Distributed algorithms in autonomous and heterogeneous networks

Bah Aladé Habib Sidi

THÈSE

Présentée pour obtenir le grade de Docteur en Sciences de l'Université d'Avignon et des Pays de Vaucluse France & de l'Université Mohammed V-Agdal Rabat - Maroc

SPÉCIALITÉ : Informatique

École Doctorale 536 « Sciences et Agro-Sciences »
Laboratoire Informatique d'Avignon (UPRES No 4128)

Distributed algorithms in autonomous and heterogeneous networks

par

SIDI Bah Aladé Habib

Soutenue publiquement le 13 Décembre 2012 devant un jury composé de :

M.	Zwi ALTMAN	Engineer R&D at France Telecom	Rapporteur
M.	VINOD KUMAR	Professeur,Alcatel Lucent Bell Labs	Rapporteur
M.	ROULLET Laurent	Ingénieur á Alcatel Lucent Bell Labs	Rapporteur
M.	EL-AZOUZI Rachid	Maitre de Conférences/HDR, LIA, Avignon	Directeur
M.	HIMMI Mohammed Majid	HDR, LIMIARF/FSR, Rabat	Directeur
M.	HAYEL Yezekael	Maitre de Conférences, LIA, Avignon	Co-directeur
M.	Philippe MICHELON	Professeur, LIA, Avignon	Examinateur (Président)
M.	BOUYAKHF El-Houssine	Professeur, LIMIARF/FSR, Rabat	Examinateur
M.	Claude CHAUDET	Maitre de Conférences, Telecom ParisTech, Paris	Examinateur

Laboratoire LIA, Avignon *Laboratoire LIMIARF*, Rabat

Preface

This thesis was carried out between the laboratory LIA, University of Avignon and LIMIARF of Mohammed V University in Rabat. In my thesis, I could during my many trips to LIMIARF Rabat meet and interact with my supervisors, which allowed me to enrich my experience in research, but also and especially my personal experience. This rich and diverse background has allowed me to produce the present work.

I would like at the end of this work to address special thanks to my supervisors, Dr. ELAZOUZI Rachid and Dr. Yezekael Hayel for all their assistance, support and guidance throughout the course of this thesis. I could learn from their advices and their combined research experiences to improve my technical skills and knowledge in order to become a careful researcher.

I wish to express my deep gratitude to Assistant Pr. Mohammed Majid HIMMI and Assistant Pr. Loubna Echabbi, my supervisors at LIMIARF and INPT, and Pr. El-Houssine BOUYAKHF Director of LIMIARF for enriching discussions that we had, to have accepted and allowed to develop this collaboration between the two universities during my thesis, for all the help, support and the efforts they have made to this thesis is carried out in the best possible conditions.

I sincerely thank all the members of the jury in particular Professor Philipe MICHELON for agreeing to chair the jury of my thesis, Professor Zwi ALTMAN, Pofessor VINOD KUMAR and Mr. ROULLET Laurent, for generously agreeing to review my thesis, for their careful reports and for all exchanges, discussions and useful interactions we had during the course of my thesis and its validation. I also thank Mr. Claude CHAUDET for agreeing to be a member of my jury.

My sincere thanks go to Dr. Majed Haddad, Mr and Mrs Julio Rojas-Mora and Sulan Wong for their considerable assistance and methodical advices, to Dr Afef FEKI for our various interactions within the project ECOSCELLS and for agreeing to host me for an extended visit at Bell-Labs, France. I thank Professor Franscesco De Pellegrini for his involvement in my thesis and for our many collaborations.

I wish to warmly thank all my colleagues, friends and members of the lab, especially Manjesh KUMAR HANAWAL and Wissam CHAHIN for their assistance and support.

Finally, I would like to thank my parents, my friends and all my family to have accompanied me unconditionally, for being by my side throughout my thesis. I do not forget all those who are dear to me, and without whom this work would not have been possible.

Abstract

Growing diversity of agents in current communication networks and increasing capacities of concurrent technologies in the network environment has lead to the consideration of a novel distributed approach of the network management. In this evolved network environment the increasing need for bandwidth and rare channel resources, opposes to reduction of the total energy consumption.

