Techniques for Design and Analysis of QoS-based Models in Partially Observable Environments

Contents

Unlike wired networks, in which the data transmission is isolated from interaction with other transmissions, in wireless networks, the medium is shared between all devices that are in the same transmission range. To overcome the interference between wireless devices, wireless networking technology has become an active research area in the last decade. Wireless networks are increasingly used with the advent of standards such as WiFi, WiMAX, Bluetooth and UMTS. There is no doubt that the next-generation wireless technologies promise higher levels of complexity.

This chapter is devoted to introduce the CR architecture and the congestion control, and to define some basic theoretical concepts, which will be used in the following chapters. The remaining sections of the chapter are structured as follows: In the next section, we present CR networks and their practical implementation. We introduce, in Section 2.2, the congestion control for wireless networks. Section 2.3 provides some insight about the decision theory, and we describe some basics of the queueing theory in Section 2.4. Section 2.5 introduces the game theory, and Section 2.6 introduces learning algorithms. We present some application of game theory, self-adaptivity and learning for wireless networks in Section 2.7. Finally, Section 2.8 concludes the chapter.

2.1 CR networks

There is a general agreement that traditional fixed spectrum allocation can be very inefficient, considering that most of the time, bandwidth that was allocated is not used and the corresponding channel is idle, which form *spectrum holes*. Although the unlicensed access to the spectrum achieves better utilization of the spectrum by using spectrum holes, (see Figure 2.1) it introduces new challenges such as: the identification of spectrum holes, the competition between SUs, etc. Note that the design of CR networks involves several disciplines, such as decision theory, queueing analysis and game theory.

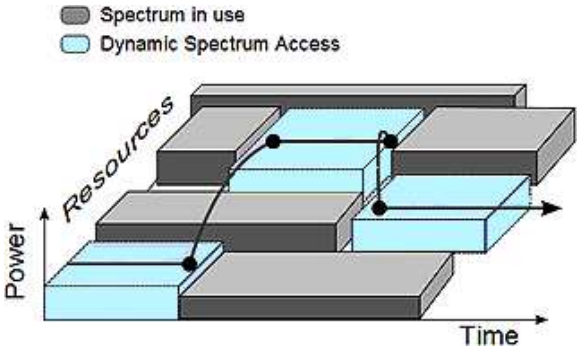

FIGURE 2.1: Wireless spectrum holes.

Furthermore, many studies showed that while some frequency bands in the spectrum are heavily used, other bands are largely unused. Note that most of the available radio spectrum was already allocated to existing wireless systems. Thus, the importance of CR paradigm aroused for allocating valuable wireless resources. The term cognition is described as the faculty of a mobile or a network to adapt its communication parameters (transmission power for mobiles or frequency for a base station) to perturbations of its environment. For instance, Ian F. Akyildiz et al. defined CR in [7] as follows:

"*A "Cognitive Radio" is a radio that can change its transmitter parameters based on interaction with the environment in which it operates*".

A big new challenge in the networking community is how to put *cognition* into networks. A radio system having this capability is called a CR, which generally uses the Software-defined Radio (SDR) technology. In fact, CR users are equipped with an SDR in order to sense and access the licensed spectrum. The SDR is considered to be the key technology that allows mobile devices to implement CR in practice. Both concepts SDR and CR are introduced in order to enhance the efficiency of the spectrum utilization in wireless systems. An SDR is defined as a reconfigurable wireless communication system that tunes dynamically its transmission parameters, such as operating frequency bands, modulation mode and transmission protocol. This adjustability can be achieved by software-controlled signal processing algorithms. The main functions of an SDR are:

- *Multi-band operation*: the ability to transmit over different frequency spectrums (cellular bands, TV bands, etc.).
- *Multi-standard support*: the ability to support different standards (GSM, WiMAX, WiFi, etc.), and different interfaces within the same standard (e.g. 802.11a, 802.11b, 802.11g in the WiFi standard).
- *Multi-service support*: the ability to support multiple types of services (3G, broadband wireless Internet, etc.).
- *Multi-channel support*: the ability to transmit and to receive over multiple frequency bands simultaneously.

