Joint Radio Resource Management in Cognitive Networks: TV White Spaces Exploitation Paradigm

Athina Bourdena¹, Prodromos Makris¹, Dimitrios N. Skoutas³, Charalabos Skianis¹ and George Kormentzas¹

¹E-mail: abourdena@icsd.aegean.gr, {pmakris, d.skoutas, cskianis, gkorm}@aegean.gr
²Dept. of Information and Communication Systems Engineering, University of the Aegean, GR-83200, Karlovassi Samos, Greece

ABSTRACT
In this book chapter, joint radio resource management (JRRM) issues in cognitive networks are discussed presenting the TV White Spaces (TVWS) spectrum exploitation use case. TVWS are portions of UHF spectrum, which will be released and interleaved according to the geographical region due to the gradual switch-off of analogue TV and the adoption of digital TV. With the availability of TVWS and their temporary lease, traditional network planning and RRM design rationale points need to be enhanced. This book chapter initially provides a state-of-the-art work for existing cognitive radio network architectures, while a reference architecture for commons and secondary TVWS trading is proposed. Subsequently, JRRM concepts for heterogeneous Radio Access Technologies’ extension over TVWS aiming to continuously guarantee the QoS, the network key performance indicators and at the same time targeting the overall highest system capacity, are presented. Finally, a thorough classification of existing admission control and scheduling techniques is provided, outlining the need for including continuously more cognitive and context-aware features in JRRM algorithms being applicable in advanced heterogeneous networking (HetNet) environments.

1 INTRODUCTION
Emerging types of wireless network services and applications, rich in multimedia content with increased requirements for network resources and guaranteed end-to-end QoS provisioning, raise the needs for higher frequency availability and create new challenges in radio-spectrum (i.e. the fundamental resource in wireless telecommunication networks) management and administration. While the utilization of advanced digital signal processing techniques enable for efficient radio-spectrum exploitation, even under the traditional “command-and-control” spectrum administration/management policy, there is a worldwide recognition that such methods have reached their limit and are no longer optimal. In fact, radio-spectrum utilization studies have resulted that most of the licensed spectrum is under-utilized (McHenry et al., 2004), and considerable parts of it would be available when both space and time dimensions are taken into account. An example of under-utilized radio-spectrum, is the so-called “television white spaces” (TVWS) that comprise of VHF/UHF frequencies, either released/freed by the digital switchover process (“Spectrum/Digital Dividend”), or being totally unexploited, mainly at local level, due to frequency planning issues and/or network design principles (“Interleaved Spectrum”) (Australian Communication &
TVWS include tenths of MHz at local/regional level (OFCOM, 2008), enable for low cost and low power systems design, provide superior propagation conditions for building penetration, while at the same time their sufficiently short wavelength facilitate the construction of resonant antennas, at a size and shape that is acceptable for many handheld devices. Therefore, TVWS are well suited for wireless network applications and services, provided by sophisticated telecommunication systems. However, the current “command-and-control” administration/management policy allows only for primary (i.e. licensed) systems to exploit TVWS for the provision of primary services, such as terrestrial digital video broadcasting (DVB-T), handheld digital video broadcasting (DVB-H), interactive (iTV), Programme Making and Special Events (PMSE), while prohibiting any other secondary transmission. Hence, the problem of spectrum scarcity, as perceived today, is due to inefficient radio-spectrum management/administration, rather than the wireless resources shortage. The envisioned frameworks and schemes that are proposed in this book chapter include policies (Unlicensed Operation in the TV broadcast bands, 2009) in which secondary (i.e. unlicensed) systems are allowed to opportunistically utilize the underused primary TV channels. TVWS spectrum exploitation paradigm will be used as a vehicle in order to address joint radio resource management (JRRM) issues in cognitive radio networks, too.

Cognitive radio (CR) techniques provide the ability in a network to share the available radio spectrum, under an opportunistic basis. Cognitive radio networks adapt their transmitter parameters, based on real-time interaction with their spectral environment, by exploiting portions of frequency bands that are unused at a specific time or a geographical location. The choice of radio spectrum selection has to be performed, under an efficient process without causing interference with other licensed systems operating in the same frequency band. A cognitive radio network can be set to transmit and receive on a variety of different frequencies, exploiting alternative access technologies supported by its hardware design. Through this ability, the most optimum radio spectrum band and the most appropriate operating parameters can be selected and reconfigured.

Towards making full use of cognitive radio networks benefits and improving spectrum efficiency, cognition, intelligent decision-making and the reconfiguration abilities are adopted based on a cognition cycle. More specifically, cognition ability includes acquisition of multi-domain environment cognition information, efficient cognition information transmission and usage. Intelligent decision making ability includes the ability of making decision adaptively, according to the dynamic changing environment and the ability of improving the end-to-end efficiency. Reconfiguration ability includes the reconfiguration of cognitive radio networks, according to previous intelligent decisions for an end-to-end efficiency purpose. Finally, cognition cycle is also exploited, in order to observe the network, perceive current network conditions and make an optimum decision.

The main learning objectives of this book chapter are the following: a) investigate JRRM issues to optimize spectrum utilization, minimize interference, guarantee fairness in TVWS access and minimize spectrum fragmentations, b) propose an abstract cognitive radio architectural framework for TVWS exploitation paradigm, c) provide means for design and implementation of enabling technologies based on cognitive radio to support mobile applications over TVWS for various spectrum sharing models, d) introduce techniques that integrate QoS/QoE aspects in dynamic spectrum management, and e) introduce context-aware and autonomous networking functionalities for JRRM algorithms in cognitive networks. The purpose of the chapter is to present ways that a traditional but always up-to-date research topic such as JRRM can be applicable in efficient TVWS spectrum exploitation. The rationale of the technical content included, is to give to intended audience an overall view of TVWS paradigm and to present specific means that general cognitive and context-awareness concepts can be integrated in related state-of-the-art frameworks towards envisioning the real evolution of cognitive networks and self-adaptive communication systems.

The remainder of the book chapter is structured as follows. In section 2, a comprehensive classification of state-of-the-art cognitive radio architectures is provided accompanied by an exhaustive categorization of existing RRM optimization algorithms, which can be used in the TVWS exploitation paradigm. Section 3 proposes a novel CR framework for TVWS exploitation based on real-time secondary spectrum market
(RTSSM) policy undertaken in the context of FP7-ICT-COGEU project (COGEU, 2012), while various TVWS allocation policies are analyzed. In section 4, the concept of joint radio resource management (JRRM) for heterogeneous radio access technologies’ (RATs) extension over TVWS is introduced by focusing on JRRM both from the spectrum broker and local resource manager (LRM) perspectives. In sections 5 and 6, the two most important LRM functionalities are extensively analyzed. More specifically, context-aware joint admission control and scheduling techniques are presented in sections 5 and 6 correspondingly. Future research trends and directions are given in section 7 providing insightful considerations about the future of cognitive networks and self-adaptive communication systems from the perspective of efficient TVWS exploitation paradigm. The chapter ends up with some concluding remarks, while extensive related bibliography list is provided for additional reading, too.

2 BACKGROUND

2.1 State of the art of existing cognitive radio network architectures

Cognitive radio concept (Mitola & Maguire, 1999) is researched and developed in response to the current wireless networks’ needs for increased spectrum availability, as well as to efficiently exploit the available radio resources (McHenry et al., 2004). Deployment of CR networks may satisfy/fulfill the increasing users’ demand for bandwidth-hungry QoS-sensitive mobile services, by allowing dynamic access to the available spectrum portions, along with on-demand utilization of the radio resources. In this respect, CR technologies allow for the establishment of frequency-agile devices, utilizing sensing mechanisms, in order to detect TVWS channels and adapt their transmission characteristics, so that primary users are not interfered by secondary ones. To achieve these, CR networks are exploiting network architectures that can be characterized either as infrastructure-based or ad-hoc, single-hop or multi-hop and centralized or distributed ones (Hossain et al., 2009). As shown in figure 1 below, distributed network architectures can be implemented in both infrastructure-based and ad-hoc cognitive radio networks. Nevertheless, a centralized dynamic spectrum access can be obtained only in an infrastructure-based network, thus a central controller is required to orchestrate the spectrum access of secondary users. In the case of a multi-hop infrastructure-based network, one of the relay stations can assume the responsibility of controlling dynamic spectrum access.

