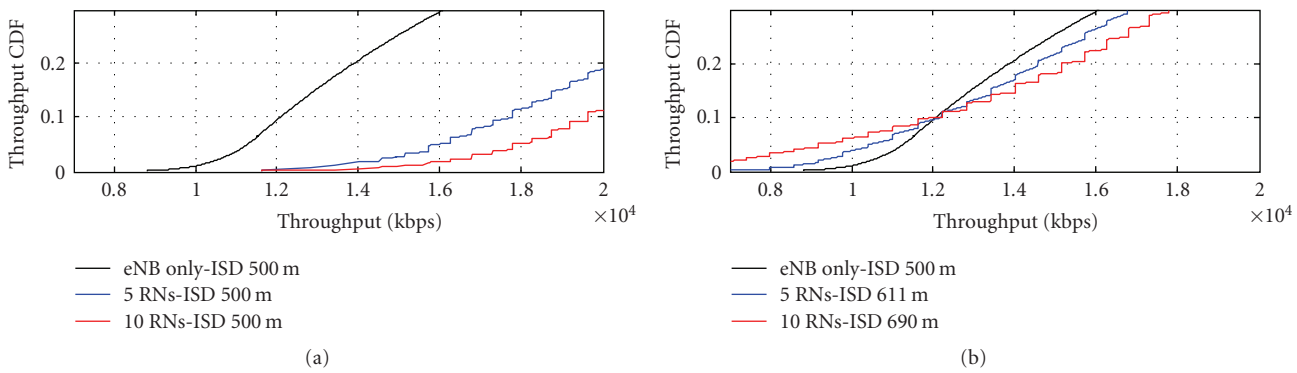


Figure 1: Translating relay throughput gains into coverage extension.

performance evaluation. Then, we describe the methodology that is employed to quantify the coverage results.

2.1. Throughput Modeling for Type 1 and Type 1b Relaying. The end-to-end (e2e) throughput experienced by a UE in a two-hop communication link (eNB-RN-UE) is given as follows:

$$\mathrm{TP}_{\mathrm{e2e}} = \min(\mathrm{TP}_{\mathrm{eNB\text{-}RN}}, \mathrm{TP}_{\mathrm{RN\text{-}UE}}), \tag{1}$$

where e2e throughput is obtained as a minimum over throughputs on the relay and access links.

In inband relaying, the donor eNB utilizes the same time-frequency radio resource pool to serve both the RNs and UEs that are directly connected to it. As well, communications on the access and relay links are time-division multiplexed. Therefore, radio resources available for the relay link will be under high competition in the macrocell eNB. Since RN transmission power is low, its coverage area is relatively small when compared to the macrocell and a UE connected to an RN will usually experience a good access link. Thus, if UE rates are not limited in the RN cells, the relay link may become a bottleneck in two-hop e2e connections.

Throughout the study, we assume a resource allocation strategy which ensures equal data flows on the relay and access links. It is found from (1) that such an allocation is optimal because both relay and access links are fully utilized. Let T_x be the portion of e2e connection resources available on link x, and let R_x be the corresponding rate. Then, we have

$$\begin{aligned} \mathrm{TP}_{\mathrm{e2e}} &= T_{\mathrm{eNB\text{-}RN}} R_{\mathrm{eNB\text{-}RN}} = T_{\mathrm{RN\text{-}UE}} R_{\mathrm{RN\text{-}UE}}, \\ T_{\mathrm{eNB\text{-}RN}} &+ T_{\mathrm{RN\text{-}UE}} = 1, \end{aligned} \tag{2}$$

where the latter equality is used to normalize the total transmission time on the relay and access links. After combining the equations in (2), we obtain the following formula for e2e throughput:

$$\mathrm{TP}_{\mathrm{e2e}} = \left(\frac{1}{\mathrm{TP}_{\mathrm{eNB\text{-}RN}}} + \frac{1}{\mathrm{TP}_{\mathrm{RN\text{-}UE}}} \right)^{-1}. \tag{3}$$

Throughout the performance analysis, (3) will be used to model the e2e throughput of Type 1 inband RNs. We emphasize that (3) takes into account the throughput limitation on the relay link.

If enough isolation between the access and relay links can be obtained, both links can be operated simultaneously, thus easing the limitations on the relay link. In such a case, it is reasonable to assume that the relay link capacity is not a limiting factor and the e2e performance is merely constrained by the access link. We have adopted this assumption for Type 1b inband relaying in order to find bounds for the performance difference with the Type 1 inband relaying approach. Thus, from this on, we assume for Type 1b relaying that

$$\mathrm{TP}_{\mathrm{e2e}} = \mathrm{TP}_{\mathrm{RN\text{-}UE}}. \tag{4}$$

2.2. Tradeoff between Different Deployments. The modeling of the previous section will be used to make throughput comparison between Type 1 and Type 1b inband relaying in the same network topology. While such a comparison is network capacity oriented, it is also of great value to carry out coverage-oriented performance comparison. For that purpose, we use the evaluation methodology of [14]. It aims to quantify the trade-off between numbers of relays and macrocell eNBs on the condition that the coverage requirement is fixed. In this paper, the cell coverage is defined in terms of the 10%-ile throughput CDF level. The 10%-ile level reflects the performance of the worst UEs, which might go easily into outage. In 3GPP, the same approach is used, but performance is measured on the 5%-ile level.

Let us describe the applied comparison methodology in the following. Assume a predefined inter-site-distance ISD_0 between macrocell eNBs and assume that RNs are deployed at the edge of each sector. Then, the additional relays will increase the system throughput with respect to the reference eNB-only deployment. Yet, if the system is scaled by increasing the ISD, then the cell edge throughput can be decreased until the new deployment admit the same 10%-ile throughput as with the reference eNB-only deployment. In the above procedure, the number of deployed RNs per sector can be varied to obtain different ISD and RN density combinations that fulfill the coverage criterion (10%-ile throughput CDF level). The explained approach is illustrated in Figure 1 which presents the throughput CDF for different deployments. The throughput gain from deploying 5 RNs and 10 RNs per sector in the network where the original ISD is 500 m translates into ISD extensions of 111 m and 190 m,

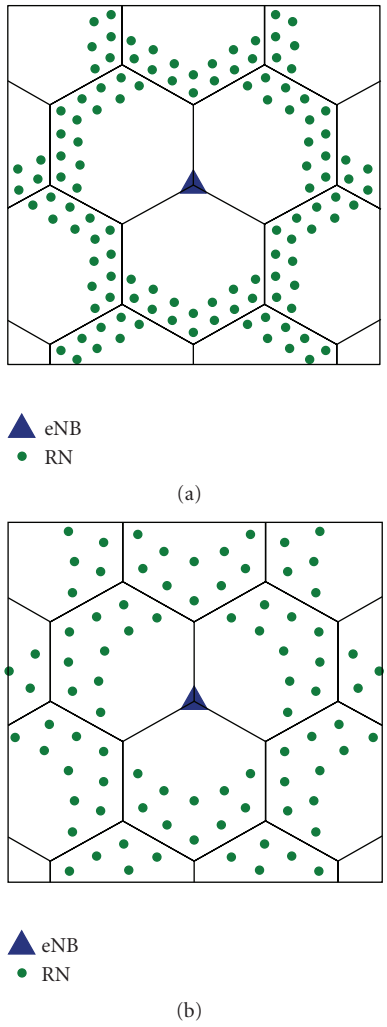


Figure 2: RN deployment in Scenario 1 (a) and Scenarios 2 and 3 (b).

respectively, while preserving the coverage performance at the 10%-ile throughput level.