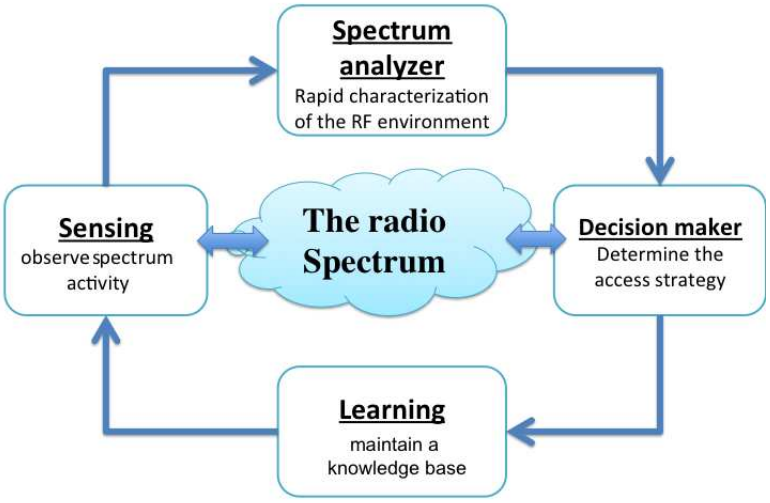

FIGURE 2.2: Components of a CR user.

A CR is aware of its environment, the internal state and predefined objectives, and looks for channel occupancy, modulation, etc., in order to make decision about its behavior.

For instance, a CR user may use SDR, so that it can reconfigure itself in order to optimize its transmission parameters. We illustrate, in Figure 2.2, the architecture of a CR node. The different components of a CR user are defined as follows:

- *An SDR-based wireless transceiver* that observes the activity of the frequency spectrum, and changes dynamically its transmission parameters.
- *A spectrum analyzer* that uses measured signals to analyze the spectrum utilization and ensure that the transmission over the spectrum is not interfered with PUs. Various signal processing techniques can be used in order to infer the spectrum usage information.
- *A decision maker* that defines the spectrum access strategy based on knowledge of the spectrum utilization. The optimal decision depends on the PUs' behavior, as well as the competitive or cooperative behavior of SUs. Different techniques, such as optimization theory, game theory and stochastic optimization, can be used in order to obtain an optimal solution.
- *A learning and knowledge extraction mechanism* that uses information of spectrum usage to understand the RF environment, i.e. the behavior of PUs. CR users maintain a knowledge base in order to adapt their transmission parameters and achieve the desired objective.

The new spectrum-licensing paradigm, initiated by the FCC in [8], promoted the idea of using the CR technology in order to face the spectrum scarcity problem. The new spectrum licensing allows unlicensed users to access the spectrum as long as they do not harm PUs, which can be achieved by spectrum sensing or power control. With the development of the CR technology, Dynamic Spectrum Access (DSA) and OSA become promising approaches that achieve major gains in the efficiency of spectrum utilization, and solving the spectrum scarcity problem. The design of DSA involves academia and industry, as well as spectrum policy makers to deal with both technical consideration and regulatory requirements. Furthermore, the development of DSA requires multidisciplinary knowledge, such as wireless communications, signal processing, optimization, artificial intelligence, decision theory, etc. For example, the competition and the cooperation between SUs accessing the same licensed bands can be modeled using game theory and utility-based techniques.

Game theory seems an ideal mathematical tool for evaluating the performance of communication systems. Since licensed channels have been opened for the unlicensed use, several works have focused on the interaction between SUs. Note that SUs may compete or cooperate with each other when accessing the spectrum. The competitive and the

TABLE 2.1: Standards for CR Aspects

Aspects	**Covering Standard Bodies**
Definition	IEEE Dyspan, ETSI, ITU-R.
Coexistence	IEEE 802.19, IEEE Dyspan.
SDR	IEEE Dyspan, SDR forum, ITU-R, OMG.
Radio Interfaces	IEEE 802.22, 3GPP.
Heterogeneous Access	ESTI, IEEE Dyspan.
Spectrum Sensing	IEEE 802.22, IEEE Dyspan.

cooperative behavior of SUs was depicted in [12], [13], [14], [15], [16] and [17]. For example, authors of [18] proposed a game theoretic framework to analyze the behavior of cognitive radios for distributed adaptive channel allocation. They defined two different objective functions for the spectrum sharing games, which capture the utility of selfish users and cooperative users, respectively. Based on the utility definition for cooperative users, they showed that the channel allocation problem can be formulated as a potential game, and thus converges to a deterministic channel allocation Nash equilibrium point. The survey paper [19] presented some some application of game theory. The survey outlines research challenges and future directions in game theoretic modeling approach in CR networks.

The potential of CR users has been recently identified by various policy [8] and [20], research [21], standardization [22], [23], and [24], and commercial organizations. The IEEE 1900 Standards Committee on Next Generation Radio and Spectrum Management was established in 2005 and jointly supported by the IEEE Communications Society (ComSoc) and the IEEE Electromagnetic Compatibility (EMC) Society. The concern of IEEE 1900 is to address key standardization issues in the emerging fields of spectrum management and advanced radio system technologies such as CR, SDR, and adaptive radio systems. Tables 2.1 and 2.2 give some standards for the CR technology. The paper [25] and references therein provide an extensive study of standards in the CR field.