Figure 1: Categories of CR Network Architectures

More specifically, if the network topology frequently differentiates/changes, CR architectures can be characterized as infrastructure or ad-hoc, while the communication along the nodes characterizes the architecture as single or multi-hop. For instance, in a single-hop, infrastructure-based CR architecture, communication among secondary users is assured, via a central controller, which is responsible for the secondary spectrum access coordination. On the other hand, in a multi-hop, infrastructure-based CR architecture, a number of base stations are utilized (i.e. relay nodes), enabling secondary users to exchange data, even though they are not in the same transmission range of each other. Furthermore, if a central controller takes the decision, regarding the spectrum allocation process, then the network architecture is a centralized one. The central controller is responsible to collect information, regarding spectrum usage of the primary users, as well as information about the transmission requirements of the secondary ones. Based on this information, an optimal allocation solution that maximizes spectrum utilization can be performed. The decisions of the central controller are communicated/broadcasted to all secondary users that are connected in the network. Alternatively, in case of distributed decision process for dynamic radio-resource exploitation, a secondary user can autonomously decide on spectrum access. Since each secondary user has to collect information, regarding the radio environment and make its decision locally, the CR transceiver of each secondary user requires greater computational resources than those required in centralized architectures. The communication overhead in this case would be smaller,
2.2 Radio Resource Management Approaches

In all cases, and no matter which architecture is utilized, vital part of CR networks is the radio resource management (RRM) entity (Bourdena et al., in press; Bourdena et al., 2011; Bourdena et al. 2012), which is considered as an optimized solution for optimally allocating network resources. While this optimization can be generally focused on optimizing either a single objective or a set of objectives, the nature of the wireless communications most of the times requires the multi-objective one. More specifically, the optimization goal in CR paradigm, is focused on formulating networking problems, as optimization ones considering resource allocation. Usually, multi-objective optimization can be performed, following either the decision making theory concept or the game theory concept. Whereas the former tries to find an optimal solution through classical mathematical rationalization, the latter considers the optimization problem, as a game theory process (Niyato & Hossain, 2007; Niyato & Hossain, 2008).

The decision making approach is based on formulating, (by maximizing or minimizing an objective function), as well as on setting equality and inequality constraints that the optimal solution is required to respect (Hossain et al., 2009). As it is illustrated in figure 2, three groups of solutions arise for this type of optimization approach:

![Figure 2: Categories of Optimization Algorithms](image)

1. Closed form solution, which can be exploited in cases that at least one solution can be expressed as a closed-form expression and it has an analytic solution. An equation is considered to be a closed-form solution if it solves a given problem in terms of functions and mathematical operations from a given generally accepted set. For example, an infinite sum would generally not be considered closed-form.

2. The next group of decision-making approach is the integer/combinatorial programming that involves parameters with integer values or with combinatorial nature, thus means that a finite number of possible solutions exist. These problems are referred as multi-objective ones and they can be solved by searching for the optimal solutions among the entire set of possible solutions. The goal of the integer/combinatorial programming is shortening the search space to a smaller one. Such an example is backtracking algorithm (Skiena, 2008) that is discussed below in the next paragraphs. In CR networks, integer/combinatorial optimization problem formulations can be used to perform efficient resource allocation methods, which meet the desired objectives when the values of some or all of the decision constraints have to be integers, such as modulation, channel allocation, and coding rate.

3. Furthermore, the last group of the decision making approaches is the mathematical programming that is used for most real-world optimization problems and can be divided into five major subfields, i.e. linear, convex, non-linear, dynamic and stochastic programming.

   a. Linear programming is used to determine a way to achieve the best solution in a given mathematical model for some list of requirements, represented as linear relationships, subject to linear equality and linear inequality constraints.

   b. The convex programming is based on convergence of the considered values towards the highest local value by exploiting the equality of the local and global optimum.

   c. The optimization process, involving non-linear objective functions and constraints, is called non-linear programming, where the key difference with the linear one is the inequality between the local optimum and the global optimum. One of the greatest challenges in non-linear programming is that several problems perform local optima. These situations are similar to the multiple peaks. It is difficult for an algorithm that tries to move from point to point only by climbing uphill, since the peak it achieves might not

but since each user has only local information, the optimal solution for spectrum access may not be achievable by all unlicensed users.
be the highest. Algorithms that propose to overcome this difficulty are termed as “global optimization” algorithms. Popular solutions for solving a non-linear programming problem are genetic algorithms and simulated annealing (Skiena, 2008) that are presented below in this section.

d. Also, the dynamic programming approach solves a high complexity problem by combining the solutions of a series of low complexity sub-problems. Dynamic programming relies on the principle of optimality, which states that in an optimal sequence of decisions or choices, each subsequence must be optimal, too.
e. The last subfield of mathematical programming is the stochastic programming, which is an optimization process that includes probabilistic elements in the problem formulation and refers to the minimization/maximization of a function in the presence of randomness in the optimization process.

On the other hand, for game theory approach, two groups of optimization solutions arise as it is illustrated in figure 2:

1. Nash equilibrium, which is a solution concept of a non-cooperative game, involving two or more players. Each player is assumed to know the equilibrium strategies of the other players, and no player, has anything to gain by changing only his own strategy unilaterally (Osborne & Rubinstein, 1994). If each player has chosen a strategy and no player can benefit by changing his/her strategy while the other players keep theirs unchanged, then the current set of strategy choices and the corresponding payoffs constitute a Nash equilibrium.

2. Pareto optimality, where no one can be made better off without making at least one individual worse off. Given an initial allocation of goods among a set of individuals, a change to a different allocation that makes at least one individual better off without making any other individual worse off is called a Pareto improvement. An allocation is defined as "Pareto efficient" or "Pareto optimal" when no further Pareto improvements can be made. Pareto efficiency is a minimal notion of efficiency and does not necessarily result in a socially desirable distribution of resources: it makes no statement about equality, or the overall well-being of a society (Barr, 2004; Sen, 1993).

At the remainder of this section, we emphasize on three specific algorithms namely backtracking, simulated annealing and genetic ones, as these seem to have the best applicability in the TVWS exploitation paradigm.

Based on the above, backtracking is a general algorithm for finding all possible solutions to some computational problem, that incrementally builds candidates to the solutions, and abandons each partial candidate ("backtracks") as soon as it determines that the partial candidate cannot possibly be completed to a valid solution. Backtracking can be applied only for problems, which admit the concept of a "partial candidate solution" and a relatively quick test of whether it can possibly be completed to a valid solution. Backtracking is often much faster than brute force enumeration of all complete candidates, since it can eliminate a large number of candidates with a single test. Backtracking is an important tool for solving constraint satisfaction problems, such as spectrum allocation in CR networks. Backtracking depends on user-given "black box procedures" that define the problem to be solved, the nature of the partial candidates, and how they are extended into complete candidates. In case of frequency allocation problem in CR, the algorithm generates all possible solutions based on the available network resources and secondary systems requests by repeatedly choosing an optimum spectrum allocation solution. It is therefore a meta-heuristic rather than a specific algorithm, although, unlike many other meta-heuristics, it is guaranteed to find all solutions to a finite problem, avoiding both repetitions and missing solutions in a bounded timeframe.

On the other hand, Simulated Annealing (SA) is a heuristic computational technique derived from statistical mechanics for finding near globally-minimum-cost solutions to large optimization problems, which can be applied in a network resource allocation process. SA algorithm gives an initial solution in the solution set with objective value, searching its neighborhood for a solution of lower value, and repeating until no further improvement is possible. The final solution obtained is optimal within its
neighborhood, or locally optimal. Neighbors of a solution are usually obtained by performing small transformations to it, called moves. The two main issues in local search are the design of effective neighborhood functions and the design of search strategies that are able to escape from poor local optima. In case of TVWS allocation, the SA algorithm tries to avoid local optima, thus creating and evaluating solutions, as a matter of spectrum utilization and spectrum fragmentation values/measurements. Finally, Genetic Algorithms (GA) are search algorithms that operate via the process of natural evolution and selection. The algorithms are initiated with a sample set of potential solutions, which then evolve towards a set of more optimal solutions. Within the sample set, solutions that are poor tend to die out, while better solutions propagate their advantageous attributes, thus introducing more solutions into the set that feature greater potential. A random mutation process allows to guarantee that a set of potential allocation solution, in case of TVWS allocation problem, will not remain stagnant and it can be efficiently filled up with numerous copies of the same solution. Moreover, GA perform better than traditional optimization algorithms because they are capable to avoid local optima, by taking advantage of an entire set of solutions spread throughout the solution space.