The different RN density and ISD combinations are referred to as ISO-performance deployments. The eNB-only deployment is referred to by the combination $(0, \mathrm{ISD}_0)$ and various RN deployments are referred to by $(N_{\mathrm{RN}}, \mathrm{ISD}_{\mathrm{ext}})$, where N_{RN} refers to the number of RNs per sector. In the example shown in Figure 1, the combinations (0 RN, 500 m), (5 RNs, 611 m), and (10 RNs, 690 m) are ISO-performance deployments.

The ISO-performance deployments are used to define the trade-off between numbers of RNs and eNBs. This trade-off is expressed in terms of an exchange ratio that will be explained in the following. Assume a reference network that covers area C and consists of N_0 macrocell sites, each composed by three sectors. If N_{RN} relays are added to each sector, starting from the outer edge of the sector (see Figure 2), then site coverage will be increasing and less sites will be needed in order to cover the area C. Let the number of sites in the RN deployment be N_{eNB}. Then, the network contains $3 \cdot N_{\mathrm{RN}} \cdot N_{\mathrm{eNB}}$ relays, and we define the exchange ratio R between RNs and macrocell eNBs by

$$R = \frac{3 \cdot N_{\mathrm{RN}} \cdot N_{\mathrm{eNB}}}{N_0 - N_{\mathrm{eNB}}}. \tag{5}$$

Thus, the exchange ratio is defined by the number of RNs that are needed to reduce the number of macrocell sites from the original number N_0 down to N_{eNB}. This ratio can be used to, for example, estimate the maximum costs of an RN site when the costs of macrocell sites are known. If, for example, $R = 30$, then costs of an RN site can be at most 1/30 of the costs of an eNB site.

Let us consider (5) more carefully. If performance evaluations are done by network level simulations, then it is usually not easy to change the number of sites in applied layout. Fortunately, the number of macrocell sites needed to cover a certain area is proportional to the square of ISD. Thus, there holds

$$R = \frac{3 \cdot N_{\mathrm{RN}} \cdot N_{\mathrm{eNB}}/N_0}{1 - N_{\mathrm{eNB}}/N_0} = \frac{3 \cdot N_{\mathrm{RN}} \cdot (\mathrm{ISD}/\mathrm{ISD}_0)^2}{1 - (\mathrm{ISD}/\mathrm{ISD}_0)^2}, \tag{6}$$

and we can use a fixed number of macrocell sites since the performance gain will be obtained through the extended ISD between macrocell eNBs. We note that the ISD increase is limited by 10%-tile throughput level constraint and multiple iterations of the network simulation might be needed before the proper ISD for a certain deployment is found.

It is worth noting that although the ISO-equivalent deployments perform similarly in terms of coverage, they may result in different exchange ratios according to the relative extension in ISD achieved by the number of RNs deployed. Then, the solution with the lowest cost is found by taking a minimum over the different achieved exchange ratios.

3. Propagation Scenarios and System Model

The performance of a mobile communication system depends largely on the radio environment and hence comparisons between different deployment alternatives and system level enhancements easily become unfair. Therefore, 3GPP has created guidelines for LTE-Advanced system evaluation methodology. The given simulation framework contains propagation and system models as well as recommended values for required parameters [5].

3.1. Propagation Models. Different environments exhibit different propagation characteristics which reflect on eNB and RN coverage areas rendering the network planning a rather challenging task. Small coverage areas may lead to high access node density and considerably high costs for operators. Hence, it is important to validate the RN deployments in different radio environments and give guidelines to the expected deployment costs. Due to increasing rate requirements, it is equally important to investigate the performance of RNs in terms of throughput in different propagation scenarios.

It was early acknowledged in 3GPP LTE-Advanced study item that the propagation modeling is of essential importance when designing and assessing different RN deployments. This fact was reflected in the 3GPP discussion on the distance-dependent path loss model which was open for quite a long time during which the model was changed several times. The first proposed 3GPP model, given in [15], consists of only a non-line-of-sight (NLOS) component and is based on the NLOS ITU-R Urban Micro model [16]. The related scenario, which we will refer to as Scenario 1 (Sc1), assumes that both UEs and RNs always experience NLOS propagation conditions to their donor eNB and, thus, the so-called single slope model of the form

$$\mathrm{PL} = \mathrm{PL}_0 + 10 \cdot n \cdot \log_{10}(R) \tag{7}$$

is applied. In single slope models like Okumura-Hata, the constant term PL_0 contains the impact of factors such as carrier frequency, and eNB and UE antenna heights, while the path loss exponent n does not usually depend on the terminal antenna height. The model in (7) is feasible for densely built areas when a UE is on the street level and the line-of-sight (LOS) probability is small. In Table 1, parameters for different 3GPP path loss models are given [5]. It is seen that the difference between constant terms in the direct (eNB-UE) and relay links is 3.6 dB. This is due to fact that RN antennas are expected to be elevated 5 m from ground level. The RN-UE path loss model shows more aggressive attenuation resulting from low RN antenna height.

The single slope model, however, is pessimistic since it does not take into account the fact that being in LOS conditions is becoming more and more probable when cell sizes are getting smaller. This is especially true when UEs are connected to RNs. Hence, the assumption of considering exclusively an NLOS connection as in [15] might be valid only in very densely populated cities. In the 3GPP evaluation framework, users are assumed to be indoors and the channel model is applied where the path loss towards the building is determined before adding the penetration loss. In many scenarios, there is an LOS connection or at least a clearly dominant direction in the channel between the RN and the building where the UE is located. Therefore, the link suffers from smaller path loss than the channel that assumes propagation over rooftops as in [16].

