ICT-MobileSummit 2009 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2009
ISBN: 978-1-905824-12-0

Advanced Radio Resource Management for IMT-Advanced in the Framework of WINNER+ Project

Jose F. MONSERRAT[1], Andreas SAUL[2], Gunther AUER[2], Thierry CLESSIENNE[3], Arif OTYAKMAZ[4], Simone REDANA[5], Rainer SCHOENEN[4], Pawel SROKA[6], David MARTIN-SACRISTAN[1], Nikolaos PAPAOULAKIS[7]
[1]*Universidad Politecnica de Valencia – iTEAM Research Institute, Valencia, Spain*
Tel: +34963877007, Fax: + 343879583, Email: jomondel@iteam.upv.es
[2]*DOCOMO Euro-Labs, Munich, Germany*
[3]*Orange Labs, Paris, France*
[4]*RWTH Aachen Univeristy – ComNets, Faculty6, Aachen, Germany*
[5]*Nokia Siemens Networks, Munich, Germany*
[6]*Poznan University of Technology, Chair of Wireless Communications, Poznan, Poland*
[7]*National Technical University of Athens, Athens, Greece*

Abstract: The race towards IMT-Advanced was recently started by ITU-R who distributed a Circular Letter asking for the submission of new technology proposals. The European Celtic project WINNER+ is bridging together experts from industry, academia and government all around Europe to devise this next fourth generation mobile, 4G. This paper presents the first set of innovative concepts for advanced Radio Resource Management that has been identified by the Innovation Group of WINNER+ for potential inclusion in IMT-Advanced. These concepts consist of promising innovative techniques that are ready to be included in current OFDMA-based cellular systems to enhance system performance. A brief description of each technique together with the relevant state of the art is provided.

Keywords: IMT-Advanced, WINNER+, Radio Resource Management, 4G.

1. Introduction

In order to kick-off the definition process for fourth generation (4G) mobile communication systems, termed IMT-Advanced, the ITU-R issued a Circular Letter calling for candidate Radio Access Technologies (RATs) for IMT-Advanced. The Recommendation M.1645 specifies the objectives of the future development of IMT-Advanced family: to reach 100 Mbps for mobile access and up to 1 Gbps for nomadic wireless access. In November 2008 some additional requirements related to technical performance for IMT-Advanced candidate radio interfaces were also described [1].

In April 2008, just after receiving the circular letter, the 3GPP organized a workshop on IMT-Advanced where it was decided that LTE-Advanced, an evolution of current LTE standard, will meet or even exceed IMT-Advanced requirements following the ITU-R agenda [2]. In order to satisfy these LTE-Advanced requirements, radio access technologies adopted in LTE Release 8 must be further enhanced, and new technical components increasing system performance should be incorporated in LTE-Advanced. Regarding radio access network technologies, some key aspects like Self-Organising Networks (SON), signalling optimisation and, most of all, Advanced Radio Resource Management (ARRM) are being discussed in the framework of 3GPP.

The WINNER+ project is aligned to the ITU-R agenda established for IMT-Advanced [3]. In fact, WINNER+ aims at developing and optimising IMT-Advanced compliant technologies that are backward compatible with LTE Release 8. Therefore, WINNER+ project intends to contribute to the future definition of LTE-Advanced. Research activity in WINNER+ is focusing on different aspects such as ARRM, spectrum sharing and its flexible usage, peer-to-peer communications and advanced antenna concepts.

This paper presents the first set of innovative concepts for Advanced Radio Resource Management that has been identified within WINNER+ project. If required, refer to [4] for further information on the performance analysis of all techniques.

The remaining of the paper is organised as follows: Section 2 addresses techniques for dynamic multidimensional resource allocation, Section 3 deals with Self Organizing Networks and, next, Section 4 presents some ideas related to efficient MultiCast (MC) and BroadCast (BC) services. Finally, Section 5 concludes the paper.

2. Dynamic Multidimensional Resource Allocation

The techniques developed in WINNER+ concerning dynamic resource allocation all deal with resource scheduling, but each of them having the focus on a different aspect. Section 2.1 considers the different requirements of different data services for the scheduling, classifies them and tries to fulfil the requirements by prioritisation applying also channel state prediction. Section 2.2 shows a technique for integrating both full- and half-duplex Frequency Division Duplex (FDD) into the WINNER+ system concept. One main aspect here is the integration of relaying, which is a key innovation for 4G. Finally, in section 2.3 cross layer aspects of the resource scheduling are investigated and an optimisation approach of the scheduling is introduced by taking the very special requirements of different applications into account.