3 TV WHITE SPACES ALLOCATION POLICIES

Although conceptually quite simple, the introduction of CR networks in TVWS represents a disruption to the current “command and control” paradigm of TV/UHF spectrum management, and therefore the exploitation of the pre-mentioned architectural/technological CR solutions is highly intertwined with the regulation models that would eventually be adopted. More specifically, command and control spectrum management regime is employed by the most regulators. This approach adopts that the regulators are the centralized authorities for radio spectrum allocation and usage decisions. The allocation decisions are often static in temporal and spatial dimensions, meaning that they are valid for extended periods of time (i.e. usually decades) and for large geographical regions (i.e. country wide). The usage of radio spectrum is often set to be exclusive, where each band is dedicated to a single network provider, maintaining interference free communication. Command and control spectrum management regime dates back when the technologies employed required interference-free mediums for achieving acceptable quality. Thus, it is often argued that the exclusive nature of the command and control approach is an artifact of outdated technologies. Among the envisaged regulation models are the “Real-time Secondary Spectrum Market” (or licensed policy) and the “Spectrum of Commons” (or unlicensed policy), as depicted in figure 3.

![Figure 3: TVWS allocation policies](image)

The “Real-time Secondary Spectrum Market” (RTSSM) (or licensed policy) may be the most appropriate solution, especially for applications that require sporadic access to spectrum and for which QoS guarantees are important. RTSSM regime adopts spectrum trading, which allows primary users (license holders) to sell/lease spectrum usage rights and secondary players to buy them (license vendees), thereby establishing a secondary market for spectrum leasing and spectrum auction. The license holder runs an admission control algorithm, which allows secondary users to access spectrum only when QoS of both primary and secondary users is adequate. The trading of secondary use may also occur through intermediaries, such as a spectrum broker, exploiting radio resource management algorithms (RRM) for determining the frequency, at which a secondary user should operate along with the economics of such transactions. TVWS availability is provided by a geo-location spectrum database to spectrum broker for specific geographical locations. Secondary users, on the other hand, dynamically request access when-and-only-when spectrum is needed and are charged based on spectrum utilization basis, as a matter of types of services, access characteristics and QoS level requests. The access types could consist of a long-term lease, a scheduled lease and a short-term lease or spot markets. Each type requires different discovery mechanisms and applies with different levels of service agreements.
On the other hand, “Spectrum of Commons” (or unlicensed policy), represents the case where coexistence with incumbent primary transmissions (e.g. DVB-T) is assured via the control of interference levels rather than by fixed spectrum assignment. In a “spectrum of commons” usage model there is no spectrum manager to preside over the resource allocation, similarly to the wireless ISM bands, where users have to fulfill the technical rules ensuring good coexistence, but do not need to negotiate with existing players. However, despite the fact that unlicensed spectrum promotes efficiency through sharing, QoS cannot be guaranteed, which is a serious problem especially for QoS-sensitive applications. Sensing techniques for reliable detection of TVWS and coexistence mechanisms for interference avoidance are the main technical challenge. Defining spectrum policies and etiquette rules to promote fairness and avoid the “tragedy of the commons” (Australian Communication & Media Authority, 2007) are also key challenges.

Spectrum sensing is a key functionality of cognitive radio systems that may incorporate power control mechanisms, able to dynamically adjust transmission power for an efficient exploitation of spectrum opportunities in TVWS. Therefore, the objective of sensing is to get reliable context information, in order to flexibly set the maximum transmission power per channel at the operating location to avoid causing interference to incumbent devices. Autonomous sensing techniques rely only on the power strength measured in specific CR locations. The decision whether a TV channel is occupied or idle is performed, by comparing the measured power strength with a threshold level. The analysis from (COGEU Deliverable 3.1, 2010) showed that it is not possible to set the maximum transmission power, simply based on the power strength detected by the CR device. Therefore, a CR network may adopt a hybrid approach, where local sensing information is combined with geo-location database information to compute the TVWS spectrum pool.

There are several advantages for the use of geo-location information to support the detection of incumbent systems. The most important is that the database stores the required information to compute the TVWS spectrum pool available in a specific location. Information, such as DVB protected areas; specifications of DVB transmitters, advance propagation models, protection rules, can be used to compute the maximum transmission power. With a database, part of the complexity associated with sensing and maximum power computation is transferred to the core network, decreasing complexity and power demand of TVWS devices. The database has the ability to be dynamically updated and continuously adjust interference protection parameters in line with the evolution of incumbent standards, e.g. DVB-T2. In addition, TVWS spectral utilization efficiency is better than using sensing alone detection. This is primarily due to the ability of geo-location enabled TVWS devices to accurately determine protected service contours.

3.1 Proposed CR framework for TVWS exploitation based on RTSSM Policy

The proposed approach considers a centralized topology with a Geo-location Spectrum database dealing directly with TVWS Devices (Spectrum of Commons policy) or with Spectrum Broker (RTSSM policy). An overview of the spectrum broker reference architecture is presented in figure 4. The centralized topology approach was adopted as the most appropriate solution in this case, since QoS guarantee is crucial in the proposed system. Furthermore, such a centralized topology enables for radio spectrum trading, establishing a secondary market for spectrum leasing and spectrum auction. Spectrum broker entity in this network topology controls the amount of bandwidth and power assigned to each secondary user, in order to keep the desired QoS and interference below the regulatory limits. In this reference model, the centralized broker is an intermediary between the Geo-location database (TVWS information supplier) and players that negotiate spectrum on behalf of spectrum users. The reference architecture supports both Spectrum of Commons and RTSSM.

Figure 4: Reference architecture for commons and secondary TVWS trading
More specifically, this approach supports a broker-based CR network architecture for the efficient exploitation of TVWS under the RTSSM regime and comprises of two core subsystems: a) a Spectrum Broker responsible for coordinating TVWS access and administrating the economics of radio-spectrum exploitation, and b) a number of Secondary Systems (i.e. mobile network operators and wireless network providers), competing/requesting for TVWS utilization. In particular, this network architecture consists of secondary systems that provide different services classes depending on the type of service, voice data, etc. According to this architecture, Spectrum Broker consists of several sub-entities, such as a Payment System, a Dynamic TVWS Allocation Mechanism (RRM module), Trading and Price Discovery mechanism (Trading module), Registration and Validation mechanism and a number of repositories. More specifically, the TVWS occupancy repository obtains information from the national database, namely the Geo-location spectrum database, which includes data regarding the available TVWS in specific locations and the maximum allowable transmission power of secondary systems per channel, in order to avoid causing interference to primary systems. The TVWS occupancy repository creates a spectrum-portfolio, including all the above-mentioned information that is advertised to bidders. Moreover, the RRM module matches the secondary systems’ requirements with available resources and thus allocates the TVWS based on QoS requirements. The TVWS allocation mechanism implements an algorithm that uses information from the Geo-location database to determine the TVWS bands and power, at which a secondary system should be allowed to operate, in order to avoid spectrum fragmentation, optimize QoS and guarantee fairness in TVWS access. Moreover, trading module is responsible to determine the revenue of Spectrum Broker, which aims to trade/lease spectrum with temporary exclusive rights to the most valuable bidder. Finally, the trading information repository hosts information about the TVWS selling/leasing procedure, as well as the spectrum-unit price to be exploited during the trading phase, creating a price-portfolio.

The proposed approach supports two alternative TVWS allocation mechanisms following either a fixed-price or an auction-based policy. In case that a fixed-price policy is adopted, an optimization algorithm (e.g. Backtracking, Simulated Annealing, Genetic Algorithm) obtains the best-matching/optimal solution by minimizing an objective function, as a matter of spectrum fragmentation and/or Secondary Systems’ prioritization (e.g. in case that some secondary technologies must be served before others). Alternatively, in the auction-based policy, the spectrum broker collects bids from the secondary systems, and subsequently determines the allocation solution along with the price for each spectrum portion from a price portfolio in order to maximize the spectrum broker profit. The auction process is then being repeated as soon as spectrum portions are available.

On the other hand, the proposed approach considers two types of TVWS Devices (TVWSD or WSD), exploiting Spectrum of Commons policy. Master devices contact Geo-location spectrum database (see figure 4) in order to obtain a set of available frequencies in their geographical location and Slave devices obtain the relevant information from master ones. The main information that needs to be communicated by master devices to the geo-location database is expected to be location, location accuracy, expected area of operation (optional), coverage area, and device type. Moreover, technical information that is required to be transferred to the master devices originate from the Geo-location database and include available frequencies (minimum requirement), maximum transmit power and the appropriate national/regional database to consult.