To address the above-explained propagation characteristics, a probabilistic dual slope model was proposed in [17] for RN-UE link. The model given in (8) is not a conventional dual slope model where a certain breakpoint distance is assumed; it considers the breakpoint through a probability and is based on measurements,

$$\begin{aligned}\mathrm{PL} &= \mathrm{Prob}(\mathrm{LOS}) \cdot \mathrm{PL}(\mathrm{LOS}) + \mathrm{Prob}(\mathrm{NLOS}) \cdot \mathrm{PL}(\mathrm{NLOS}),\\ \mathrm{PL}(\mathrm{LOS}) &= \mathrm{PL}_{\mathrm{LOS}} + 10 \cdot n_{\mathrm{LOS}} \cdot \log_{10}(R),\\ \mathrm{PL}(\mathrm{NLOS}) &= \mathrm{PL}_{\mathrm{NLOS}} + 10 \cdot n_{\mathrm{NLOS}} \cdot \log_{10}(R).\end{aligned} \tag{8}$$

The corresponding model, which will be referred to as Scenario 2 (Sc2) throughout this paper, assumes a mixed LOS/NLOS modeling of the access channel; for parameters, see Table 1. The path loss on the access link is a weighted combination of two, LOS and NLOS, components, where the weighting factor decays as the UE-RN distance increases. The model in [17], however, does not consider environments, where users in a macrocell deployment may experience LOS propagation conditions with their donor eNB.

Finally, the 3GPP propagation Scenario 3 (Sc3), which is based on the model in [5], considers environments with better propagation conditions as compared to both models in Scenario 1 and Scenario 2. This scenario applies probabilistic dual slope model on all three links. It defines an LOS probability function versus the UE-eNB or UE-RN distance, and according to a random probability factor, the UE could be in LOS or NLOS propagation conditions. The model, thus, accounts for the case where UEs are in LOS condition with their eNB or RN, and as well for cases where a UE might be over-the-corner and, hence, might sometimes enjoy an LOS condition and in a nearby place an NLOS condition.

The scenarios reflect three different possible propagation conditions where relays can be deployed. Parameters of Table 1 have been accepted by 3GPP partners for performance evaluation of relay deployments in LTE-Advanced [5] and will be adopted throughout this paper.

3.2. System Model. The simulated network is represented by a regular hexagonal cellular layout with 19 trisectored macrocell sites. The RNs in each sector admit regular outdoor deployment at the sector border. Applied RN deployments will be discussed in Section 4 in more details. Simulation parameters follow the parameter settings agreed in 3GPP [5] and are summarized in Table 2. According to 3GPP performance evaluation guidelines, indoor users are assumed to be distributed with equal probability over the sector area and a 20 dB penetration loss is added on the access and direct links. The full-buffer traffic model is adopted. We note that due to outdoor RN deployment the penetration loss does not occur on the relay link (eNB-RN). Interference in the network is neglected, as the study investigates the performance of RNs in coverage-limited scenarios. Shadowing and fast fading are implicitly accounted for by a 30 dB extra margin on each link.

The link throughput in the system is calculated from signal to interference-plus-noise ratio (SINR) by using a mapping

$$\mathrm{TP} = \mathrm{BW} \cdot B_{\mathrm{eff}} \cdot \log_2\left(1 + \frac{\mathrm{SINR}}{\mathrm{SINR}_{\mathrm{eff}}}\right), \tag{9}$$

where BW is the system operation bandwidth, and B_{eff} and $\mathrm{SINR}_{\mathrm{eff}}$ are the so-called bandwidth and SINR efficiencies [18, 19]. The approximation in (9) is obtained from the well-known Shannon link capacity after scaling by two parameters. This approximation method was introduced in [18], and it has been later used in 3GPP evaluations [19]. Bandwidth and SINR efficiencies depend on the antenna configuration, and we apply values which are given in Table 2. In simulations, we have used a -7 dB limit on SINR so that a UE will be in outage if it experiences SINR levels less

Table 1: Propagation Models.

Distance	Channel Models R [km]
Scenario 1 (Sc1)	
	eNB-UE Link PL = 128.1 + 37.6 $\log_{10}(R)$ *RN-UE Link* PL = 140.7 + 36.7 $\log_{10}(R)$ *eNB–RN Link* PL = 124.5 + 37.6 $\log_{10}(R)$
Scenario 2 (Sc2)	
	eNB-UE Link PL = 128.1 + 37.6 $\log_{10}(R)$ *RN-UE Link* PL = Prob(LOS) PL(LOS) + [1−Prob(LOS)] PL(NLOS) PL(LOS): 103.8 + 20.9 $\log_{10}(R)$, PL(NLOS): 145.4 + 37.5 $\log_{10}(R)$ ISD *500 m-Urban Model* Prob(LOS) = 0.5 – min(0.5, 5 exp(−0.156/R)) + min(0.5, 5 exp(−R/0.03)) ISD *1732 m-Rural Model* Prob(LOS) = 0.5 – min(0.5, 3 exp(−0.3/R)) + min(0.5, 3 exp(−R/0.095)) *eNB-RN Link* PL = 124.5 + 37.6 $\log_{10}(R)$
Scenario 3 (Sc3)	
	eNB-UE Link PL(LOS): 103.4 + 24.2 $\log_{10}(R)$, PL(NLOS): 131.1 + 42.8 $\log_{10}(R)$ ISD *500 m-Urban Model* Prob(LOS) = min(0.018/R, 1)(1− exp(−R/0.063)) + exp(−R/0.063) ISD *1732 m-Suburban Model* Prob (LOS) = exp(−(R − 0.01)/0.2) *RN-UE Link* PL(LOS): 103.8 + 20.9 $\log_{10}(R)$, PL(NLOS): 145.4 + 37.5 $\log_{10}(R)$ ISD *500 m-Urban Model* Prob(LOS) = 0.5 − min(0.5, 5 exp(−0.156/R)) + min(0.5, 5 exp(−R/0.03)) ISD *1732 m-Suburban Model* Prob(LOS) = 0.5 − min(0.5, 3 exp(−0.3/R)) + min(0.5, 3 exp(−R/0.095)) *eNB-RN Link* PL(LOS): 100.7 + 23.5 $\log_{10}(R)$, PL(NLOS): 125.2 + 36.3 $\log_{10}(R)$ ISD *500 m-Urban Model* Prob(LOS) = min(0.018/R, 1)(1− exp(−R / 0.072)) + exp(−R/0.072) ISD *1732 m-Suburban Model* Prob(LOS) = exp(−(R − 0.01)/0.23)

Table 2: Simulation Parameters.