The licensed spectrum can be utilized by SUs through either OSA or Dynamic Spectrum Sharing (DSS). In the first approach, SUs access licensed channels only when PUs are not using them. Using the DSS, SUs are allowed to use simultaneously the spectrum with PUs, as long as their transmissions do not cause harmful interferences with PUs.

The main challenge for CR networks is to locate *spectrum holes* and distribute them efficiently. The surveys [7], [26] and [27] provide a summary about recent works and design issues in CR networks.

TABLE 2.2: IEEE Dyspan Working Groups

IEEE 1900.1	Terminology and concepts for next generation radio systems and spectrum management.
IEEE 1900.2	Interference and coexistence analysis.
IEEE 1900.3	Conformance evaluation of SDR software modules.
IEEE 1900.4	Architectural building blocks enabling network device distributed decision-making in heterogeneous wireless access networks.
IEEE 1900.5	Policy language and policy architectures for managing CR, and for DSA applications.
IEEE 1900.6	Spectrum sensing interfaces and data structures for dynamic spectrum access and other advanced radio communication systems.
IEEE P1900.7	Radio interface for white space dynamic spectrum access radio systems supporting fixed and mobile operation.
IEEE 802.22	Wireless communication at 54-863 MHz. It has an arrangement related to the identification of PUs and defining power levels so as not to interfere with adjacent bands. It is targeting at using CR techniques to allow sharing of the TV spectrum with broadcast service.

2.2 Congestion control in wireless networks

With the increase of the heterogeneity and the complexity of the Internet, the standard TCP congestion control mechanism becomes inefficient (see [28] and [29] for example). The main reasons, for this inefficiency, is that congestion signals are only indicated by packet loss, and TCP uses fixed Additive Increase Multiplicative Decrease (AIMD) algorithm to adapt the congestion window size. Nevertheless, the window size should be changed according to the network environment and the media content. Note that physical impairments of the wireless transmission medium increase the complexity of designing a media-aware congestion control for wireless environments.

Despite of the success of TCP, the existing congestion control is considered unsuitable for delay-sensitive, bandwidth-intense, and loss-tolerant multimedia applications, such as real-time audio streaming, video-conferences etc. (see [9] and [11]). There are three main reasons for this:

- First, TCP is error-free and trades transmission delay for reliability. In fact, packets may be lost during transport due to network congestion and physical impairments. TCP keeps retransmitting them until they are transmitted successfully, even with a large delay. Note that although multimedia packets are successfully received, they are not decodable if they are received after their respective delay deadlines.

- Secondly, TCP congestion control adopts an AIMD algorithm, which linearly increases its congestion window size per Round-Trip Time (RTT) when there is no packet loss, and multiplicatively decreases the congestion window size when packet loss occurs. This results in a fluctuating TCP throughput over time, which significantly increases the end-to-end packet delay, and leads to worse performances for multimedia applications [11].
- Finally, standard TCP congestion control is based on network performance metrics (namely QoS metrics) and not on a subjective metric of the quality perceived by the user (measured through the QoE). In wireless systems, where the environment has an important impact on the quality of multimedia applications, a QoE-based congestion control for TCP is welcome.

Some variant of TCP was proposed, such as TCP Vegas [30] and FAST TCP [31], using the RTT values for the congestion indication. Note that the RTT usually increase before packet losses occur when the network is congested. FAST TCP is developed at the Netlab, California Institute of Technology and now being commercialized by FastSoft. It is compatible with existing TCP algorithms, requiring modification only to the computer which is sending data.

The key idea of designing a wireless TCP is to distinguish the cause of packet loss [28]. Many schemes are proposed in the literature. For example, TCP Veno [32] estimates the backlogged packet in the buffer of the bottleneck link, as illustrated in Algorithm 1. It determines the optimal throughput the network can accommodate based on the minimal RTT, denoted $BaseRTT$. The difference between the optimal throughput and the actual throughput can be used to derive the amount of backlogged packets in the queue of the bottleneck link. TCP Veno suggests that the loss is said to be random if the number of backlogged data is below a threshold β, and congestive otherwise.

Algorithm 1 TCP Veno Algorithm: distinguish the cause of packet loss [32]

when packet loss is detected by fast retransmit:
if $(DIFF < \beta)$ **then**
  $ssthresh = cwndloss \times (4/5)$;
  //where $DIFF = (cwnd/BaseRTT - cwnd/RTT) \times BaseRTT$
  //random loss (due to bit errors) is most likely to have occurred
else
  $ssthresh = cwndloss/2$;
  // congestive state is most likely to have occurred,
  //even there occurs random loss at this time
end if
when packet loss is detected by retransmit-timeout timer:
ssthresh is set to half of the current window ;
slow start is performed; // performs the same action as in Reno

2.3 Decision-making models

Whether we make it consciously or not, every day we make several decisions. Frequently, it is not trivial to make the right decision for some problems. Usually, decisions we take have not only immediate results or outcomes, but impact also our future decisions. Unless we take into account both present and future impact of our decisions, we may not achieve good overall performances. We study, in the following section, a decision model, useful for studying a wide range of multi-stage optimization problems.