This thesis focuses on application of distributed mechanisms and learning methods to allow for more autonomy in the heterogeneous network, this in order to improve its performances. We are mainly interested in energy efficient stochastic mechanisms that will operate in a distributed fashion by taking advantage of the computational capabilities of all the agents and entities of the network. We rely on application of Game theory to study different types of complex systems in the distributed wireless networks with dynamic interconnectivity.

Specifically, we use the stochastic reinforcement learning tools to address issues such as, distributed user-network association that allows achieving an efficient dynamic and decentralized radio resource management. Then, we combine access selection procedures with distributed optimization to address the inter-cells interferences coordination (ICIC) for LTE-advanced networks using dynamic power control and design of fractional frequency reuse mechanisms. Moreover we address in non-hierarchical networks, more precisely in Delay Tolerant Networks (DTNs), decentralized methods related to minimization of the end-to-end communication delay. In this framework we are interested, in addition to Nash equilibrium, to the notion of evolutionary stable equiliria in the different context of Evolutionary Games, Markov Decision Evolutionary Games and Minority Games. As the major parts of our work includes testing and validations by simulations, eventually we present several implementations and integrations materials for edition of simulation platforms and test beds.

Résumé

La diversité croissante des différents agents constituant les réseaux de communication actuels ainsi que capacité accrue des technologies concurrentes dans l'environnement réseau a conduit à la prise en compte d'une nouvelle approche distribuée de la gestion du réseau. Dans cet environnement réseau évolué, le besoin en accroissement de la bande passante et en ressources rares, s'oppose à la réduction de la consommation énergétique globale.

Dans notre travail nous nous intéressons à l'application de mécanismes distribués et de méthodes d'apprentissages visant à introduire d'avantage d'autonomie dans les réseaux hétérogènes, mobiles en particulier, tout en améliorant les performances par rapport aux débits et à la qualité de service. Notre étude se concentre principalement sur l'élaboration de mécanismes distribués stochastiques et énergétiquement efficaces en profitant des capacités de calcul de tous les agents et entités du réseau. Divers outils de la théorie des jeux nous permettent de modéliser et d'étudier différents types de systèmes dont la complexité est induite par la grande taille, l'hétérogénéité et le caractère dynamique des interconnexions. Plus spécifiquement, nous utilisons des outils d'apprentissage par renforcement pour aborder des questions telles que l'attachement distribué des utilisateurs permettant une gestion dynamique, décentralisée et efficace des ressources radio. Nous combinons ensuite les procédures de sélection d'accès à des méthodes d'optimisation distribuées du type gradient stochastique, pour adresser le problème de coordination des interférences intercellulaires (ICIC) dans les réseaux LTE-A. Cette approche se base sur un contrôle de puissance dynamique conduisant à une réutilisation fractionnaire des fréquences radios. Par ailleurs nous adressons dans les réseaux decentralisés non-hiérarchiques, plus précisément les réseaux tolérants aux délais(DTNs), des méthodes décentralisées liées à la minimisation du délai de transmission de bout en bout. Dans ce cadre nous nous intéressons, en outre des équilibres de Nash, à la notion d'équilibre évolutionnairement stables dans différents contextes de jeux évolutionnaires, jeux évolutionnaires décisionnels markoviens et jeux de minorité. Enfin, la majeure partie du travail effectué se rattachant aux tests et validations par simulations, nous présentons plusieurs éléments d'implémentations et d'intégrations liés à la mise en place de plateformes de simulations et d'expérimentations.

Mots clés : Réseau de télécommunication sans fil, évaluation des performances, algorithmique distribuée et protocoles de communication.