Parameter	Value
System Parameters	
Carrier Frequency	2 GHz
Bandwidth	10 MHz
Highest Modulation Scheme	64-QAM ($R = 9/10$)
Penetration Loss	20 dB on eNB-UE and RN-UE links
Bandwidth Efficiency	0.88
SINR Efficiency	1.25
Thermal Noise PSD	−174 dBm/Hz
SINR lower bound	−7 dB
eNB Parameters	
eNB Transmit Power	46 dBm
eNB Elevation	32 m
eNB Antenna Gain	14 dBi
eNB Antenna Configuration	Tx-2, Rx-2
eNB Noise Figure	5 dB
eNB Antenna Pattern	$A(\theta) = -\min[12\,(\theta/\theta_{3\text{dB}})^2, A_m]$ $\Theta_{3\text{dB}} = 70^\circ$ and $A_m = 25$ dB
UE Parameters	
UE Antenna Configuration	Tx-1, Rx-2
UE Antenna Gain	0 dBi
UE Height	1.5 m
UE Noise Figure	9 dB
Relay Node Parameters	
RN Transmit Power	30 dBm
RN Elevation	5 m
RN Antenna Configuration	Tx-2, Rx-2
RN-eNB Antenna Gain	7 dBi
RN-UE Antenna Gain	5 dBi
RN Antenna Pattern	Omni-directional
RN Noise Figure	5 dB

than −7 dB and then the throughput is set to zero. This limit is introduced due to control channel requirements.

4. Performance Evaluation and Analysis

In this section, the performances of Type 1 and Type 1b RNs are analyzed in terms of coverage extension and network capacity. First, we discuss RN deployments in the different propagation environments, then a bound on the number of RNs that can be deployed is deduced and, thereafter, we present the coverage extension and throughput results.

4.1. Relay Node Deployment. In LTE-Advanced, RNs will be deployed by operators in contrary to, for example, femtocells for which random deployment by end users may take place. Considering, as well, that RNs will be deployed to provide coverage improvements and a more homogeneous user experience over the cell area, we have systematically deployed RNs at the cell edge. Deployment is done in a way that no coverage gaps are left and overlapping between neighboring RN cells is minimized. The unnecessary holes in RN coverage would deteriorate the relative coverage extension gain from RN deployments, while the overlap between RN cells would lead to inefficient RN deployment.

The number of RNs required to cover the cell edge, however, depends significantly on the propagation environment since RNs will exhibit different coverage areas in different propagation conditions. Hence, careful planning of RN deployment is required.

Figure 2(a) presents RN deployments in Scenario 1, whereas RN deployments in Scenario 2 and Scenario 3 are shown in Figure 2(b). The deployments are nearly optimum in terms of the coverage improvement capabilities of the deployed RNs. In both Figures 2(a) and 2(b), 2 tiers of RNs are deployed. A tier of RNs is defined as the sufficient number of RNs required to cover the cell edge without introducing RN coverage gaps. In this study, 7 RNs and 14 RNs, which constitute, respectively, 1 and 2 tiers, are deployed in Scenario 1 environments. The second tier is another set of RNs deployed closer to the eNB, however, still in a way where no RN coverage gaps are occurring between the two tiers. In Scenario 2 and Scenario 3, either 5 RNs or 10 RNs are deployed. The different number of RNs per tier reflects the different coverage areas of RNs in different propagation environments. To keep the comparison fair, the same deployments are used for both Type 1 and Type 1b inband RNs.

Tables 3 and 4 present, respectively, the RN coverage areas for the three considered scenarios in networks with ISD 500 m and ISD 1732 m. We note that these ISD values have been widely used in 3GPP for the urban and suburban test cases. It is found that in case of 500 m ISD, the first tier of RNs covers about 20% of the macrocell sector area in Scenario 1 while in Scenario 2 the coverage is around 40%. Similar behavior is experienced in 1732 m ISD case. We recall that the coverage area is defined by cell selection according to the received signal strength and, hence, the better propagation conditions on the access link in Scenario 2 due to the LOS component render observed larger coverage area. This conclusion is also valid for Scenario 3, but there, the potential LOS component occurs also in the direct link between eNB and UE and, therefore, RN coverage areas are somewhat smaller than in Scenario 2 for ISD 500 m case. In case of networks with ISD 1732 m, however, Scenario 3 models a suburban environment where UEs towards the sector edge experience, with very high probability, NLOS propagation conditions to the eNB which are worse than those in Scenario 2. This leads to a larger coverage of RNs in Scenario 3 in ISD 1732 m case.

If two tiers of relays are employed, then the area covered by RNs is increasing, but we note that coverage cannot be doubled by doubling the number of relays. This is due to fact that RN coverage area will become smaller when it is moved closer to the eNB. We consider this phenomenon in the next subsection.

4.2. Bound on the Number of Deployed Relays. From site coverage extension perspective, the outer edge of the macrocell sector offers the most attractive locations for relay deployments. Therefore, relays are usually deployed in tiers starting from the macrocell edge; see Figure 2, where RN deployments are illustrated. The number of relays in tiers is selected so that relays cover the whole macrocell edge. By adding new tiers between macrocell center and the first tier, the area covered by relays can be increased. Yet, the coverage gain from additional relay tiers will become smaller tier by tier. More specifically, while the relay link budget will define the maximum distance between the eNB and the first tier RNs, eNB interference on the access link may set a practical minimum distance between the eNB and RNs. To describe this in more details, we consider a user between the eNB and the RN in a location where the received powers from both are equal. Then,

$$\frac{P_{\text{eNB}}}{L_{\text{eNB-UE}}} = \frac{P_{\text{RN}}}{L_{\text{RN-UE}}}, \tag{10}$$

where P_{eNB} and P_{RN} refer to the transmission powers of the eNB and the RN, respectively, and $L_{\text{eNB-UE}}$ and $L_{\text{RN-UE}}$ refer to the path losses on the links indicted by the subscript. Let us consider only distance-dependent path losses and antenna gains for simplicity. Then,

$$\begin{aligned} L_{\text{eNB-UE}} &= G_{\text{eNB}} \cdot \alpha_{\text{eNB}} \cdot D_{\text{eNB}}^{\beta_{\text{eNB}}}, \\ L_{\text{RN-UE}} &= G_{\text{RN}} \cdot \alpha_{\text{RN}} \cdot D_{\text{RN}}^{\beta_{\text{RN}}}, \end{aligned} \tag{11}$$

where D_{eNB} and D_{RN} are distances from a UE to the eNB and the RN, respectively, $(\alpha_{\text{eNB}}, \beta_{\text{eNB}})$ and $(\alpha_{\text{RN}}, \beta_{\text{RN}})$ are path loss model parameters, and G_{eNB} and G_{RN} are the eNB and RN antenna gains, respectively. From (11) we solve

$$D_{\text{RN}} = D_{\text{eNB}}^{\beta_{\text{eNB}}/\beta_{\text{RN}}} \cdot \left(\frac{\alpha_{\text{eNB}}}{\alpha_{\text{RN}}} \cdot \frac{G_{\text{RN}}}{G_{\text{eNB}}} \cdot \frac{P_{\text{RN}}}{P_{\text{eNB}}} \right)^{1/\beta_{\text{RN}}}. \tag{12}$$