**3.2 Performance evaluation results**

Towards verifying the validity of the proposed RRM algorithms and the capacity of the proposed CR network architecture for efficient TVWS exploitation and QoS provisioning within the RTSSM policy, a decision making process was implemented, by exploiting Simulated Annealing, Genetic Algorithms, Backtracking and Pruning algorithms. A number of several experimental tests was designed and conducted, under controlled-conditions (i.e. simulations) evaluating the performance of the above algorithms, as a matter of spectrum fragmentation and simulation time. The experimental test-bed consists of a TVWS Occupancy Repository, which keeps records about UHF/TV frequencies that can be utilized.
by LTE secondary systems. Information in this repository was built around actual/real spectrum data
gathered within the framework of the ICT-FP7 “COGEU” project (COGEU, 2012), concerning TVWS
availability between 626MHz (Ch.40) and 752MHz (Ch.60) around Munich in Germany. It should be
noted that in the simulation tests that were conducted, fixed-price mode was selected, based on a single
spectrum-unit price that was applied for every TVWS trading process.
In this context, the simulation scenario includes seven LTE secondary systems with different radio
characteristics that were simultaneously competing for the available TVWS (see figure 5) during 4
different time periods. LTE systems operate under Time-Division-Duplexing (TDD) mode, while a
different QoS level was adopted for each system, based on specific services requirements. This QoS level
was respected, by the optimization algorithms for the fixed-price mode, during radio spectrum allocation
process. Additionally, for each new simulation period (namely as Time Period in the experimental tests)
secondary systems with different QoS expectation were entering the test-bed, under a fixed schedule,
requesting access to the available (at the given Time Period) TVWS. The technical specifications of such
LTE secondary systems are presented in table 1.

Figure 5: Time Periods of simulation scenario

From table 1, it comes that there are two major types of services defined with guaranteed bit rate (GBR)
and non-guaranteed bit rate (Non-GBR). GBR services are real-time applications, such as conversational
voice and video, while Non-GBR services include P2P and Web applications. For a GBR service, a
minimum amount of bandwidth is reserved by the proposed system and the network resources provision
is guaranteed, by taking into account specific QoS requirements. GBR services should not experience
packet losses or high latency in case of network congestion. On the other hand, Non-GBR services are
provided under a best effort scheme and a maximum bit rate is not guaranteed on a per-service basis.
Based on the above-mentioned simulation scenario, four time periods were defined as depicted in figure
5.

Table 1: Technical Specifications of each Secondary System

<table>
<thead>
<tr>
<th>Secondary System</th>
<th>Services Provided</th>
<th>Bandwidth (MHz)</th>
<th>Priority/QoS Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTE 1</td>
<td>TCP-based services (GBR)</td>
<td>20</td>
<td>Medium</td>
</tr>
<tr>
<td>LTE 2</td>
<td>P2P (Non-GBR)</td>
<td>5</td>
<td>Low – Best Effort</td>
</tr>
<tr>
<td>LTE 3</td>
<td>Internet (Non-GBR)</td>
<td>20</td>
<td>Low – Best Effort</td>
</tr>
<tr>
<td>LTE 4</td>
<td>Video (GBR)</td>
<td>20</td>
<td>High</td>
</tr>
<tr>
<td>LTE 5</td>
<td>Video (GBR)</td>
<td>10</td>
<td>High</td>
</tr>
<tr>
<td>LTE 6</td>
<td>P2P (Non-GBR)</td>
<td>5</td>
<td>Low – Best Effort</td>
</tr>
<tr>
<td>LTE 7</td>
<td>Video (GBR)</td>
<td>5</td>
<td>Medium</td>
</tr>
</tbody>
</table>

Figure 6 depicts the performance evaluation results obtained in every time period for each RRM
implementation. From the upper plot, it can be verified that all algorithms provide an acceptable
fragmentation score, taking into account that: a) the value “0” represents an “un-fragmented” spectrum,
while when moving towards “1” spectrum becomes more-and-more fragmented, i.e. there exist many
blocks of unexploited frequencies. The lower plot represents a qualitative comparison among
Backtracking (with and without Pruning technique), Simulated Annealing and Genetic Algorithm, as a
matter of the duration of the simulation before obtaining the optimum solution. From this plot, it can be
observed that Simulated Annealing performs slightly better in comparison to the other algorithms,
obtaining faster the best-matching solution in a shorter simulation time.

Figure 6: Simulation results
3.3 Applicability of the proposed solutions

Towards proving the applicability of the proposed solutions, a TVWS occupancy repository was implemented, acting as a geo-location database that holds information about secondary networks operating over TVWS around Munich area in Germany (COGEU Deliverable 3.1, 2010). The main objective is to keep the track of the allocated radio spectrum, provide data to the TVWS allocation mechanism for the provision of QoS to the secondary players in the relevant frequency bands and advertise the TVWS portfolio available for trading. The technical characteristics of the deployed secondary networks, operating in TVWS are stored in the repository. Such information entities are needed, in order to evaluate current and future requests from spectrum buyers (i.e. players). The repository is divided in three time periods, two during the day, and one for the night. During the day period, specific sites are on the market, and existing Mobile Network Operators can lease TVWS to operate extra LTE-downlink carriers. An auction-based approach is used for the day period. During the night period, spectrum demand is lower than spectrum supply; therefore the available TVWS are leased with a fixed/benchmark price, for instance, to a wireless network operator that implements smart metering applications. The TVWS occupancy repository is the unit that contains information on active TVWS networks and their operational parameters. The repository carries all information required to compute mutual interference between TVWS systems. This repository also hosts the methods to manage the database and generate events or reacts on external events relevant for the management of TVWS systems.

4 JOINT RADIO RESOURCE MANAGEMENT FOR HETEROGENEOUS RADIO ACCESS TECHNOLOGIES’ EXTENSION OVER TVWS

The system operation is based on three layers/entities, as depicted in figure 7, each one denoting a significant process for the resource allocation. The layers of the system consist of the Local Resource Manager (LRM), the Spectrum Manager (SM) and the Spectrum Broker (SB).

*Figure 7: Layers of system operation*

The LRM is responsible for the disposal/assignment of spectral resources within the area of each secondary system. More specifically, LRM calculates the required bandwidth needed for each class, taking into account the radio link operation and the traffic load. Depending on the requests sent by secondary users through the LRM, the spectrum manager of each secondary system assigns to them the TVWS resources. Moreover, each spectrum manager sends information to the Spectrum Broker based on the requested bandwidth of each secondary system, the load handled, and the priority of classes. It also sends a negotiation request, in case that a secondary system requests for more bandwidth than the initial needs for bandwidth. The Spectrum Broker is responsible for conducting the spectrum allocation process, either utilizing a fixed-price or an auction-based approach, based on negotiations and requests for required bandwidth. The Spectrum Broker of the proposed CR network architecture initially advertises data regarding spectrum portions that are available to be leased to secondary systems, as well as relevant maximum allowable transmission power thresholds. This information originated from the Geo-location database, is hosted within the TVWS Occupancy Repository. Thus, the Spectrum Broker firstly advertises the spectrum-portfolio and the price-portfolio to the secondary systems, in order to be informed for the transmission characteristics and the call price of the TVWS spectrum. After this stage, bidders (i.e. secondary systems) send/define their needs/bids for the spectrum of interest, as well as the offered price, in case of auctions. Spectrum Broker collects all interests/bids and the dynamic TVWS allocation module analyzes and processes them as a matter of secondary systems technical requirements and the locally available TVWS channel characteristics (cf. figure 4). For each spectrum portion/fragment, Spectrum
Broker creates and maintains a list with interest/bids per time period, namely as auction-portfolio, in order to choose the most valuable bidder for each specific time slot, in case of auction process, or to assign TVWS to secondary systems that cause the least spectrum fragmentation, in case of fixed-price. The auction portfolio is also analyzed/elaborated by Trading Module, taking into account a spectrum-unit price or call price (e.g. cost per MHz) that is based on spectrum-auction policies. Finally, an optimized solution combining the dynamic TVWS allocation results and the Trading Module output is obtained, enabling Spectrum Broker to sell/assign TVWS frequencies to the corresponding secondary systems under the RTSSM regime/policy. In other words, Spectrum Broker is responsible for obtaining the best-matching solution, through an optimization-based process, which constitutes a NP-hard problem, thus an approximation algorithm is required in order to solve either the fixed-price or the auction-based process.