2.3.1 Markov decision process

We focus, in this section, on the sequential decision model, Markov decision process (MDP), where the decision maker, usually called agent or controller, makes decisions sequentially. We denote by decision epoch, every time the agent has to make a decision. At every decision epoch, the agent observes the state of the system and chooses an action. Choosing an action in a given state has mainly two results: the agent receives a reward, and the system evolves to a possibly different state at the next decision epoch. We formulate an MDP problem as follows:

- *Decision epochs*: Denote by $\mathcal{T}$ the set of decision epochs. If this set is finite, the decision problem is said to be *finite horizon* problem, otherwise it is called an *infinite horizon* problem.
- *States*: At every decision epoch, the system occupies a state $s(t)$. $\mathcal{S}$ denotes the set of all possible states.
- *Actions*: We denote the set of actions for each state s by $\mathcal{A}_s$, and the set of all possible actions is referred to as $\mathcal{A} = \cup_{s\in\mathcal{S}}\mathcal{A}_s$.
- *Immediate reward* $\mathcal{R} : \mathcal{S} \times \mathcal{A} \to I\!R$: We denote by $r_t(s, a)$, defined for state $s \in \mathcal{S}$ and action $a \in \mathcal{A}_s$, the real-valued function that assigns, for a given decision epoch t, a value as outcome for taking the action a in the state s. If $r_t(s, a)$ is positive, it is called reward function. Otherwise, it is called cost function.
- *Transition probabilities* $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$: When the agent takes the action a in the state s, the system state in the next decision epoch is determined by transition probabilities $p_t(.|s, a)$. We usually assume that $\sum_{j\in\mathcal{S}} p_t(j|s, a) = 1$.
- *Decision rules*: Decision rules are functions $d_t : \mathcal{S} \to \mathcal{A}$, which specify the action choice when the system is in the state s at the decision epoch t.

A decision rule is said to be *Markovian* if it depends on previous system states and actions only through the current state of the system, and said to be deterministic if it determines the action to be chosen with certainty. We define, in the following, strategies for agents in our decision problem.

Definition 2.1. A policy, contingency plan or strategy specifies the decision rule to be used at every decision epoch. A policy $\pi = (d_1, d_2, \cdots)$ is a sequence of decisions, one for every decision epoch. We denote Γ the set of all possible policies.

Definition 2.2. We call a stationary policy, a policy that determines the action to be chosen depending on the system state, regardless of decision epochs. A stationary policy has the form $\pi_s = (d, d, \cdots)$, and we denote by Γ_s the set of all stationary policies.

The utility function, denoted U, represents the satisfaction of the agent. Note that the agent is trying to maximize its utility function if we have considered a reward function in the instantaneous reward, or trying to minimize its utility function if we have considered a cost function in the instantaneous reward. Specifically, there are three types of utility functions: the total expected reward, the average expected reward, and the discounted expected reward, defined as follows:

- The total expected reward: $V = \sum_{t \in \mathcal{T}} r_t(s, a)$.
- The average expected reward: $V = \liminf_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} r_t(s, a)$.
- The discounted expected reward: $V = \sum_{t \in \mathcal{T}} \gamma^t r_t(s, a)$, where γ is a discount factor.

Note that MDP is not designed to solve decisions problems when the system state is partially observable. Hopefully, decision problems for partially observable environment can be modeled using a Partially Observable Markov Decision Process (POMDP) framework.

2.3.2 Partially observable Markov decision process

The POMDP is a very general and powerful framework, extending the application of MDPs to a wider range of problems. Smallwood and Sondik proposed the first exact POMDP algorithm in 1971, [33]. They proposed the value iteration algorithm to solve POMDP problems (see [33], [34] and [35]). Note that a POMDP is an MDP, in which agents are unable to observe the system state. The agent's goal remains to maximize the expected future rewards.

A POMDP can be described as a tuple $\langle \mathcal{T}, \mathcal{S}, \mathcal{A}, R, \mathcal{P}, \Omega, O \rangle$ where:

- $\mathcal{T}, \mathcal{S}, \mathcal{A}, R$ and $\mathcal{P}$ describe an MDP.
- Ω is a finite set of observations an agent can experience of its world.
- $O : \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\Omega)$ is the observation function, which maps actions and states to a probability distribution over possible observations.