If we use parameters of Scenario 1 from Tables 1 and 2 and assuming that the RN is in the boresight direction from eNB, then we obtain

$$D_{\text{RN}} \approx 0.095 \cdot D_{\text{eNB}}^{1.02}. \tag{13}$$

From the macrocell geometry, we find that if ISD is 500 m, then the macrocell range is 333 m. If RN is located close to the cell edge, then D_{eNB} is around 300 m since D_{RN} would be then about 28 m. Thus, RN range is quite small even for first tier relays in Scenario 1. Furthermore, if RN is placed close to the midpoint between cell edge and eNB, and assume $D_{\text{eNB}} = 160$ m, for example, then we have $D_{\text{RN}} = 15$ m, and RN coverage becomes truly small. Fortunately, the situation in Scenario 2 and Scenario 3 will be more favorable for relaying, see, for example, Tables 3 and 4.

4.3. Coverage Extension: ISO-Performance. In what follows, the coverage extension capabilities of Type 1 and Type 1b RN deployments are investigated in different scenarios in ISD 500 m and 1732 m networks, respectively.

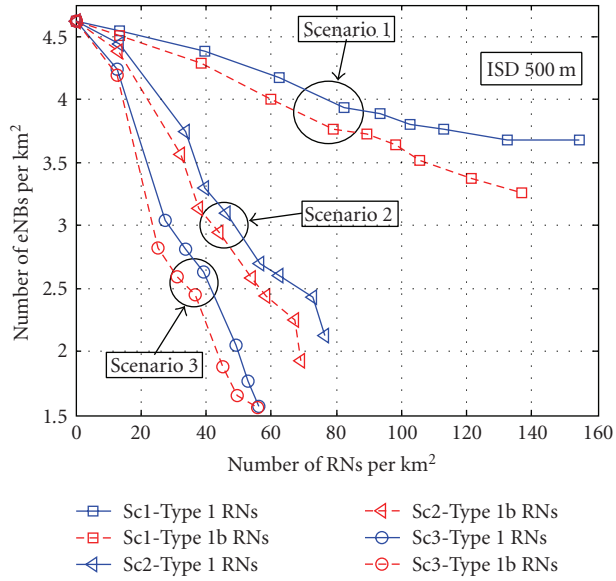


Figure 3: ISO-performance curve - ISD 500 m.

Table 3: RN Deployment Characteristics, ISD 500 m.

RN tiers per sector	Scenario	Number of RNs	Total RN coverage area [%]
1 tier	Sc 1	7	19.5
	Sc 2	5	39.5
	Sc 3	5	35.5
2 tiers	Sc 1	14	33
	Sc 2	10	65
	Sc 3	10	53.5

Table 4: RN Deployment Characteristics, ISD 1732 m.

RN tiers per sector	Scenario	Number of RNs	Total RN coverage area [%]
1 tier	Sc 1	7	21
	Sc 2	5	39
	Sc 3	5	48
2 tiers	Sc 1	14	36.5
	Sc 2	10	61.5
	Sc 3	10	69.5

Figure 3 presents the ISO-performance of Type 1 and Type 1b RNs for different scenarios in urban environments. Two conclusions are clearly deductible. On one hand, the ISO-performance differences between Type 1 and Type 1b RNs are small in all three scenarios. In terms of coverage extension, the resources consumed by the relay link have limited effect on the system performance. On the second hand, there is a significant difference between the ISO-performances of RN deployments when using different path loss models. The required numbers of RNs in ISO

combinations of Scenario 3 are less than those in Scenario 2, which in turn is much more favorable for relaying than Scenario 1. When comparing Scenario 1 and Scenario 2, the better performance on the access link of Scenario 2 due to LOS probability clearly reflects on the performance of all UEs connected to RNs. We emphasize that the direct link path loss between eNB and UE as well as the path loss between eNB and RN are the same for Scenario 1 and Scenario 2. Thus, for both Type 1 and Type 1b relays the difference between Scenario 1 and 2 ISO curves of Figure 3 is a direct consequence of the difference in the access link attenuation between RN and UE.

While comparing Scenario 2 and Scenario 3, it should be noted that the access link between RN and UE admits similar model in these scenarios while direct and relay link models are different. The better performance of Scenario 3 is mainly due to improved direct link. However, the LOS probability model in Scenario 3 significantly improves the performance of UEs close to the eNB, that is, those experiencing relatively good throughput levels anyway, and results in moderate performance improvement for the UEs at the edges between eNB and RN coverage areas. These are the UEs that contribute the most to the 10%-ile throughput CDF level. On the contrary, as in Scenario 2, the access link model of Scenario 3 ensures significantly better UE throughput for users connected to RNs. Also, Type 1 inband relays can utilize the increased access link performance well in Scenario 3 since LOS probability is introduced to the relay link. Added to that, simulations show that UEs contributing to percentile levels higher than the 10%-ile in throughput CDF are much better in Scenario 3 than those at the 10%-ile level when compared to the relative performance of both groups (UEs below 10%-ile level and those above 10%-ile level) in Scenario 2. Knowing that these UEs will eventually contribute to the 10%-ile as the ISD increases, a higher margin of ISD extension can be supported in Scenario 3. These reasons lead to better coverage extension capabilities for RNs in Scenario 3 as compared to Scenario 2.

Table 5 presents the exchange ratio, calculated using (6), between the RNs and the macrocell eNB in the three considered scenarios. The values in Table 5 are selected based on the possible ISO-performance combinations so that the exchange ratio is minimized, that is, the selected (N_{RN}, ISD) combination is the one which achieves the largest ISD extension relative to the number of RNs and eNBs deployed per unit area. Simulations show that in Scenario 1, deploying 7 Type 1 RNs achieves an extension of 42 m, whereas deploying 7 Type 1b RNs results in a 54 m ISD extension. In Scenario 2, the best exchange ratio is achieved by deploying 4 or 7 Type 1 RNs, which achieves ISD extensions of 92 m and 154 m, respectively. For Type 1b RNs, 4-RN deployment is the most cost-efficient deployment and it results in an ISD extension of 107 m. For Scenario 3, both Type 1 and Type 1b RNs present the most cost-efficient ISO combination with 3 RNs per sector resulting in ISD extensions of 106 m and 140 m, respectively.