Even though optimized TVWS spectrum allocation policies have to be adopted from a business logic point of view, dynamic joint radio resource management (JRRM) techniques need also to be implemented at the Local Resource Manager (LRM) level. As shown in figure 7, each secondary system, via its Spectrum Manager module, allocates the required spectrum to its LRM entities and the latter serve all the mobile end users being in a specific geographical area. More specifically, secondary systems can be considered LTE, WiMAX, WiFi, HSPA, GSM systems etc. while at the secondary user level (cf. figure 7), each LRM is deployed in one of the heterogeneous radio access technologies’ (RATs) base stations. For example, in figure 7, the first spectrum manager entity depicted at the left hand side, could reside in an LTE system, which serves many geographical areas by its corresponding e-NodeBs. In each e-NodeB, a LRM entity is deployed mainly dealing with interference management, admission control and scheduling issues. The second spectrum manager entity depicted at the right hand side could represent a WiMAX system following the same rationale with the pre-mentioned LTE system. LRM entities, apart from implementing admission control and scheduling algorithms, send feedback to their inter-related spectrum managers, too. This context information being delivered enables the LRM entities to provide optimal QoS/QoE services to the mobile end users and the whole proposed CR network architecture to keep the network key performance indicators at acceptable levels.

Given the fact that many heterogeneous radio access technologies (RATs) coexist in a specific geographical area, available radio resources need to be jointly managed (Falowo & Chan, 2008). When radio resources are jointly managed, joint call admission control algorithms are needed for making radio access technology selection decisions. Scheduling techniques need also to follow context-aware mobile and wireless networking (CAMoWiN) principles (Makris et al., in press). In the following two sections (i.e. sections 5 and 6), a thorough overview of the state-of-the-art admission control and scheduling algorithms and techniques are presented. These JRRM techniques are complementary to the ones being adopted at the spectrum broker and spectrum manager layers and have already been discussed in the previous sections. The main differential feature of the JRRM techniques being implemented at the LRM level, is that the decision making procedures should be extensively dynamic since decision loops take place even every 1 ms (e.g. LTE-A scheduler). As discussed in section 3, other CR-related phases are also included and interconnected with RRM modules and these are network monitoring, context acquisition, TVWS carrier’s assessment, trading and price discovery modules etc. These phases can be carried out in larger time intervals (e.g. every one hour/day) as they are more business logic oriented and thus decision making procedures are made within a timeframe of several hours, days or even months (e.g. spectrum allocation policies, accounting strategies, policies for personalization services, etc). Even though self-organized networking (SON) principles are gaining ground via 3GPP LTE-A latest standardization activities (3GPP, 2011), there are a lot of things to be done by the research community until all RRM-related processes of the proposed CR network architecture become cognitive.

In the following two sections of this book chapter, we discuss on the two main components of the Local Resource Manager, namely the Admission Control and the Packet Scheduling procedures. Through a brief survey of the related literature, we identify the main features that each of these components should have, as well as potential building blocks for their implementation.
5 CONTEXT AWARE JOINT ADMISSION CONTROL & RESOURCE PARTITIONING TECHNIQUES

According to the descriptions having been made in the previous section regarding the LRM functionalities (figure 7) and specifically the joint call admission control (JCAC)-related operations (figure 8), we will further present a classification of JCAC techniques being applicable in heterogeneous wireless access networks emphasizing in emerging cognitive and context-aware enhancements recently proposed in the international literature. Based on the efficient TVWS exploitation paradigm presented in this book chapter, when radio resources are allocated to LRM entities, JCAC schemes are needed in order radio resources to be jointly managed. JCAC is a major category of JRRM and is responsible for deciding whether an incoming call/service can be accepted or not and which of the available radio access networks (RANs) is/are most suitable to accommodate the incoming call/service. The decision strategy for the admission of new connections consists of guaranteeing that the sum of the minimum rate requirements of all accepted connections does not surpass specific thresholds. This strategy guarantees that the resources available to the scheduler (cf. section 6) are sufficient to provide the QoS requirements for all accepted connections. The major factors that need joint management from a JCAC perspective are: a) multiple heterogeneous RATs (cf. section 4), b) multiple user groups with diversified priorities and needs, and c) multiple service groups with diversified QoS requirements. Furthermore, the major network key performance indicators that have to be taken into account and are the objectives of JCAC in cognitive networking are (Falowo & Chan, 2008): a) guarantee QoS/QoE requirements (data rate, delay, jitter, PER, BER) of accepted calls/services, b) minimize call blocking and dropping probabilities, c) maximize operators’ revenues, d) maximize radio resource utilization, e) maximize user satisfaction by granting additional resources beyond those required in the initial AC process, f) minimize the number of handoffs from one RAT to another, and g) uniform distribution and balancing of the total network load. From the above, it can be easily inferred that there are trade-offs when trying to satisfy some of the pre-referred objectives. As already mentioned in section 2.2, multi-objective decision making rationale can be applied in such kind of problems. In this section, we further elaborate on various partitioning techniques ending up by describing a roadmap for future JCAC implementations being applicable in next generation networking environments (e.g. 3GPP heterogeneous networks (HetNets), mobile cloud, etc). Partitioning refers to various algorithmic techniques that can be applied to a pool of resources trying to fulfill multiple and diversified objectives. In the following, complete sharing, complete partitioning, hybrid/virtual partitioning and advanced related context-aware approaches are described.

Figure 8: The main components of the Local Resource Manager

5.1 Complete Sharing and Complete Partitioning

Complete Sharing (CS) is the most trivial technique and considers one unique pool of resources, which is common for all combinations of user groups, service groups and available RATs. A new call is admitted into the system, if there are adequate resources from the unique common pool, otherwise it is rejected. That is, when the total network resources get to their limits, a new call will be blocked while a handoff call will be dropped. CS is a first-come-first-served (FCFS) non-prioritization scheme and thus adopts the simplest resource allocation policy. Its major advantages are implementation simplicity and high radio resource utilization. However, CS does not provide any QoS differentiation and thus has poor QoS performance (Fang & Zhang, 2002).

In Complete Partitioning (CP), the overall resources are partitioned into several parts according to a combination of RAN, user group and service type and a new call request is rejected, if the resource mapped to the corresponding combination is used up (Chen et al., 2006). In other words, by the term “partitioning”, we mean that a fixed capacity is allocated to each combination and thus no resources from one partition can be allocated to more than one combination. The size of each partition is defined according to a priori knowledge that the system already acquires taken from extensive past statistical measurements. From this kind of measurements, various mobility and load traffic patterns are derived and
thus a good calculation of the size of each fixed partition can be done. Whenever a radical change in JCAC-related key performance indicators is observed, a network administrator can manually calibrate various parameters assuring the system’s proper operation. The main advantage of CP is that it has good QoS performance and ensures the fairness of different priority calls, but due to the fixed partition policy, the radio resource utilization can be severely decreased. Another main drawback is that there is no elasticity/dynamicity in the size of the partitions and no intelligence is included in the JCAC process. Towards providing more intelligence in both CS and CP schemes, many algorithmic proposals have been made in the literature during the last decade. *Guard channel* is one of the earliest techniques and it proposes to reserve some extra capacity for prioritized calls (e.g. handoff calls, high-priority user/service groups, etc) by implementing a static threshold. Higher utilization can be achieved through dynamic adaptation of the threshold according to the network state by adopting the *fractional guard channel* scheme (Li et al, 2004; Niyato & Hossain, 2005). As the dynamic adaptation depends on the radio resource utilization, various acceptance probabilities can become smaller, when utilization is high and vice versa. Enhanced proposals based on fractional guard channel are *multi-threshold resource reservation* schemes, which implement multiple dynamic thresholds in order to assign different priorities to multiple combinations of service calls (Ogbonnwan & Li, 2006; Makris & Skianis, 2008). *Thinning* algorithms (Fang, 2003) follow the same rationale by supporting multiple types of services and calculating the admission probability based on the priority and the current traffic situation. Finally, in *queuing priority* schemes, when utilization reaches 100%, high-priority calls are queued and are served when some radio resources become available. In this case, the queuing delay has to be inter-related with other types of delay imposed by scheduling process (cf. section 6). The main drawback of queuing priority scheme is that is needs a lot of buffers to deal with real-time multimedia traffic. It also needs a sophisticated scheduling mechanism in order to meet the QoS requirements of delay-sensitive calls (Falowo & Chan, 2008).