2.1 – Utility based scheduling

Many works have been using Utility Theory to propose solutions for the Radio Resource Allocation (RRA) algorithms. Utility theory performs the optimization of a utility-pricing system, which is established based on the mapping of some performance criteria (e.g. rate, delay) or resource usage (e.g. sub-carriers, power) into the corresponding pricing values.

2.1.1 Traffic-Aware Score-Based Scheduling

A Score-Based (SB) scheduler [5] is considered as a candidate technique for the WINNER+ system. However, the proposed scheduler should be designed not only to support multiple users simultaneously but also to serve Real-Time (RT) and Non-Real-Time (NRT) traffic for a user at the same time at any given scheduling time instant. Thus, two scheduling factors, the delay requirements (priority) of scheduling, which may be represented by a time-utility function, and the efficiency of radio resource usage, should be taken into account for designing the scheduler. Hence, a Traffic-Aware Score-Based (TASB) scheduling algorithm has been proposed that uses a modified SB criterion including the priority and time-utility factors of all queued packets to minimize the scheduling delay introduced to RT services. The optimization criterion is defined as follows:

$$i(n) = \arg\min_{j=1,..,K} \frac{s_j(n)}{1+\alpha\sum_{m=1}^{L}\beta_{j,m}U'_{j,m}(t)}, \qquad (1)$$

where $\beta_{j,m}$ is the priority class factor of packet m, $U'_{j,m}(t)$ is the derivative of the time-utility function $U_{j,m}(t)$ of packet m at time t, L is the total number of packets from user j queued for scheduling, α is a constant defining the impact of packet urgency on the resource allocation process, and $s_j(n)$ is the score of user j calculated for resources block n as specified in [5]. The score corresponds to the rank of user's current transmission rate $r_j(n)$ among the past values $\{r_j(n-1), r_j(n-2), ..., r_j(n-W+1)\}$ observed over a window W.

The time-utility function for each packet class depends on the remaining time to the deadline for packet scheduling. Two time-utility functions have been considered: sigmoidal and parabolic function for the RT services, and linear function for NRT services.

The main advantage of the proposed algorithm is that it minimizes the introduced scheduling delay keeping the efficiency of the system (achieved throughput) at the same level as the original SB algorithm. Moreover, the possibility of "tuning" the influence of the urgency factor on the resource allocation process gives a great flexibility, allowing switching from the original SB algorithm to the minimum-delay solution.

2.1.2 Prediction

In this section we present how the scheduling at current time k can be performed not only based on the scheduling in an observation window W but predicting the scheduling in a prediction window W.

A predictive scheduler aims at assigning resources at time k on the basis of the prediction of users scheduling in a time window from k to $k+L-1$. At current time k it schedules the users whose channel dynamics, according to the predictor, would make QoS provisioning using future radio resources more difficult.

The scheduling in the prediction window can be computed recursively (see Figure 1) by applying to the time slot j any QoS scheduler, such as the Utility based Predictive Scheduler (UPS) [5] or the Traffic-Aware Score-Based algorithm (TASB) illustrated in section 2.1.1, beginning from $j=k+L-1$ and decreasing j until the time $j=k$ is scheduled. At each algorithm iteration the scheduler is applied in a window from $k-W$ to j.

During the processing of time j the predicted scheduling from $k-W$ to j is the result of the algorithm applied for the user scheduling at the previous scheduling time (k-1), while from j to $k+L$-1 it has been already updated by the user scheduling procedure at the current time k (as shown in Figure 1).

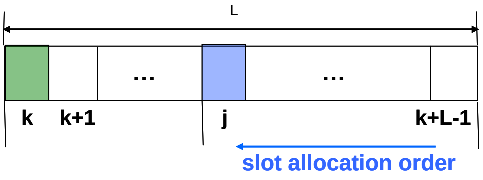

Figure 1: Scheduling in the prediction window L

2.2 – Relay Capable Scheduling for Combined Full-/Half-Duplex FDD

In beyond 3G systems, as WINNER II proposal, the support of half-duplex FDD is taken into account, but the initial concept has to be reconsidered, because it may not be appropriate to integrate neither both full- and half-duplex nor relaying. Figure 2 shows the MAC super-frame structure of the WINNER II system for half-duplex operation [6].