Contents

List of Figures

List of Tables

Introduction

General context

In first generations of communication networks, one generally considers centralized intelligence in the network at the opposite of distributed systems. This choice was done especially in order to reduce as much as possible the computational burden and complexity in equipments of the networks end-users. This way of conceiving the networks and in particular hierarchical networks (like cellular mobile networks), is generally suitable when there is only one technology monitored by a single operator of the network. In the area of mobile communications, mainly in mobile cellular networks, there have been a prominent evolution with the advent of several new technologies of communication based on improved techniques and enhanced devices. From the first to the third generation of mobile communications technologies, considerable improvement of transmission rates has allowed the introduction of new applications and development of considerable amount of services for the end-users. As a result of this dramatic increase of the network capacity, the nature of the end-users equipments in the network has then extended from simple land-line phone and mobiles phones, to smart phones and even laptops, fridge, appliances ... with the capability to communicate between each other. From another side, in the domain of computer networks, ad-hoc networks have moved from the intranet to wider areas, thanks to the occurrence of the INTERNET and evolutions of processor capacities at the end-user. Those two main types of communications networks, namely mobile and computer networks, have thus constantly evolved to eventually merge inside the core and even the air interface of the communication networks as we know today(the Global Unified Network). As a matter of fact, several new technologies are now forced to coexist inside a common network area (local area networks access points, hot-spots, mobile networks Bases Stations, etc.). As a consequence of this heterogeneity interoperability has become a crucial aspect in the evolved new paradigm of mobile communications for example and yielded from the research community and standardization organisms the release of several new standards see [64] and [30]. From the network management standpoint, incumbent operators no longer detain the monopoly of the market and one has witnessed during recent years raise of new comers as well as growth of virtual operators in the sector of telecommunications all around the world. Users are thus offered access to several concurrent technologies with more or less similar capacities and monitored by different operators of the network. As an example, in the mobile cellular networks, the HSDPA, EV-DO and UMTS mobile communication standards have been observed to provide comparable performances to end-user in the reverse link[58]. Similarly, from HSPA+, WiMAX 1.5 to current fourth generation of mobile communication technology, LTE, analogous performances in terms of network capacity was announced inside a tight range of values [43][125]. The current communication systems have thus become more and more complex and have dramatically evolved toward heterogeneous network architectures with a high demand of traffic making centralized management stringent and mostly suboptimal.

Motivation and general overview

The question of how to manage and how to take efficiently advantage of the pool of available resources has then become prominent. Indeed, there has been a growing interest on how users should be assigned to exploit those resources more efficiently. Many proposals appeared in the network architecture as well as in the design of new protocols. Some approaches are based on joint Radio Resource Management (RRM) in order to efficiently share the scarce network resource with the objective to take advantage of all the heterogeneous resources available. But resource management for QoS provisioning in heterogeneous wireless network is a complex task because of heterogeneity of policies and mechanisms used for QoS provisioning in different wireless networks, along-with highly probabilistic behavior of network traffic. Moreover, the presence of the multiple radio access technology provides roaming capability in different wireless networks through vertical handovers. Those handover operations may cause significant degradation to QoS provisions though. To address this problem, authors propose in the literature new approaches such as joint radio resource (JRRM) mechanism to achieve an efficient usage of a joint pool of resources. For example, [68] and [69]propose a framework for a JRRM based on fuzzy neural methodology and reinforcement learning algorithms. Apart from advances in the network architecture, major advances in the field of electronics on the mobile user side have allowed for more computational capabilities at mobile agent's terminals (the so called Smart-phones). For example the user equipment in 3Gs networks is capable of operating on several accesses at the same time using the multi-homing technique. Furthermore, integration of proximity communications unlike,infrared or bluetooth on another hand has encouraged the development of peer to peer like communications with mobile phones in ad hoc mode where mobile devices can communicate between each other without resorting to an access point.

Next generation wireless systems are indeed expected to enable versatile and flexible communications between mobile and autonomous users even in case where no infrastructure is available. In these regimes, in fact, due to nodes' mobility, network topology may change rapidly and in a non-deterministic way in time. All customary network functionalities such as topology discovery, routing and messaging have therefore to be handled by the mobile nodes, thereby creating the need for efficient decentralized algorithms. The design of such algorithms under continuous topology changes and using only local information available at mobiles requires a specific design effort. On one hand, high mobility and frequent network partitioning rule out Internet routing protocols which operate poorly under uncertain networking conditions, high mobility and frequent network partitioning. But, on the other hand, many users carry advanced computing devices such as smart-phones, netbooks, etc. As we just mentioned, such devices are equipped with wireless interfaces so that it is possible to sustain communication by leveraging intermediate nodes acting as relays, the so called *carry-store-and-forward* strategy. Messages can thus arrive at their destination thanks to the mobility of some subset of nodes that carry copies of the message stored in their local memory. The idea of networks with such characteristics has been introduced in literature as Delay (or Disruption) Tolerant Networks (DTNs) [5, 99, 70]. The DTN thus comes as a solution to