As the agent does not directly observe the global state of the system, it infers the global system state based on past observations and actions that can be summarized in a belief vector $\omega(t) = \{\omega_1(t), ..., \omega_{2^N}(t)\}$, where $\omega_j(t)$ is the conditional probability that the system state $s(t) = j$.

Note that a POMDP may be reduced to an MDP over the belief space. Specifically, we define, in the following, an important property of the value function for a POMDP optimization: the Piecewise Linear and Convex (PWLC) property.

It is due to Smallwood and Sondik [34] that the value function $V(\lambda(t))$ is shown to be convex and piecewise linear, as illustrated in Figure 2.3, where $\lambda(t)$ denotes the belief vector at the time slot t. In the example illustrated in Figure 2.3, the domain of $V(\lambda(t))$ is partitioned into a finite number of regions. Each region is characterized by a Υ-vector. Note that the value function is given by the inner product of $\lambda(t)$ and a vector $\Upsilon_i(t)$, where $\lambda(t)$ is in the region characterized by the vector $\Upsilon_i(t)$. The belief vector is transformed into a possibly different point in the space of belief at the succeeding time slot, depending on actions and observations. Note that the domain of $V(\lambda(t-1))$ is partitioned into tree regions at the time slot $t-1$, and become partitioned into four regions at the time slot t. The PWLC property of the value function is the key element for designing an optimal solution for POMDP problems.

Definition 2.3. The value function $V(\lambda(t))$, where $\lambda(t)$ is the belief vector, is said to be PWLC if it can be represented by a finite set of $|S|$-dimensional vectors, $\Upsilon = \{\Upsilon_1, \Upsilon_2, \cdots\}$, such that $V(\lambda(t))$ is the inner product of the belief vector and a Υ-vector.

We present, in the following section, one of the major approaches of programming that is usually used in order to solve MDP and POMDP problems.

2.3.3 Dynamic programming

The Dynamic Programming (DP) techniques transform complex problems, such as MDP and POMDP, into sequences of simpler subproblems. The key idea of the DP is the multi-stage nature of the optimization procedure. Richard Bellman introduced the term *dynamic programming* in 1940s. He refined this concept to the modern meaning in 1953

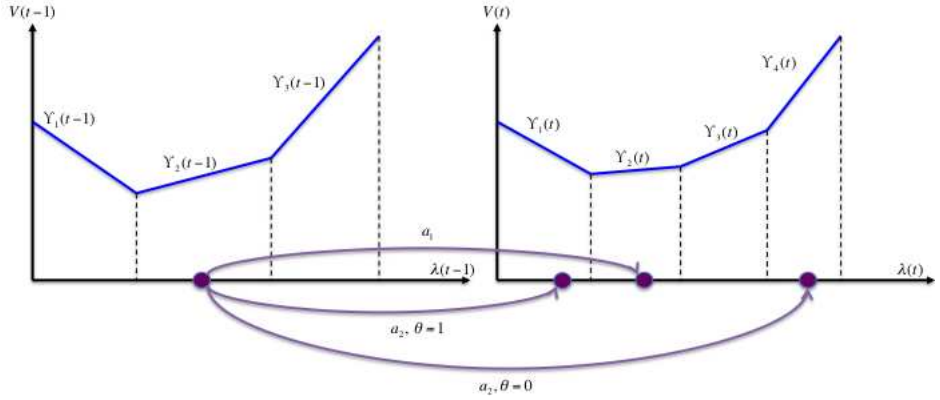

FIGURE 2.3: The structure of the value function at the time slots $t-1$ and t.

[36] for decision problems. The optimality of the DP solution results from the following principle of optimality:

Definition 2.4. In an optimal sequence of decisions or choices, each subsequence must also be optimal.

MDP and POMDP problems are solved, with the DP, by using Bellman's equations, which are also called DP equations or optimality equations.

Definition 2.5. The Bellman's equations are expressed as follows:

- The total expected reward: $V^{\pi}(s) = r(s, \pi(s)) + \sum_{s' \in \mathcal{S}} p(s'|s, \pi(s)) V^{\pi}(s')$.
- The average expected reward: $g_u(s_0) + V^{\pi}(s|s_0) = r(s, \pi(s)) + \sum_{s' \in \mathcal{S}} p(s'|s, \pi(s), s_0) V^{\pi}(s'|s_0)$, where $g_u(s_0)$ is a constant that depends on the initial state s_0.
- The discounted expected reward: $V^{\pi}(s) = r(s, \pi(s)) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, \pi(s)) V^{\pi}(s')$, where γ is a discount factor.

2.4 Queueing analysis

Basically, the queueing theory is the mathematical study of waiting lines or queues. Note that the queueing theory was applied in diverse fields. Specifically, the queueing theory represents an important mathematical tool for computer and network analysis. For example, the queueing analysis may answers the following questions:

- What is the packet delay at routers?
- What is the fraction of packets that will be lost?