The exchange ratios presented in Table 5 reflect the performances shown by the ISO plots of Figure 3 and values can be used to compare Type 1 and Type 1b RNs in terms of costs. In Scenario 1, the deployment of Type 1 RNs is appealing if Type 1 RN cost is lower than 93/120 times that of Type 1b RN. The cost should be lower than 26/30 and 15/18 times that of Type 1b RNs in Scenario 2 and Scenario 3, respectively. Note that the cost stands for the Total Cost of Ownership (TCO), and in such a case, the extra cost incurred to insure enough antenna isolation on the access and relay links in Type 1b RN deployments should be taken into account.

Table 5: Coverage extension evaluation, ISD 500 m.

Considered Scenario	Best exchange Ratio	
	Type 1 RN	Type 1b RN
Scenario 1	120	93
Scenario 2	30	26
Scenario 3	18	15

Investigating the RN performance in the different scenarios, inband Type 1 RNs are appealing, costwise, if the RN TCO is less than 1/120 times that of a macro eNB. In Scenario 2 and Scenario 3, the cost limitation diminishes significantly down to 1/30 and 1/18 times that of an eNB. Similar behavior is noticed for Type 1b RNs, where the cost-appealing ratio falls from 1/86 in Scenario 1 to 1/26 and 1/15 times the cost of an eNB in Scenario 2 and Scenario 3, respectively.

4.3.2. Rural/Suburban Models (ISD 1732 m). Figure 4 shows the ISO-performance plots for Type 1 and Type 1b RNs in different propagation conditions and ISD 1732 m network. Type 1 and Type 1b RNs provide the same performance in terms of coverage extension, as they have overlapping ISO curves. This is an important indicator of the fact that the inband relay link for Type 1 relays does not incur noticeable costs on resource utilization provided that RNs are used to extend network coverage on rural/suburban areas. The exchange ratios between RNs and macro eNBs are presented in Table 6. Performance evaluation results of Figure 4 and Table 6 show a one-to-one exchange ratio between Type 1 and Type 1b RNs.

An important result from the coverage extension study in both ISD 500 m and 1732 m cases is that the required number of Type 1b relays is not much smaller than that of Type 1 relays. This is mainly due the fact that RNs are located outdoors, and even in Scenario 1 they admit a very good relay link towards the serving eNB. If network dimensioning is done assuming indoor users, then the radio resource usage by the backhaul of an outdoor Type 1 RN is relatively small. Type 1 relay deployments, as compared to Type 1b relay deployments, are hence well justified since they incur less costs than Type 1b relays.

While comparing the system performances in different scenarios, similar behavior is visible as in the ISD 500 m network: The performance differs drastically between the scenarios. Exchange ratios vary from 86 RNs per eNB in Scenario 1 to 24 RNs per eNB in Scenario 3. The exchange ratio for Scenario 1 is achieved by deploying 7 RNs, Type 1 or Type 1b, per sector, which results in a 201 m ISD extension.

Table 6: Coverage extension evaluation, ISD 1732 m.

Considered Scenario	Best Exchange Ratio	
	Type 1 RN	Type 1b RN
Scenario 1	86	86
Scenario 2	28	28
Scenario 3	24	24

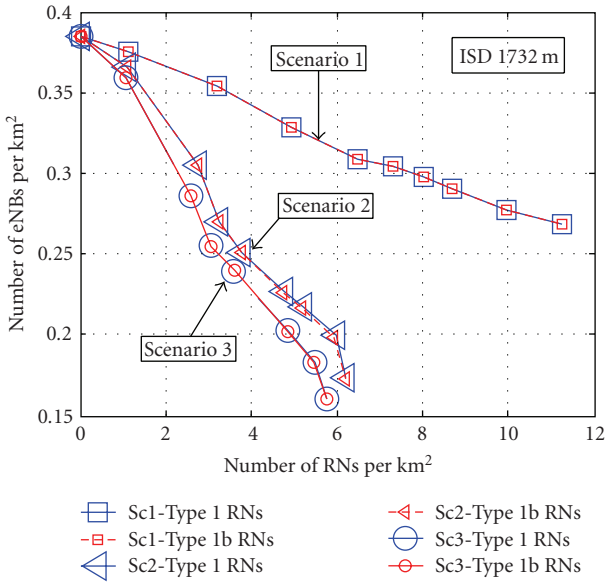


Figure 4: ISO-performance curve - ISD 1732 m.

In Scenario 2, RN deployments with 5 RNs per sector achieve the best cost-efficient ISO-performance deployment and give an ISD extension of 414 m. The exchange ratio in Scenario 3 is obtained by deploying 4 RNs which achieve an ISD extension of 396 m. These performance differences among scenarios are due to applied path loss models and, therefore, the discussion and conclusions are the same as in case of ISD 500 m network.

Finally, we note that the exchange ratios were presented in connection with the corresponding RN deployments. If further coverage is required, a tradeoff between ISD extension and higher RN costs relative to the eNB may be considered.

4.4. Cell Throughput. In the following, the average cell throughput for the different scenarios will be presented for networks with ISD 500 m and 1732 m, respectively. The performance of 1-Tier and 2-Tier Type 1 and Type 1b relay deployments will be investigated and compared.

4.4.1. Urban Models (ISD 500 m). Figure 5 presents the CDF of the cell throughput for Type 1 and Type 1b relays assuming Scenario 1. It is clear that for both 7-RN and 14-RN deployments, Type 1b RNs outperform Type 1 RNs, specifically at the high throughput levels, where Type 1 RNs do not offer any throughput improvements over the macro eNB-only deployment. This is due to the relay link, which

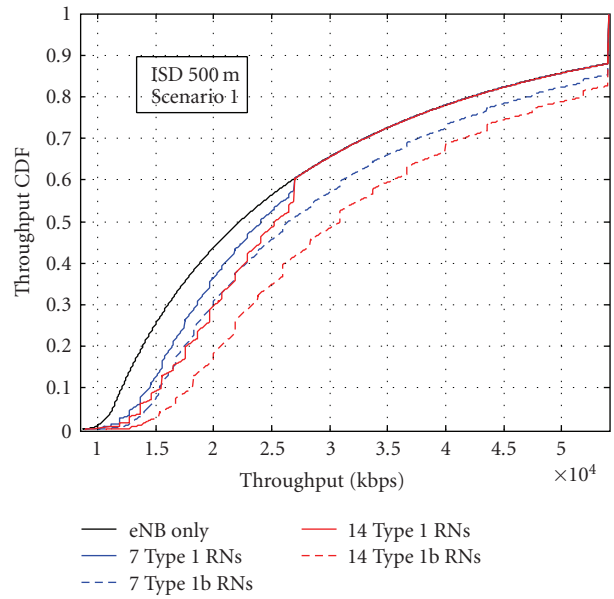


Figure 5: Throughput CDF-Scenario 1, ISD 500 m.