the design of a new network architecture where the mobile user plays a central role in the structure of the network. We will address particularly the DTNs in the third part of our thesis. This new paradigm of the network has then motivated the interest in decentralized systems which come to define how it is possible to establish a trade-off between network performances and signalization load under the dynamics of the ever changing environment. This new highly heterogeneous and dynamic environment thus raise among several others, the problems of agent coordination for overall performances optimizations, global energy consumption through signaling or distributed computations for self-organization and interoperability. However, if we resort to distributed computing it will be important to analyze how will perform a decentralized approach of the network in this predesignated environment. We will expect from the study of decentralized systems, to motivate the design of more flexible networks and technologies, which will adapt to the dynamic of the environment and introduce some level of fairness while serving a large population of mobile clients. Nevertheless, the major downside of the decentralized architecture comes from the fact that, the more the system is distributed the more the computation burden at distributed agents is increased, which can generate an additional energy cost for computation. In order to skirt this problem mainly constrained by fast draining and very limited life time of batteries in mobile devices, several approaches such as automatic screen light management, the execution of optimized codes in the mobile, clouds phones and many others was proposed in the literature. However, all those technical solutions that rely mainly on the optimization of internal processing performances are still insufficient. In [12] authors developed throughout their thesis several methods for energy conservation and optimization related to sleep mode management and network card operations management. Indeed one of the major causes observed of the energy consumption in smart phones is due to data upload/download using the network resources of the mobile [88].

But our approach is different. In our thesis, we seek to introduce some distributed intelligence in the network, through the design and implementation of new strategic behaviors and protocols for the network agents, this in order to optimize the overall network performances. As a matter of fact, in distributed network architectures the optimization of energy expenditure at a mobile does not totally depend on the solely behavior of the mobile itself but also depends by the behavior of the population of other mobiles with which it is in interaction. This imposes to the user some strategic behavior toward its surrounding environment and poses a coordination problem which is interesting to address in the dense architecture of current networks where the cost of cooperation can be exponential. Users of the networks then need to find autonomously their optimal strategies. We will study how distributed mechanisms and learning methods can be used to allow for more autonomy in the heterogeneous network. Learning is a crucial aspect of intelligence as it allows individuals to learn by they own the accurate behavior or strategy to adopt in face of a particular situation. More specifically, machine learning is a scientific discipline that is concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases [119]. Machine learning has emerged as a vibrant discipline with objective of developing machines with learning capacities. There are many effort in different disciplines, all contributing in the direction of improving the intel-

ligence of machines in several dimensions. As previously motivated future network management techniques should be able to minimize human intervention, set their own parameters, optimize these parameters, and heal problems by themselves when they occur. This requires to improve the intelligence of devices to be able to select the best action by repeated interactions with unknown random environment. One of the most important branch of Artificial intelligence, is the reinforcement learning which allows devices or software agents to learn their behaviors based on feedback from the environment and automatically determine the ideal behavior in order to maximize its performance. For these reasons, we are interested to use the tools of reinforcement learning to develop some learning schemes for the distributed network environment. The first part of this thesis will then be dedicated to a survey on reinforcement learning algorithms.

Our objective thus lies in the definition of new methods in order to improve the network performances, this using different tools such as Game theory, Markov and Markov decision processes, Queuing theory, design of stochastic algorithms, optimization and control theory. Our central argument is to design some reinforcement learning techniques as a tool to study the different problems related to decentralized networks. We will address problems such as resource allocation and optimal power control mechanism for energy efficiency. As described in figure 1 we will investigate in this thesis frameworks in which learning methods in a distributed environment can be applied to study interaction between several actors in a situation of game. Let's briefly define what we understand by the notion of game. We define a Game as a situation where smart agents (featured with the capability of making decisions) interact strategically with the objective of maximizing their own profit. This assumption which tends to assimilate mobiles devices to decisional agents is very important for application of Game theory models in networking. In fact, the development of more and more complex algorithms on board of mobile phones makes the user equipment more autonomous which makes this constraining assumption valid. One can simply consider that strategical decision schemes can be implemented on board of the mobile agent. In a game individuals are given to select actions among a predefined set of actions given a predefined context where Game theory is used to analyze the interactions between players and generally one tends to compute an equilibrium strategy profile of the game given those interactions. The application of Game Theory(GT) in communication networks although recent is not novel and has been successfully employed in particular in wireless and communication networks. All along this thesis we will refer to several tools from game theory which will be defined as they are introduced in the document.