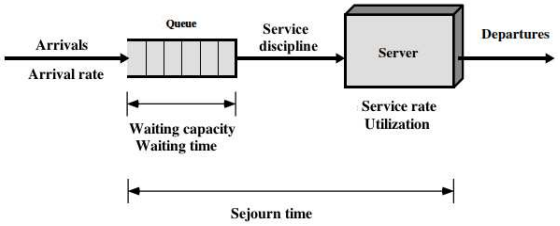

FIGURE 2.4: Single-server queueing model.

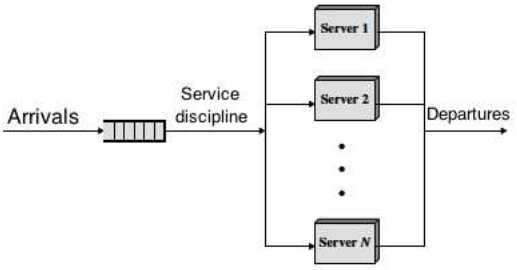

FIGURE 2.5: Multi-server queueing model.

- What is the optimal size of the buffer?

A queueing model can be either single-server (see Figure 2.4) or multi-server (see Figure 2.5), and is characterized by:

- *The arrival process*: it is usually assumed that the arrival times (of packet for example) are independent and have a common distribution. A Poisson process arrival is defined by exponential inter-arrival times.
- *The service times*: they are usually assumed to be independent and identically distributed, and independent of the inter-arrival times.
- *The service discipline*: there are many possibilities for the order in which costumers enter service (FIFO, LIFO, Random, priority, PS, etc.).
- *The service capacity*: the number of servers helping customers.
- *The waiting capacity*: the number of costumers that can be present in the system simultaneously.

Despite of the complexity of the queueing theory, its application for the performance analysis of wireless networks may be remarkably straightforward.

Kendall introduced a shorthand, four-part notation $a/b/c/d$ to characterize these queueing models. The first letter determines the inter-arrival time distribution, the second one determines the service time distribution, the third letter specifies the number of servers, and the last one represents the waiting capacity of the system. For example, the letter G denotes a general distribution, an exponential distribution is denoted by the letter M, and D denotes deterministic distribution. Some examples are M/M/1, M/M/c, M/G/1, M/M/c/K. Moreover, we have the very special PASTA [37] property:

Definition 2.6. For $M/\cdot/\cdot$ queueing systems with Poisson arrivals, the PASTA property holds: arriving customers find on average the same situation in the queueing system as

an outside observer looking at the system at an arbitrary point in time. More precisely, arriving customers observes the system in its stationary regime.

The major performance measures, in the analysis of queueing models, are:

- The distribution of the waiting time and the sojourn time of a customer. The sojourn time is the waiting time plus the service time.
- The distribution of the number of customers in the system.
- The distribution of the busy period of the server.

2.5 Game theory

2.5.1 Overview

In this section, we present some basics of the game theory. The game theory models the behavior of multiple players in interaction. It provides mathematical tools for studying conflicts and cooperation between rational players. Note that rational players are players *wanting more rather than less of a good*. The rationality is widely used as an assumption of the behavior of individuals in micro-economic models, and appears in almost all decision-making models.

There are several applications of game theory. If players know only their local state, the non-cooperative game may be adapted by players. In non-cooperative games, players act individually in order to maximize their own payoff. If players care about the long-term benefits, the repeated game may be employed in order to take into account future rewards. If a group of players cares about mutual benefits, the cooperative game may be employed. In fact, in cooperative games, coalitions of players, having joint actions, are formed in order to maximize a mutual utility. Finally, a stochastic game is a dynamic game with probabilistic transitions, played by one or more players.

We define a game by the following components:

- *A set of players*: $N = \{1, \cdots, n\}$.
- *A set of actions*: $A = \cup_{i \in N} A_i$, where A_i is the set of all possible actions for the player i.
- *An utility function*: We define the utility function for player i, $u_i : A \rightarrow \mathbb{R}$, the player preference. We denote vector of utility functions for all players by $\mathbf{u} =$

$(u_1, \cdots, u_n) : A \to \mathbb{R}^n$. Note that the utility function represents the desirability of an action for players. An utility function for a given player assigns a number for every possible outcome of the game with the property that higher (or lower) number implies that the outcome is more preferred.

In the following, we define strategies for a player in the game.

Definition 2.7. A strategy of a player defines the action the player will select in every distinguishable state of the world. In repeated games, the strategy of a player is a set of decision rule, one for each stage of the game, that specify the action to be chosen.