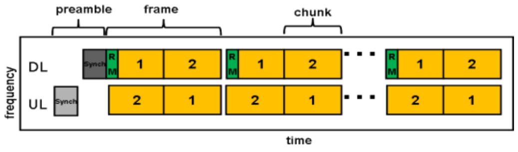

Figure 2: WINNER II half-duplex groups

This super-frame comprises a preamble for synchronisation followed by eight frames each containing two chunks or resource blocks. Each frame begins with a resource map (RM) where the resource allocation information, i.e., which time and frequency slot is reserved for which user terminal (UT), is broadcasted to the UTs associated to a BS. The chunks are used to realise the half-duplex operation of a link. There are two half-duplex groups, say 1 and 2. UTs belonging to group 1 receive in the first half of a frame and transmit in the second half whereas UTs of group 2 do it the other way round. Full-duplex terminals, of course, may transmit and receive all the time. Obviously, in this scheme a UT of group 2 can never receive the resource map since it is always transmitting during the resource map broadcast. Moreover, this scheme is unfair concerning the resource distribution between the two groups since group 1 gets fewer resources in the DL than group 2.

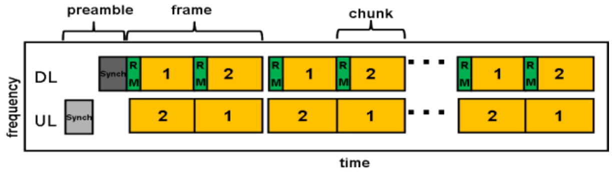

Figure 3: Two resource maps per frame

One way to enable half-duplex scheduling is to precede group 2 in the DL by a RM, hereby ensuring that UTs in group 2 are able to receive resource allocation. This is shown in Figure 3 where a second RM is inserted before the second chunk in a frame. The super-frame structure has twice as many frames compared to Figure 2 while the duplex groups switch from frame to frame. This can lead to a waste of resources if few terminals operate in half-duplex FDD mode, because full-duplex terminals do not need RM differentiation into two groups. To reduce the overhead to only one resource map per frame, the order of the duplex groups could be alternated from frame to frame, i.e., in frame *N* duplex group 1 would receive the RM and in frame *N*+1 group 2 receives the map According to this information contained in the RM spans the contents of two chunks. This scheme can be further developed by interchanging chunks of consecutive frames leading to Figure 4. Each group receives a resource map every other frame and both chunks in a frame belong to the same group.

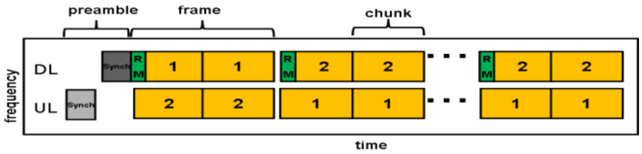

Figure 4: Alternating duplex groups over frames

The integration of relaying increases the complexity of half-duplex scheduling. In the WINNER II system concept [6] it is assumed that RNs behave like a UT towards the BS and behave towards their UTs like a BS. Accordingly, two consecutive task phases of RNs can be identified that switch from frame to frame, namely “BS” and “UT”. RNs are supposed to operate in full-duplex mode. Therefore, the link between RN and BS is easy to schedule. The BS must only ensure that the RN is in a phase where it is acting as a “UT” when it is scheduled by the BS to receive the resource map. During the “BS” phase the RN schedules UTs of the associated half-duplex groups. In the phase marked “UT” the RN operates as UT to the BS. The interesting phase is “BS” where the RN acts as a BS. Since scheduling is done not only for one frame, but must also consider future frames the RN scheduler must take into account the time phases where it is in “BS” role. A half-duplex UT that is scheduled must not switch from frame to frame for UTs directly connected to a BS, but it must switch from “BS” phase to “BS” phase of its serving RN. Since it is, in general, possible that this “BS” task pattern is not fixed, the resource scheduler must switch the UT group to be scheduled, independently, from absolute frame numbers within a super-frame. This becomes clear from Figure 5. The arrows point to the frames that are scheduled during the RM phases. For UTs that are directly served by a BS the duplex groups alternate from frame to frame as visible from the first row in Figure 5. However, for UTs served by a RN the duplex groups alternate on the basis of the "BS" task phase of the RN. This is illustrated in the bottom row of Figure 5. Clearly, there is a gap of one frame between the downlink and uplink phases of both duplex groups and a gap of three frames between consecutive DL and UL phases, respectively. This leads to an increased delay and a lower maximum data rate for UTs owing to the gaps mentioned. The scheme introduced can easily be extended to a scenario with more than two hops.