5.2 Hybrid and Virtual Partitioning

The problem with CS and CP schemes is that they are not flexible enough in order to cope with all emerging JCAC challenges (e.g. multi-homing, innovative mobile services/business models, etc). Nowadays, users’ demands are not only restricted in enjoying different types of services from various heterogeneous RATs but they continuously and increasingly demand for more flexibility and elasticity in order their QoE to be enhanced. Innovative business models are also pushing towards this direction as an individual user may have more than one profile according to the mobile device he/she uses, his/her location, etc. As a result, the number of the corresponding combinations referred in section 5.1 has become very large and conventional partitioning schemes cannot handle the incurred algorithmic complexity.

*Hybrid/virtual partitioning* is a JCAC scheme, which manages to combine the advantages of CS and CP and strikes a balance between unrestricted sharing in CS and unrestricted isolation in CP. More specifically, hybrid/virtual partitioning scheme behaves like unrestricted sharing when the overall traffic is light and like complete isolation when the overall traffic is heavy. Hence, the best characteristics of CS and CP under different loadings are combined (Yao et al., 2004). The general structure of a partition in hybrid/virtual partitioning is explicitly explained in (Skoutas et al., in press). More specifically, each partition has two main parts namely the “commonly shared” and the “reserved” area. Each partition is allowed to accept “external” service calls (i.e. calls which were initially aimed to be served by other partitions). However, this has to be performed in a controlled manner in order to prevent the flooding of the partitions with external calls. While an “external” service call can be placed only at the commonly shared area, a “native” service call (i.e. a call which is mapped to be served by the specific partition) can be placed in any of the defined areas. By considering a dynamically changing reservation factor, the “commonly shared” area of the partition can be obtained by subtracting the “reserved” area capacity from the total capacity of the partition. As a result, as the reservation factor increases, the “commonly shared” area decreases thus accepting fewer “external” service calls and vice versa. It has to be noted that an accompanied preemption scheme is needed in order the robustness of the overall hybrid/virtual
partitioning scheme to be supported. Hence, the main drawback is that the preemption scheme may lead to lower utilization.

Towards confronting this drawback, many supplementary algorithmic proposals have been made. Spillover-partitioning algorithms are proposed in (Yilmaz et al., 2010), where utilization is improved by sharing certain partitions among different service calls. QoS degradation (Niyato & Hossain, 2005; Chou & Shin, 2004; Wan et al., 2005) is also a well-known technique and is used in situations of network congestion. For example, when the network becomes congested, the amount of bandwidth allocated to some of the ongoing calls (also called degradable calls) is revoked to accommodate more incoming calls so that call dropping/blocking probabilities can be maintained at the target level without affecting resource utilization maximization targets. In case of light traffic load, some revenue maximization algorithms have been proposed such as (Chen et al., 2006). In order to increase the resources utilization, some calls (also called upgradable) can be allocated with more resources. A typical example is web browsing or file downloading. In this scenario, the mobile user can enjoy better QoS (by first giving his consent for being excessively charged) and at the same time the operator can increase its revenues.

5.3 Advanced Context-Aware Approaches

As previously described, hybrid/virtual partitioning family of JCAC techniques seems able to adequately and simultaneously satisfy most of the JCAC-related objectives in cognitive networking (Falowo & Chan, 2008). However, research community envisions even more challenges regarding the future-networking continuum, which need to be addressed. For example, JCAC-related architectural innovations proposed by various IEEE standards like P1900.4 (IEEE P1900.4, 2008) and 802.21 (IEEE Standard 802.21, 2009) stress the need for dealing with distributed radio resource usage optimization issues from an overall system perspective. Radio resource management in LTE-Advanced networks including related emerging challenges in HetNets (Lopez-Perez et al., 2011), machine-to-machine communications (M2M) (Zheng et al., 2012), device-to-device communications (D2D) (Yu et al., 2011) and cooperative communications (Elkourdi & Simeone, 2011) are also fields of research that are continuously gaining ground. Finally, novel JCAC principles have to be stressed for mobile cloud computing environments as the idea of integrating cloud computing into heterogeneous mobile and wireless networking is promising, too (Makris et al., 2012; Hoang et al., 2012).

Regarding the afore-mentioned JCAC-related architectural innovations, the main objective is to define an appropriate system architecture and protocols, which will facilitate the optimization of radio resource usage by exploiting context information exchanged between network and mobile terminals, regardless of their support for multiple simultaneous links and dynamic spectrum access. The “Distributed Radio Resource Usage Optimization” use case introduced in (IEEE P1900.4, 2008), contains many cognitive and context-aware JCAC-related building blocks, while reconfiguration and self-management features play a critical role, too.

Regarding LTE-Advanced innovations, the main breakthrough lies in the fact that the case of mobile terminals (MTs) being directly (i.e. via one hop) connected to base stations (BSs) of heterogeneous RATs in order to acquire access to services is not a panacea. In fact, many wireless/wired nodes can be relays of information, while small base stations (e.g. femtocells) can operate as relays, too (Elkourdi & Simeone, 2011). Moreover, the concepts of M2M and D2D communications introduce the idea of MTs communicating directly with each other over M2M/D2D links, while remaining control under BSs. Due to this potential, location-aware and geo-referenced services can be developed and thus novel JCAC design has to be adopted.

In mobile cloud computing/networking, the main novelty feature, which has to be stressed is that a JCAC framework has to simultaneously take into account both: a) wireless/radio access resources pool and b) computing resources pool for data processing/storage aiming at flexible virtualized infrastructure sharing solutions. That is, there is no sense in allocating only networking resources to MTs, because there may not be corresponding sufficient computing resources to support the ongoing calls/services. Finally, joint design and optimization of access and backhaul is needed and hence JCAC modules have to be accordingly enhanced.
6 CONTEXT AWARE SCHEDULING TECHNIQUES

Following the centralized spectrum sharing model described at section 4, the spectrum distribution to the secondary systems is based on a dynamic negotiation procedure, which involves the Local Resource Manager, the Spectrum Manager and the Spectrum Broker. Thus, the spectrum available to the end users of a secondary system is periodically adapted to their requirements in a cognitive manner by utilizing either a fixed-price or an auction-based approach. As discussed at the previous section, within each secondary system a CAC scheme should be employed in order to maintain the traffic load within a manageable range, while a packet scheduling mechanism is responsible for the efficient sharing of the available capacity among the ongoing traffic flows.

This centralized dynamic spectrum approach can only be used with infrastructure based wireless networks, where a central controller is able to coordinate the spectrum access of the secondary users. Consequently, the most likely candidates for the exploitation of TVWS under this scheme are Mobile Broadband Wireless Networks (MBWN) such as LTE and Mobile WiMAX which, despite their diverse origin, are slowly converging towards the same design goals. Regarding the packet transmission process, MBWNs have adopted similar features, such as Fast Scheduling, Hybrid ARQ (HARQ) and Adaptive Modulation and Coding (AMC) and as a result, it is feasible to identify a theoretical scheduling framework with wide applicability (Skoutas & Rouskas, 2010).

The main trade-off when designing a context aware scheduling mechanism is between the speed and accuracy of decision making. In general, the scheduler’s computational complexity and the optimality of the scheduling decisions are both reduced when decreasing the amount of the information that is required for each scheduling decision. While low computation complexity is essential for supporting the fast scheduling feature of mobile broadband and wireless networks (MBWNs), it is expected that future context aware scheduling schemes will be required to process increasingly more data. In order to circumvent this problem we have to define heuristic scheduling disciplines which will help us to avoid, to the extent possible, the use of computationally expensive, exhaustive search algorithms (Skoutas & Rouskas, 2009).

Another important trade-off that should be considered when designing a wireless scheduler is the one between the efficient utilization of the available bandwidth and the fair sharing of resources among the users. In wireless networks it is always more efficient, in terms of system's throughput, to share the bandwidth only among users with good channel conditions. However, this policy can be proven to be highly unfair for users that experience low SINR (Signal to Interference plus Noise Ratio) for extended time periods. Therefore, the scheduler should be able to maximize the utilization of the bandwidth and at the same time to maintain fairness among all users by utilizing compensation mechanisms (Capozzi et al., in press).