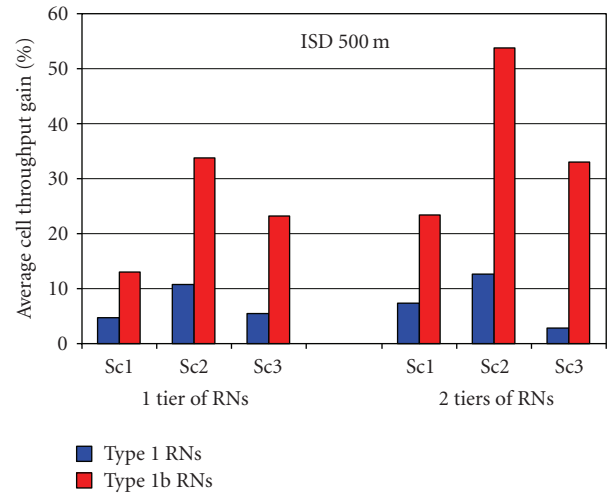


Figure 6: Average cell throughput gains in RN deployments, ISD 500 m.

limits the e2e throughput over the two-hop link at high SINR levels to half of that that can be achieved on the direct link. This can be seen also from (2) and (3), where the e2e throughput of a UE served by a Type 1 RN is formulated. Thus, only direct link users close to eNB contribute to the throughput CDF on very high throughput levels. Therefore, the peak rate limitation in RN cells is best observed from the step in the CDF plots of Figure 5 at 27 Mbps, where the performance of Type 1 RN deployments converges to that of the eNB-only.

According to simulations, a 7-RN deployment with Type 1 relays provides 20% improvement over the eNB-only deployment at the 10%-ile throughput CDF level, whereas Type 1b relays provide 27% gain. The observed gains at the 50%-ile level are 8% and 18% from Type 1 and Type

1b RN deployments, respectively. The clear trend is that gains from Type 1b relays against Type 1 relays become larger with increased throughput CDF levels. This is due to the fact that the relay link creates a bottleneck limiting the e2e throughput when access link radio conditions become better. The conclusion is that UEs connected to Type 1 RNs can in general experience better throughput than in eNB-only deployment, but only Type 1b relaying clearly increases the number of users that admit extremely high throughput. The latter conclusion is backed up by Figure 6 that presents the achieved average cell throughput gains with respect to the eNB-only reference case. We note that 2-tier deployment can be used to increase the cell rate when employing Type 1b relays, but in case of Type 1 relays the gain from the second tier is small. It is worth noting that in the coverage investigations the large throughput performance difference between Type 1 and Type 1b was not well visible in the coverage extension capabilities of relays because the difference at the 10%-ile throughput level is small.

Since results of Figures 5 and 6 assume Scenario 1, it is worth to briefly discuss about the throughput in other scenarios. After comparing the performance of RNs in all three scenarios, we notice that relaying benefits can significantly differ, as was the case in the coverage extension study. RN deployments in Scenario 2 show better performance than in Scenario 1; introducing LOS component to the path loss model of the access link improves the performance of relay deployments whereas the eNB-only performance does not change. When comparing Scenario 2 and Scenario 3, it was found that RN deployments in the former scenario achieve higher relative gain in average cell throughput compared to the eNB-only deployment. This is due to the considerably high throughput levels of eNB-only deployments in Scenario 3 where the UEs close to the eNB experience an LOS connection and hence achieve much higher throughputs as compared to those in Scenario 2. Such UEs contribute significantly to the average cell throughput as compared to those on the cell edge. With the RN cells providing slightly better capacity improvements in Scenario 3 when compared to those in Scenario 2, the relative capacity improvement by RNs in the former is then smaller.

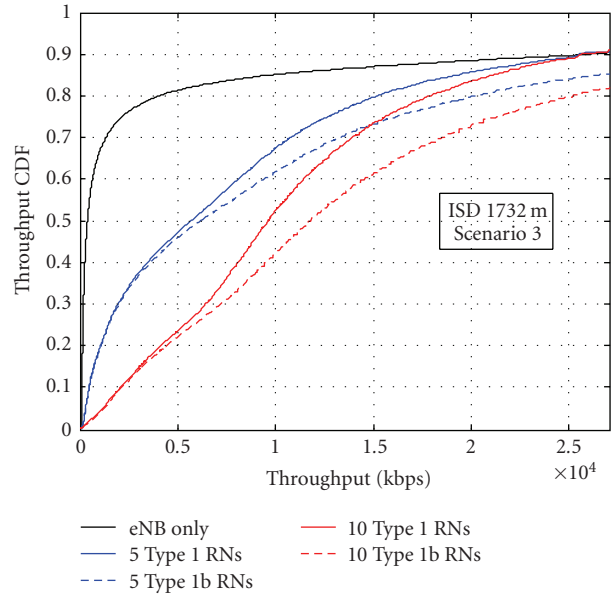


Figure 7: Throughput CDF-Scenario 3, ISD 1732 m.

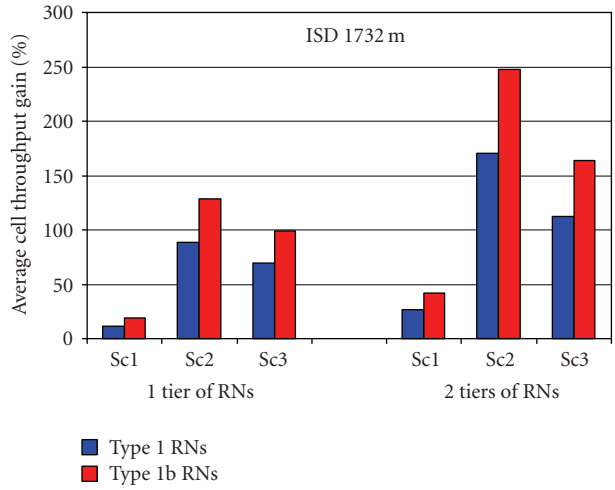


Figure 8: Average cell throughput gains in RN deployments, ISD 1732 m.

4.4.2. Rural/Suburban Models (ISD 1732 m). Figure 7 presents the throughput CDF plots of 1-tier and 2-tier RN deployments assuming Scenario 3 propagation model. It is clearly observed that Type 1 and Type 1b RNs show the same performance at the low percentile throughput CDF levels.

This confirms the previous result from the coverage extension study; resources consumed for the inband relay link in Type 1 RN deployments do not impact the performance of cell edge UEs.

As in networks where ISD = 500 m, Type 1b RNs outperform Type 1 RNs at mid to high throughput levels where the relay link becomes a limiting factor as it requires large portion of the resources. It is, however, worth noting that the relative gain from Type 1b RNs as compared to Type 1 RNs is small.