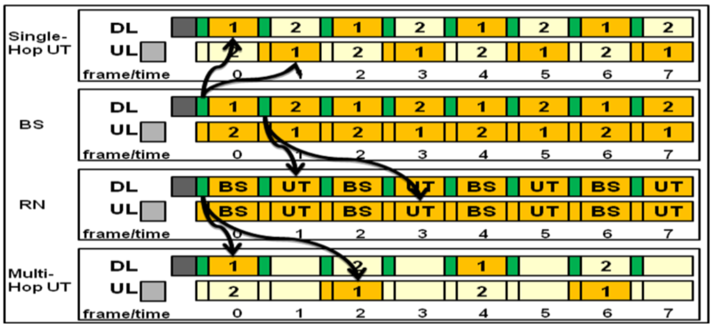

Figure 5: Frame structures of all station types and their dependencies

2.3 Cross-Layer Optimization (CLO) Between Link and Application (APP) Layer

In LTE-Advanced the wireless radio interface is shared between mobile users. In a conventional system without cross-layer optimization (CLO), overloading the system in peak situations with particularly high traffic is to be avoided. Therefore, the average contribution of Quality of Service (QoS) sensitive traffic to the overall system load has to be kept much lower than the maximum load that the system could support. Most of the traffic can be served only on a best effort (BE) basis.

In contrast to that, we are considering a system including CLO. By CLO the resource allocation at the link layer and the resource consumption of the application (APP) layer are controlled and matched by an optimizer. This allows a better utilization of the resources in a situation where many users shall be served with a required QoS. Instead of locally maximizing the efficiency of each single layer with CLO the overall efficiency of the system is improved. In our approach CLO is not replacing the conventional system design. It is rather considered as optional add-on for improving the resource efficiency of eNodeB's with a high traffic load.

An indispensible condition for the considered CLO is that the applications can adapt their data rates in a certain range following the decision of a cross-layer optimizer. In practice, from application point of view this requirement could be fulfilled, e.g., by encoding a video or audio stream with different data rates and quality levels. Enabling technologies are for example transcoding at the eNodeB, or using modern codecs, for example based on the MPEG-4 advanced audio coding (AAC) standard with bit slice arithmetic coding (BSAC), or like the scalable video coding (SVC) extension of the H.264 advanced video coding (AVC) standard, etc. Scalable codecs allow an easy rate adaptation with graceful quality degradation by dropping certain parts of the stream, e.g., at the eNodeB.

At the cross-layer optimizer we describe the link layer in terms of long-term average link capacities and data rates [5]. For the APP layer the objective is to improve the user perceived quality, which can be described by mathematical models like the objective mean opinion score (MOS) [7] and depends, e.g., on the possible data rates. Using these long-term properties of applications and channels allows limiting control signalling to a few times per second. For the considered CLO delay issues are sufficiently handled at the scheduler can be neglected at the optimizer.

An optimizer maximizing the sum of the MOS of all users can improve the average perceived quality significantly [7]. In the WINNER+ project we are investigating different objectives, e.g., based on minimax theory. As published in [5], setting up the optimization problem to guarantee a certain minimum perceived quality allows serving more than two times the number of users for the chosen parameters in comparison to a max-sum-MOS approach. Moreover, differentiating priority classes, e.g., to support ordinary and premium users, can be done easily and effectively while maintaining high fairness among the users of each class with the proposed approach.

3. Self-Organizing Networks

Self-organisation allows the network to detect changes, make intelligent decisions based upon these inputs, and then implement the appropriate action. The systems must be location and situation-aware, and must take advantage of this information to dynamically configure themselves in a distributed fashion.

In the first phase of WINNER+ four advanced techniques have been devised. Two of them are directly related to the concept of self-organization. The first one, described with more detail in this paper, proposes a new mechanism to reduce the level of interference in the system. The second one is related to the automatic response of the system against congestion situations. Besides, two additional cognitive tools are provided. These tools, one for traffic characterisation and another for traffic prediction, allows a higher degree of cognition in the system and therefore could be used by other techniques willing to improve their performance.