Furthermore, the scheduler should be capable of providing different QoS levels to the current and future applications and services. This feature is very important as we move towards all-IP network architectures, which, inherently, are not well suited for QoS demanding services. The QoS is usually expressed as a set of minimum or maximum allowable values for specific metrics such as end-to-end latency, jitter, Bit Error Rate (BER) and minimum guaranteed bit rate. With the advent of cognitive networks, more network metrics are made available, enabling thus the scheduling mechanism to provide personalized services and take into account constraints such as energy consumption. In the following, we will briefly discuss the evolution of packet scheduling from a simple fair capacity sharing problem in the first network deployments to a multi-parameter optimization problem in current context aware wireless networks.

6.1 Fundamental Scheduling Disciplines

The first scheduling schemes were designed for low QoS demanding services, which were transmitted over wireline networks. Therefore, the goal they had to accomplish was the equal sharing of the available capacity among the end users. The Round Robin (RR) approach, which aims to assign the shared channel for equal amount of time to each traffic flow is a typical example of this category of schedulers. RR can
be combined with a Resource Preemption (RP) mechanism in order to achieve QoS differentiation between low priority and high priority traffic flows (Capozzi et al., in press).

Generalized Processor Sharing (Parekh & Gallager, 1993) is a more advanced scheduling discipline, which aims to provide differentiated allocation of the available capacity among the ongoing traffic flows. According to the GPS discipline, each traffic flow receives at each scheduling period a fraction of the total capacity, which is proportional to a positive real number (weight) assigned to the flow. While GPS is more efficient than RR and RP, it cannot be applied to real networks as it assumes that the transmitted packet traffic is infinitely divisible. The Weighted Round Robin (WRR) scheduler (Fattah & Leung, 2002) is a simple and realistic approximation of the GPS scheduling discipline. According to WRR, the number of transmitted packets from each non-empty queue is proportional to the respective weight of the queue. However, the use of WRR can be problematic when the size of the data packets is variable. A better approximation of GPS for packet transmission is Packet by Packet GPS (PGPS), also known as Weighted Fair Queuing (WFQ) (Fattah & Leung, 2002), which emulates the GPS operation. Specifically, PGPS emulates a hypothetical GPS server and on each packet arrival, a virtual service time is calculated according to the GPS discipline. Then, the packets from different flows are sequenced and arranged for transmission in increasing order of their virtual departure times. The Earliest Due Date (EDD) scheduler (Capozzi et al., in press) is a scheme able to provide QoS differentiation based on packet delay. EDD associates each incoming packet with a deadline and then it serves the packets following an increasing order of their respective deadlines. Shortest Time to Extinction (STE) (Panwar et al., 1988) is a scheduling discipline similar to EDD, which discards the packets that exceed their deadlines instead of keeping them in the queue.

The abovementioned scheduling schemes are able to provide a basic form of QoS differentiation together with a worst-case delay guarantee in the case of leaky bucket constrained sources. However, as they are designed for wireline networks, they are not aware of the variations of the wireless channel capacity and therefore they cannot be directly applied to wireless networks.

6.2 Wireless Channel Aware Scheduling Schemes

Moving in the era of wireless networks, the scheduler designers tried to transfer the wireline scheduling disciplines in the wireless environment. Thus, for example, the GPS discipline was evolved towards this direction, producing a class of Wireless Fair Queuing (WiFQ) schemes such as Channel condition Independent Fair Queuing algorithm (CIF-Q), Idealized Wireless Fair Queuing algorithm (IWQ) and Server Based Fairness Approach (SBFA) (Fattah & Leung, 2002). The main idea behind all these schemes is the same; a GPS based scheduler is emulated in an error free environment and used as a reference model for the actual rate allocation. A compensation mechanism ensures that flows, which are lagging in comparison to their error-free service model will be compensated in the subsequent scheduling periods.

Credit based (CB) schemes (Kam et al., 2001) are also developed for the sharing of a wireless channel and their scheduling discipline resembles, in a sense, to that of Wireless Fair Queuing. The CB schemes assign to each ongoing traffic flow a guaranteed transmission rate, which, as in WFQ, is a fraction of the available capacity. The service priority of the flows is then defined based on a priority variable, which is called the credit. The credit of each flow is calculated as the difference between the amount of data that should have been transmitted (i.e. error free service model), and the actual amount of data that has been transmitted so far from the specific flow. All the schemes that are following rate-based disciplines, such as CB and WFQ, are able to provide data rate guarantees to the ongoing connections as long as a perfect power control is assumed.

Following a similar evolutionary path as GPS, the earliest due date (EDD) scheduling discipline has been introduced to the wireless environment through the Feasible-EDD (FEDD) scheme (Shakkottai & Srikant, 2002). As a variant of EDD, FEDD transmits at each scheduling period the packets with the lowest deadlines; however, in contrast to EDD, it does not consider all the ongoing traffic streams but only those with good channel conditions. A Proactive variant of EDD (PEDD) (The et al., 2003) takes
into account expected future changes in the state of the wireless channel and adjusts accordingly the packet’s deadline. A main disadvantage of both FEDD and PEDD is that they assume ideal knowledge of the channel conditions. Thus, Realistic PEDD (R-PEDD) and R-PEDD+ (Kong & Teh, 2004) pointed out the need for separate probing mechanisms for the acquisition of the required channel information.

The scheduling schemes of this category are adapted to the wireless environment and are able to provide QoS differentiation. Nevertheless, they are still missing important features and they cannot be directly applied to the current and future mobile wireless broadband networks.

6.3 Context Aware Scheduling Schemes

Since the advent of MBWNs, applications and services were classified into QoS classes according to the delay sensitivity of the corresponding traffic, and hence it has become apparent that future packet schedulers should be able to support delay based QoS differentiation. Another fundamental feature of MBWNs is fast scheduling, which dictates that the scheduling decisions should be performed in a fast and accurate manner. Moreover, a scheduler should be aware of the Channel State Information (CSI) in order to able to select at each transmission interval the proper modulation scheme that offers the required Bit Error Rate (BER). Hence, a modulation scheme with reduced constellation can be used over noisy channels while a higher constellation modulation scheme can be used when the channel conditions are good, providing thus increased data rates. The conclusion that emerges from the above is that the early wireless scheduling schemes discussed previously, are not suitable for the current and future MBWNs. Rate based algorithms such as WiFQ cannot be used in MBWNs because they are not able to provide delay based QoS differentiation while others which can provide such kind of QoS differentiation (i.e. FEDD, PEDD and R-PEDD) are too complex and their use can be problematic.

A scheduling algorithm with Dynamic Priority Assignment (DPA), proposed in (Skoutas & Rouskas, 2009) is one of the first approaches that aimed to combine all the required characteristics of a MBWN scheduler and at the same time to provide a deterministic delay bound to each connection. DPA is aware: a) of the available shared capacity, b) the number of queued packets at each queue, c) the state of the wireless path and d) the available transmission power. DPA utilizes the processing gain of CDMA networks instead of AMC in order to achieve the required BER. Dynamic Hybrid Scheduler (DHS) (Skoutas & Rouskas, 2010) is a more generalized scheduling approach that utilizes AMC and extends the DPA discipline so that it can be applied to the most of the currently evolving broadband wireless technologies.

More recent approaches (Piro et al., 2011; Esmailpour & Nasser, 2011) exploit the concept of combining scheduling disciplines in a two level scheduling process which offers a more accurate and effective handling of the shared capacity. Thus, the desired QoS can be preserved together with a better utilization of the available resources. Furthermore, one could additionally consider multi-cell scheduling aiming to achieve efficient resource allocation and inter-cell interference mitigation in multi-cell environments. At (Pateromichelakis et al., in press), the evolution of interference management techniques is studied and their common features and differences are discussed. Scheduling in relay assisted MBWNs is also a topic that becomes to draw the attention of the scientific community. Future scheduling schemes should be aware of the characteristics of the upcoming relay enhanced cellular networks. Centralized schemes (Salem et al., 2010) as well as fully distributed scheduling algorithms (Suzhi & Zhang, 2012) have been recently presented in the related literature aiming to reduce outage probability and preserve user fairness, which is crucial in such environments. Finally, one should also consider the need for more energy-efficient wireless communications. At (Feng-Seng et al., 2012), a scheme that reduces the energy consumption at the mobile terminal by scheduling its transmissions into fewer time slots is proposed while at (Wang et al., 2012) a number of techniques that can be employed in future green mobile networks are presented.
7 FUTURE RESEARCH DIRECTIONS