This is observed in Figure 8, where the gains in average cell throughput are presented. It is clearly observed that gains from both 1-Tier and 2-Tier RN deployments are similar in case of Type 1 and Type 1b RNs. However, it is worth noting the significant difference in RN performance in the three scenarios. RN deployments perform remarkably better in Scenario 2 and Scenario 3 as compared to Scenario 1. This is due to the severe propagation loss on the access link in Scenario 1 where all UEs are assumed to be in NLOS conditions with the donor RN. Comparing Scenario 2 and Scenario 3, similar conclusions and justifications still hold as in networks with ISD 500 m.

Note that significant capacity improvements are achieved by RN deployments in all scenarios as compared to eNB-only deployments.

5. Conclusion

We compared the performance and feasibility of Type 1 and Type 1b Relay Node (RN) deployments within the LTE-Advanced framework. As performance measures, we used the cell coverage, that is, the exchange ratio between numbers of relays and evolved Node Bs (eNBs), and cell throughput. The evaluation was carried out for three different 3GPP propagation models in networks with inter-site-distances (ISDs) 500 m and 1732 m.

First, it was analytically shown that the number of relays within a macrocell admits a practical upper bound since a relay coverage area decreases when it is moved closer to the eNB. Yet, it was noticed that the RN coverage depends largely on the applied path loss scenario. This has been also acknowledged by 3GPP where three different path loss scenarios have been introduced for LTE-Advanced performance evaluations. It was shown through system level simulations that relay deployments exhibit significantly different performance in the different scenarios in terms of both coverage and capacity-oriented evaluation criteria. RNs provided significant gains in scenarios where users exhibit good propagation conditions to the RN, whereas modest gains were achieved in other scenarios.

As the relaying overhead in case of Type 1 relays was expected to be the limiting factor, we investigated the effect of inband relay backhaul link in the three 3GPP-originated scenarios. Results show that the difference in coverage performance between Type 1 and Type 1b relay deployments is small making the Type 1 RN deployment attractive, especially in the case where network dimensioning assumes indoor users and outdoor relays. Actually, in suburban networks with ISD 1732 m, Type 1 and Type 1b RN deployments provide equal coverage performance. If ISD 500 m urban Scenario 1 is assumed, then Type 1b relays provide slightly better coverage performance than Type 1 relays. Yet, it should be noted that 3GPP path loss Scenario 1 admits extremely aggressive signal attenuation between RN and UE making the use of small outdoor cells less favorable in general.

The investigations have shown that the capacity performance of Type 1 and Type 1b inband relays differs significantly. Especially in ISD 500 m network, Type 1 relays clearly lag behind Type 1b relays in throughput performance. The throughput difference is largest in high-throughput regimes.

References

[1] Circular Letter 5/LCCE/2, "Invitations for Submission of Proposals for Candidate Radio Interface Technologies for the Terrestrial Components of the Radio Interface(s) for IMT-Advanced and Invitation to Participating in Their Subsequent Evaluation," March 2008, http://www.itu.int/.

[2] 3GPP RP-090939, "3GPP Submission Package for IMT-Advanced," http://www.3gpp.org/.

[3] TR36.913, "Requirements for Further Advancements for Evolved Universal Terrestrial Radio Access (E-UTRA)," v9.0.0, December 2009, http://www.3gpp.org/.

[4] P. E. Mogensen, T. Koivisto, K. I. Pedersen et al., "LTE-advanced: the path towards gigabit/s in wireless mobile communications," in *Proceedings of the 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace and Electronic Systems Technology (VITAE '09)*, pp. 147–151, Aalborg, Denmark, May 2009.

[5] TR 36.814, "Further Advancements for E-UTRA: Physical Layer Aspects (Release 9)," v2.0.0, March 2010, http://www.3gpp.org/.

[6] E. Lang, S. Redana, and B. Raaf, "Business impact of relay deployment for coverage extension in 3GPP LTE-advanced," in *Proceedings of IEEE International Conference on Communications Workshops (ICC '09)*, Dresden, Germany, June 2009.

[7] K. Doppler, S. Redana, M. Wódczak, P. Rost, and R. Wichman, "Dynamic resource assignment and cooperative relaying in cellular networks: concept and performance assessment," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 475281, 14 pages, 2009.

[8] P. Moberg, P. Skillermark, N. Johansson, and A. Furuskär, "Performance and cost evaluation of fixed relay nodes in future wide area cellular networks," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, pp. 1–5, Athens, Greece, September 2007.

[9] B. Timuş, "Cost analysis issues in a wireless multihop architecture with fixed relays," in *Proceedings of the 61st IEEE Vehicular Technology Conference (VTC '05)*, pp. 3178–3182, Stockholm, Sweden, May-June 2005.

[10] A. So and B. Liang, "Effect of relaying on capacity improvement in wireless local area networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '05)*, pp. 1539–1544, New Orleans, La, USA, March 2005.

[11] A. Bou Saleh, S. Redana, B. Raaf, and J. Hämäläinen, "Comparison of relay and pico eNB deployments in LTE-advanced," in *Proceedings of the 70th IEEE Vehicular Technology Conference (VTC '09)*, Anchorage, Alaska, USA, September 2009.

[12] R. Schoenen, W. Zirwas, and B. H. Walke, "Capacity and coverage analysis of a 3GPP-LTE multihop deployment scenario," in *Proceedings of IEEE International Conference on Communications Workshops*, pp. 31–36, Beijing, China, May 2008.

[13] R. Schoenen, R. Halfmann, and B. H. Walke, "An FDD multihop cellular network for 3 GPP-LTE," in *Proceedings of the 67th IEEE Vehicular Technology Conference (VTC '08)*, pp. 1990–1994, Marina Bay, Singapore, May 2008.

[14] T. Beniero, S. Redana, J. Hämäläinen, and B. Raaf, "Effect of relaying on coverage in 3GPP LTE-advanced," in *Proceedings of the 69th IEEE Vehicular Technology Conference (VTC '09)*, Barcelona, Spain, April 2009.

[15] TR 36.814, "Further Advancements for E-UTRA: Physical Layer Aspects (Release 9)," v1.0.0, Januaray 2009, http://www.3gpp.org/.

[16] "Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced," Document 5D/TEMP/99 (Rev.1), http://www.itu.int/.

[17] TR 36.814, "Further Advancements for E-UTRA: Physical Layer Aspects (Release 9)," v1.0.1, March 2009, http://www.3gpp.org/.

[18] P. Mogensen, W. Na, I. Z. Kováes et al., "LTE capacity compared to the shannon bound," in *Proceedings of the 65th IEEE Vehicular Technology Conference (VTC '07)*, pp. 1234–1238, Dublin, Ireland, April 2007.

[19] H. Holma and A. Toskala, *LTE for UMTS*, John Wiley & Sons, New York, NY, USA, 2009.