2.5.2 The Nash equilibrium

The most famous property of game theory is the Nash Equilibrium [38]. The NE is an action vector such that there is no individual benefit from unilateral deviation.

Definition 2.8. The NE is defined as a set of strategies (one for each player), having the property that there is no increase in the utility of any player if it chooses a different action, given other players' actions. Note that $\mathbf{u}^* = (u_1^*, u_2^*, ..., u_N^*)$, is a NE if:

$$\forall i \in \{1, \ldots, N\}, \quad u_i^* = \arg\max_{u_i} R_i(u_i, \mathbf{u}_{-\mathbf{i}}^*). \tag{2.1}$$

2.5.3 Hierarchical game

When there is some hierarchy or priority between players in the game, the latter may be modeled using a hierarchical game. Specifically, players are divided into two sets: leaders and followers. In fact, there are two stages in the game: leaders choose, first, theirs actions, and then followers choose their actions based on observations of leaders' actions. One of the proprieties of such game is the Stackelberg Equilibrium, which is a situation where neither leaders nor followers have incentive to change theirs actions.

Definition 2.9. The Stackelberg equilibrium is defined as a couple of strategy profiles $(\mu^*, \mathbf{u}^*)$, where the strategy μ^* maximizes the utility of the leaders, and $\mathbf{u}^*$ is the best response of followers to leaders' strategies.

Stackelberg game formulations were already proposed in the CR literature (see for example [39] and [40]), as the natural hierarchy between PUs and SUs is very similar to the hierarchy between leaders and followers.

2.5.4 Partially observable stochastic games

Partially observable stochastic games [41] can be considered as an extension of stochastic games for partially observable environments [42]. It is also very closely related to the model of an extensive game with imperfect information [43]. Furthermore, POSG can be seen as an extension of a POMDP for the multi-user context [33]. In fact, POSG focus on self-interested users in partially observable environments. A POSG is a tuple $< \mathcal{I}, \mathcal{S}, \{b_0\}, \{\mathcal{A}_i\}, \{\mathcal{O}_i\}, \mathcal{P}, \{\mathcal{R}_i\} >$ defined by:

- $\mathcal{I}$ is a set $\mathcal{I} = \{1, ..., N\}$ of N players.
- $\mathcal{S}$ is a finite set of states.
- b_0 represents the initial state distribution.
- $\mathcal{A}_i$ is a finite set of actions for player i.
- $\mathcal{O}_i$ is a finite set of observations for player i.
- $\mathcal{P}$ is a set of Markovian state transition and observation probabilities, where $\mathcal{P}(s', \mathbf{o}|s, \mathbf{a})$ denote the probability of taking the joint action $\mathbf{a}$ in state s results in a transition to the state s' and the joint observation $\mathbf{o}$.
- $\mathcal{R}_i : S \times A_1 \times \cdots \times A_n \to I\!R$ is a reward function for the player i.

2.6 Learning

In the architectures of future networks, where mobiles manage their communication parameters autonomously (power, frequency, ...), it is important to study learning algorithms that allow mobiles to use efficiently network opportunities. There are many applications of learning-based algorithms in the literature such as sharing spectrum in cognitive networks, routing protocols in ad hoc networks or as the distribution of traffic between operators. Specifically, MDP problems can be solved by many online reinforcement learning approaches, which can be classified into two categories: model-based approaches (e.g. RTDP [44] and Prioritized Sweeping [45]), and model-free approaches (e.g. Q-learning [46] and SARSA [47]). A model-based learning approach builds empirical models of the state evolution and the resulting reward based on interaction experiences, and applies standard DP algorithms such as value iteration or policy iteration to solve it. In contrast to the model-based approach, a model-free approach directly learns the optimal policy without specifying any model of the state evolution and reward function. There were some friendly debates within the reinforcement learning community

as to whether model-based or model-free could be shown to be clearly superior to the other (see [48] and [49]). However, all these reinforcement learning approaches suffer from the well-known curse of dimensionality problem, meaning that a practical MDP problem involves an enormous state and action spaces, which significantly impacts the complexity and the convergence time to solve the problem.

2.7 Some applications of game theory, self-adaptivity and learning in wireless networks

In this dissertation, we focus on the MAC layer, and we study the wireless spectrum management. Specifically, CR has been considered as a promising technology to enhance the radio spectrum efficiency via opportunistic transmission at link level. Note that locating frequencies that are not utilized by PUs, at a given time slot, represents the main challenge in designing CR networks. Moreover, SUs' transmissions depend not only on opportunities available in the licensed spectrum, but also on the competition with each other. Note that if CR users support multimedia applications, such as video streaming, VoIP or online gaming, they must be able to guarantee some QoS requirements. We further focus, in this dissertation, on self-adaptive wireless networks, where CR users are energy-efficient and have some QoS requirements that must be guaranteed.