This section discusses emerging research trends providing insightful considerations about the future of cognitive networks and self-adaptive communication systems from the perspective of efficient TVWS exploitation paradigm. Several related ongoing projects co-funded by the EU Commission pave the way for new research innovations in the future. Apart from COGEU project’s concepts (COGEU, 2012), which are presented in this book chapter, QUASAR (QUASAR, 2012), QoSMOS (QoSMOS, 2012) and SACRA (SACRA, 2012) projects provide future roadmaps regarding JRRM challenges for TVWS exploitation paradigm. More specifically, COGEU project proposes: a) JRRM techniques to optimize spectrum utilization, minimize interference, guarantee fairness in TVWS access and integrate QoS aspects in dynamic spectrum management, b) ways that reliable data delivery can be realized by routing and transport protocols across regions of different spectrum availability, and c) protocols that will allow system players to efficiently negotiate spectrum information parameters with a centralized spectrum broker. QUASAR provides specific and reasoned proposals to go beyond the current regulatory frameworks defined in specifications of various federated organizations such as FCC (Federal Communications Commission) in USA and CEPT-SE (Conference of Postal and Telecommunications Administrations Spectrum Engineering) in Europe, while many national regulators from the globe such as ACMA (Australia), KCC (Korea), Ofcom (UK), iDEA (Singapore) can exploit the project’s recent research outcomes, too. Regarding JRRM concepts, QUASAR proposes: a) a cognitive management architecture to implement decision-making processes as well as to support mobility and QoS provisioning at the radio access/link level and b) a cognitive spectrum management framework accompanied by corresponding RRM algorithms to optimize spectrum utilization. QoSMOS project main focus is to make use of technology and service neutral spectrum opportunistically. It proposes an overall TVWS exploitation framework applicable to both centralized and distributed CR architectures including: a) a cognitive manager for resource management (CM-RM) to manage the problem of efficiently enforcing QoS for coexisting heterogeneous wireless networks with intermittently available spectrum resources and b) a cognitive manager for spectrum management (CM-SM). SACRA project mainly focuses on physical-layer studies by designing hardware components to support the CR approach viewing the TVWS exploitation problem from an energy-efficient perspective, too. It proposes a cognitive RRM inner and outer loop architecture, which enables collection of key information from bands sensed by the SACRA sensing system to make decisions on data partitioning across spectrum bands and on resource allocation to secondary users within TVWS bands. JRRM rationale is also adopted to determine the amount of resources to allocate to each link via optimization algorithms enhanced by prediction and learning techniques.

In addition to the above-mentioned short-term future research directions, there are some longer-term future research insights incurred by the evolution of cognitive networks and TVWS exploitation paradigm. In (Wu S. H. et al., 2012), a Cognitive Radio Cloud Networking (CRCN) model is proposed that is able to support CR access in TVWS. Making use of the flexible and vast computing capacity of the cloud, the proposed cloud infrastructure has virtually unlimited resources to collect, analyze, process, and coordinate the massive and dynamic CR communications activities. Moreover, a primitive CRCN prototype is presented, which integrates functions of cooperative spectrum sensing (CSS), dynamic spectrum access (DSA), mobility management and QoS provisioning on a unified cloud platform. Conclusively, scalable cloud-based CSS and JRRM schemes remain crucial to justify the feasibility of the envisioned CRCN concept in the support of large-scale public access.

Extensive research efforts are also expected to take place in the field of worldwide trends in regulation of secondary access to TVWS using CR techniques, as indicated in (Nekovee M. et al., 2012). Some of these regulatory trends are: a) elaboration on consensus points regarding CEPT SE43 and FCC rules, b) geo-location databases versus cooperative spectrum sensing trade-off investigations, c) alternative options for secondary licensing, d) aggregate interference management in TVWS, and e) generalize TVWS exploitation paradigm to other potential candidates such as military and radar bands. In any case, all these regulation mentality shift needs to be jointly addressed by regulators, industry and academia.
Promising applicability area candidates for TVWS exploitation can also be machine-to-machine (M2M) and rural broadband communications as the state-of-the-art experience has indicated that white spaces are better suited to deploying new network infrastructures and not peer-to-peer communications (Webb W., 2012). In (Yan Z., 2012), the use of CR technology in M2M communications from different point of views, including technical, applications, industry support, and standardization perspectives is motivated. Cognitive M2M system coexistence in TVWS is a new research challenge, which needs careful design to ensure fair and efficient sharing among heterogeneous users, while JRRM concepts have to be effectively mapped to new network system design prerequisites.

8 CONCLUSION

Nowadays, cognitive radio is being intensively researched for proper access to the TV White Spaces, which become available on a geographical basis after the gradual switch-off of analogue TV and adoption of digital TV. Due to the excellent propagation conditions of the released UHF/VHF band, efficient TVWS exploitation paradigm is seen as an opportunity for new services and business. In the context of FP7 ICT COGEU project (COGEU, 2012), the general idea is to move away from the binary choice of optimizing current spectrum (not always possible) or buy new spectrum with exclusive rights (too costly) by including a third option, which is the secondary use of TVWS (e.g. new spectrum commons, real-time secondary spectrum market and auction-based market). In this book chapter, a centralized CR topology is proposed with a spectrum broker trading with various secondary systems. Various spectrum sharing models and TVWS allocation policies are also investigated. Apart from our proposed architectural innovations, we further focused on emerging JRRM challenges. Indeed, with the availability of TVWS and their temporary lease, the traditional concepts of network planning and RRM have to be significantly enhanced. More specifically, as TVWS extend the pool of available radio resources for every heterogeneous RAT, it is necessary to manage them on the general context of JRRM (i.e. new TVWS carriers and legacy carriers from each heterogeneous RAT). We emphasized on context-aware JCAC and joint scheduling techniques by providing a comprehensive classification of past, state-of-the-art and emerging algorithmic proposals having been made in the international literature. Means of advanced context-aware and cognitive approaches are also introduced regarding the applicability of novel JRRM design and development in emerging research fields, such as LTE-Advanced systems, cooperative communications and mobile cloud computing/networking environments. Hence, by reading this book chapter, researchers from both academia and industry can effectively identify technical challenges regarding their ongoing/future work on JRRM issues in a broad range of cognitive and context-aware mobile and wireless networking area.

ACKNOWLEDGMENTS

The work presented in this paper has been undertaken in the context of the project COGEU (Cognitive Radio Systems for Efficient Sharing of TV White Spaces in European Context). COGEU is a Specific Targeted Research Project (STREP) supported by the European 7th Framework Programme, Contract number ICT-248560, Project duration 1st January 2010 to 31st December 2012 (36 months). The authors would like to acknowledge the contributions of their colleagues from the COGEU consortium.

REFERENCES


FP7-ICT-249060 SACRA Project http://www.ict-sacra.eu/


ADDITIONAL READING SECTION


KEY TERMS AND DEFINITIONS

TV white spaces (TVWS). TVWS are the unexploited portions of radio spectrum in the TV bands, such as the guard bands between broadcasting channels and channels freed up, by the transition from analogue to digital terrestrial television.

Joint Radio resource management (JRRM). RRM is the system level control of co-channel interference and other radio transmission characteristics in wireless communication systems. RRM involves strategies and algorithms for controlling parameters, such as transmission power, data rates, handover criteria, modulation scheme and error coding scheme. Joint Radio Resource Management (JRRM) algorithms define ways of achieving an efficient usage of a joint pool of resources belonging to
several radio access networks.

**Dynamic spectrum management (DSM).** DSM also referred to as dynamic spectrum access (DSA), is a set of techniques based on theoretical concepts in network information theory and game theory that is being researched to improve the performance of a wireless communication network and efficiently exploit radio spectrum resources.

**Real Time Secondary Spectrum Markets (RTSSM).** RTSSM policy used in cognitive radio systems adopts spectrum trading, which allows primary users to sell/lease spectrum usage rights and secondary players to buy them, thereby establishing a secondary market for spectrum leasing and spectrum auction.

**Context awareness (CA).** CA refers, in general, to the ability of computing systems to acquire and reason about the context information and adapt the corresponding applications accordingly.

**Joint Call Admission Control (JCAC) algorithms.** JCAC algorithms are one subset of JRRM algorithms, which decide whether an incoming call can be accepted or not in a wireless access network. They also decide which of the available radio access networks is most suitable to accommodate the incoming call.

**Fragmentation factor.** Fragmentation factor defines the optimal spectrum usage over time in order to avoid that TVWS will be divided into discrete fragments.