3.1 – Contention Free Dynamic Slot Allocation in Cellular Networks

When several users in a shared channel simultaneously attempt to access a given time-frequency resource block (slot), collisions due to co-channel interference (CCI) are encountered. Reservation protocols, such as reservation ALOHA (R-ALOHA) [8] and packet reservation multiple access (PRMA) [9], divide the available resources to idle and reserved slots. For R-ALOHA idle slots are allocated in contention and reserved slots are protected from CCI as follows [8]:

- **Contention**: If the slot is sensed idle a packet is transmitted to contend with other users for an unreserved slot. In case of collision the packet is retransmitted.
- **Reservation**: Upon successful reception the receiver broadcasts an acknowledgment. This ACK reserves the slot and other users do not use that slot in future transmissions.

In wireless networks, slot reservation translates to an exclusion region around an active receiver. A competing communication link is denied access to a reserved slot if its transmitter is located within the exclusion region; otherwise the slot may be concurrently accessed by both links. An efficient realization of R-ALOHA in decentralized wireless networks is provided by the busy signal concept, where the receiver acknowledges successful reception by means of a time-multiplexed busy burst. The range where the busy-signal power, I_{nk}^{b}, observed at slot *n* of frame *k*, exceeds a pre-defined threshold I_{th}. notifies a potential transmitter that it is within the exclusion region of an active receiver. While a busy burst enabled reservation protocol ensures uncontested use of reserved slots, collisions in contention are encountered, caused by simultaneously accessed idle slots from adjacent cells.

The proposed cellular slot allocation and reservation (CESAR) protocol utilize inter-cell coordination by resource partitioning so to mitigate the contention phase typically encountered in distributed slot reservation protocols, such as R-ALOHA. CESAR grants access to unreserved slots based on two conditions: 1.) the slot is sensed idle and 2.) a pre-defined resource partitioning pattern issues an access right to a given cell. While resource partitioning ensures that at most one cell may access an idle slot at a time, cyclically shifting the partitioning pattern allows each cell to successively contend for all slots. Hence, CESAR imposes no restrictions on the amount of resources one cell may allocate, and therefore overcomes the limitations of classical inter-cell resource partitioning based on static frequency reuse planning.

For resource partitioning with frequency reuse factor *R*, cells are organized into *R* pre-defined cell groups, such that adjacent cells are in different cell groups *G*. Destructive interference from nearby cells is mitigated by assigning mutually orthogonal slots to different cell groups, while cells that belong to the same cell group *G* spatially reuse resources. CESAR utilizes the cyclically shifted resource partitioning pattern

$$g_{nk} = n+k \mod R \qquad (2)$$

that associates slot *n* at frame *k* to cell group $G = g_{nk}$.

The proposed CESAR policy accomplishes two objectives: cell *c* retains all previously reserved slots, identified by the reservation indicator $\rho_{nk}^{c} = 1$; in addition, cell *c* is granted contention free access to idle slots that satisfy (2). Cell *c* may access slot (*n*,*k*) if the following condition is met

$$g_{nk} = G \quad \text{and} \quad I_{nk}^{b} \le I_{\text{th}} \quad \text{or} \quad \rho_{nk}^{c} = 1 \qquad (3)$$

In an unreserved slot (*n*,*k*) of cell *c* is sensed idle, if all active out-of-cell receivers are outside the exclusion range of cell *c*, such that $I_{nk}^{b} < I_{th}$. Otherwise slot (*n*,*k*) is sensed busy. Then cell *c* is denied access for slot (*n*,*k*) in (3), regardless the outcome of (2).

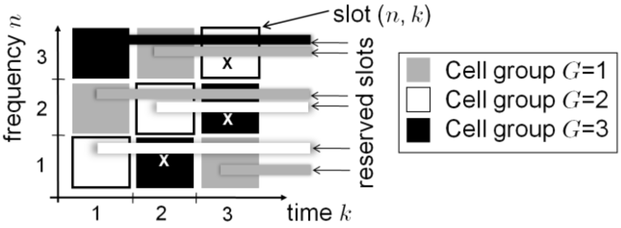

Figure 6: CESAR working principle: slots are successively accessed by virtue of cyclically shifted resource partitioning patterns, provided that the interference induced in adjacent cells is sufficiently low

The working principle of CESAR is illustrated in Figure 6. Initially at frame *k*=1 all slots are idle, $\rho_{nk}^{c} = 0$, $\forall n, c$, so that (3) allows each cell to initially allocate N_s/R slots, where N_s is the number of slots per frame. In subsequent frames the cyclic shift of the partitioning pattern (2) allows adjacent cells to successively probe slot *n*. To this end, slot *n* previously reserved by cell *c*, may be accessed by cell *i*, if the interference induced to cell *c* is sufficiently low, such that $I_{nk}^{b} < I_{th}$ in (2), giving rise to spatially reused slots that are concurrently reserved by adjacent cells. After $k \ge R$ frames all slots are either reserved or busy so that CESAR converges to a steady state.