Furthermore, we focus on the transport layer, and we study the congestion control in wireless networks under some QoS and Quality of Experience (QoE) constraints. Note that network users ignore, generally, the buffers' occupation level, which depends on the throughput of all users transmitting over the network. Specifically, we focus, in this dissertation, on the design of foresighted congestion control mechanisms for wireless networks, which are aware of the media content. We describe, in the following, some applications of game theory, self-adaptivity and learning in wireless networks.

2.7.1 Cognitive radio

During this dissertation, we address different layers of the protocol stack. We focus, first, on a low level of the protocol stack, the MAC layer. The first, contribution of this dissertation is an OSA policy for CR networks. In fact, we consider a system composed of several channels, where only one channel is shared between all SUs. Note that SUs have also the aptitude to sense licensed channels, and use one of them if it is idle. Specifically, we consider both the slotted and the non-slotted models, and we study the OSA as a queueing system. Thereafter, we consider that SUs may decide individually whether to sense licensed channels or to use the dedicated band. We prove the existence of a NE

between SUs, and we compare the performance of SUs at the NE with the performance of the global system, managed through a centralized controller, using the price of the anarchy (PoA).

The second contribution is a POMDP-based OSA mechanism for CR networks. In fact, we consider that SUs take into account energy consumption and QoS requirements, which were often ignored in existing OSA solutions. Specifically, we formulate the problem using a POMDP framework with an average reward criterion, and we assume that SUs may decide to use another dedicated medium of communication (such as 3G) in order to transmit their packets. We derive some structural properties of the value function, and we show the existence of optimal OSA policy in the class of threshold strategies.

Moreover, we propose two learning and knowledge extraction mechanisms. Most of researches in the OSA area assume that some information such that statistics about the activity of PUs are priory known by SUs, which may not be realistic in decentralized systems. In practice, CR users base on learning methods to get insight about the Radio Frequency (RF) environment. Specifically, we present two learning-based protocols to estimate licensed channels' dynamics: rate estimator, and transition matrices estimator.

The last, but not the least, contribution at the MAC layer is a non-cooperative OSA for CR networks. In fact, as SUs spend energy for sensing licensed channels, they may choose to be inactive during a given time slot in order to save energy. Then, there exists a tradeoff between large packet delay, due to the presence of PUs and collisions between SUs, and high-energy consumption (spent for sensing and transmitting over licensed channels). We study this problem considering a two levels approach. Firstly, we consider several SUs competing in order to access licensed channels, and we study the NE among these SUs. The NE is obtained by using a Linear Program (LP). We identify a paradox in this CR context: when licensed channels are more occupied by PUs, this may improve the spectrum utilization by SUs. Second, based on this observation, we propose a Stackelberg formulation, where a network manager may increase the occupation of licensed channels in order to improve the average throughput of SUs. We prove the existence of a Stackelberg equilibrium that maximizes the average throughput of SUs.

2.7.2 Transport layer

We focus on the transport layer and we highlight the following contributions: The first contribution is a media-aware congestion control mechanism. In fact, we consider several end-to-end users sharing the network. As users ignore the congestion status at bottleneck links, we model the congestion control using a POMDP framework. Moreover, we prove the existence of an optimal stationary policy, and we derive some structural properties

of the value function. Thereafter, we propose a low-complexity learning-based algorithm that can be implemented on mobile devices having a limited computational capacity.

The second contribution, at the transport layer, is a QoE-aware congestion control for conversational services in wireless environments. In fact, standard TCP congestion control is based on network performance metrics (namely QoS metrics) and not on a subjective metric of the quality perceived by the user (measured through the QoE metrics). Therefore, we propose an end-to-end QoE-based congestion control mechanism that maximizes the subjective quality of multimedia through Mean Opinion Score (MOS) feedbacks from receivers.

2.8 Conclusion

In this chapter, we have introduced some theoretical concepts that will be useful for the analysis of wireless networks in partially observable environments. We have presented some applications of the game theory, self adaptivity and learning in partially observable environment. Specifically, we study the OSA at the MAC layer in CR networks and the self-adaptive congestion control at the transport layer. Since the static spectrum allocation has been shown not efficient and unable to manage the increasing number of wireless users, a new licensing scheme is being developed allowing the dynamic access to the spectrum in order to improve the spectrum utilization, through CR approaches. Nevertheless, implementing the CR technology introduces new challenges about the management of the wireless spectrum. To achieve such goal, several disciplines can be involved, such as decision theory, queueing analysis and game theory. We study in the following part of this thesis the impact of the OSA mechanisms on the performance of SUs. Specifically, the next chapter focuses on the performance of SUs through a queueing analysis. We consider both the centralized and the decentralized models.