4. Efficient Multicast and Broadcast Services

The WINNER system aims at providing an efficient transmission of Multimedia Broadcast and Multicast Services (MBMS) to several users simultaneously in the future communication mobile network in such a way that the entire cell is covered. With this end three emission modes can be considered. The broadcast mode consists in a unidirectional point-to-multipoint (p-t-m) transmission of MBMS data from a single source to all users in an MBMS area. The multicast mode allows the p-t-m transmission of the same service in the same area from the same source but to a multicast group. Finally, the unicast mode allows the bidirectional point-to-point (p-t-p) transmission to only one user.

From a point of view of deployment scenarios, the p-t-p/p-t-m switching algorithm will depend on the networking mode. Two scenarios of broadcast networking have been considered in WINNER+: the Multi-Frequency Networking (MFN) and the Single Frequency Networking (SFN). The results indicate that SFN is serious candidate for broadcasting in fact with SFN the terminal can derive benefit from the neighbouring cell, and without consuming excessive resource in term of allocated sub-carriers and receiver complexity. [5] gives an example of criterion for the exclusive choice: the number of active users in the network obtained by a counting process. The results indicate that only the nearest third part of the cell is concerned by the switching process so to make this process more feasible, many conventional physical features, as beamforming, must be used to improve the p-t-p performance.

5. Conclusions

This paper has described some promising techniques for RRM to be implemented in the future IMT-Advanced systems, highlighting the impact of these techniques in the WINNER+ system architecture. Three innovative concepts have been identified, namely Dynamic Multidimensional Resource Allocation, Self-Organizing Networks and Efficient Multicast and Broadcast Services. A careful description of those concepts has been provided together with their repercussion in the system. It is worth highlighting that neither technique has an effect on the current WINNER architecture concept (very similar to LTE) as defined in the WINNER II project. Therefore, the system in mind is OFDMA-based, with both TDD and FDD modes available and with the capability of using multi-hop transmission by means of relay nodes, either fixed or mobile. System architecture encompasses just a set of base stations (or eNodeB’s, according to the 3GPP nomenclature) and a gateway node that connects the system with other IP-based networks. Only potential slight modifications on this flat architecture are envisaged to cope with the needed data interchange among base stations.

Acknowledgements

This article has been written in the framework of the CELTIC project CP5-026 WINNER+. The authors would like to acknowledge the contributions of their colleagues.

References

[1] ITU-R Report M.2134, “Requirements related to technical performance for IMT-Advanced radio interface(s),” November 2008.
[2] 3GPP TR 36.913, Requirements for Further Advancements for E-UTRA (LTE-Advanced),” June 2008.
[3] WINNER+ Project CELTIC CP5-026, http://projects.celtic-initiative.org/winner+/
[4] CELTIC CP5-026 WINNER+, Deliverable D1.1, “First set of best innovations in advanced radio resource management,” February 2009.
[5] T. Bonald, “A Score-Based Opportunistic Scheduler for Fading Radio Channels,” in Proceedings of the European Wireless Conference, Barcelona, February 2004.
[6] IST-4-027756 WINNER II, Deliverable D6.13.14 “WINNER II system concept description,” Nov. 2007.
[7] S. Khan, S. Duhovnikov, E. Steinbach et al., “Application-driven Cross- layer Optimization for Mobile Multimedia Communication using a Common Application Layer Quality Metric,” in Second International Symposium on Multimedia over Wireless (ISMW'06), Vancouver, Canada, Jul. 2006.
[8] W. Crowther, R. Rettberg, D. Walden, S. Ornstein, and F. Heart, “A System for Broadcast Communications: Reservation-ALOHA,” in Proc. 6th Hawaii Int. Conf. Sys. Sci., Jan. 1973, pp. 371–374.
[9] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, and B. Ramamurthi, “Packet Reservation Multiple Access for Local Wireless Communications,” IEEE Trans. Commun., vol. 37, no. 8, pp. 885–890, Aug. 1989.