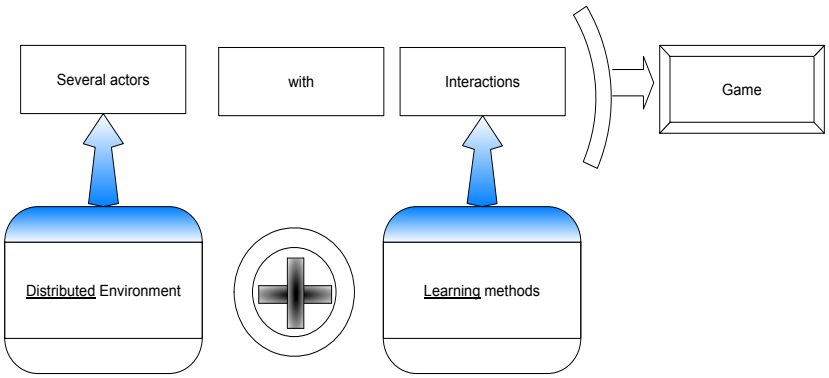


Figure 1: *Connexion between distributed learning and game framework*

Contributions and Organization of the thesis

The different contributions of this thesis can be separated over two main aspects of networks performances optimization. First we identified and focused on the problems of distributed resources allocation including network load management and mitigation of inter-cell interference load in a distributed fashion for mobile cellular networks. Second we focused on the problem of optimal routing strategies and forwarding control mechanisms design in DTNs under the constraint of overall energy consumption for ad hoc networks. We can split our contributions in the following points :

Mobile cellular networks : Here we propose some mechanisms for interoperability in a context of small cells or coexistence between two different technologies. We propose a novel approach to achieve an efficient dynamic and decentralized radio resource management in heterogeneous wireless network scenarios. We base on reinforcement learning techniques to solve a problem related to user network association. The network framework is set and clearly defined relying on propositions of tight and loose coupling of concurrent network architectures in the cellular network environment. The objective of this approach is to perform dynamically and efficiently, regarding overall network performances, offloading or load-balancing between different network severs or coexisting severs in general. Specifically our approach applies in a network situation where several mobiles users are in presence of different communications technologies and need to select in an autonomous way the access to be connected with. Note that in heterogeneous networks mobile users are frequently subject to such situations. In the heterogeneous environment, an implicit interaction thus takes place between the population of mobile users through the different servers or access nodes. This network scenario models as a non-cooperative game between users where the mobile agents observe their utilities as a function of the throughput obtained in result of the action taken. In this game users actions are to select the sever on which to send their tasks. During the transient states of the system, players interact over several rounds of iterations and we rely on a distributed reinforcement learning algorithm, implemented on board of each device, to drive the whole system to a steady state corresponding to the Nash Equilibrium of our non-cooperative game. This approach thus provides by the

design of an autonomous decision making scheme, a stochastic adaptive solution to the user-network association problem. Furthermore, we design a fractional frequency reuse technique for inter-cell interference coordination(ICIC) in an OFDMA small cells network using game theoretical setting. We define a Stackelberg game and propose a hierarchical algorithm to allow distributed convergence towards the Stackelberg equilibrium. Here the idea is to combine our approach for access selection with stochastic power control for interference management. Recent works such as fractional frequency reuse [47, 107] and soft frequency reuse [2] allowing users in different channel conditions to benefit from different reuse patterns have been proposed in this subject. Still, all of these schemes are static interference management approaches, where a specific reuse pattern is predetermined a priori by a network operator at off-line. Our approach goes in the same line with [26] where authors use the Gibbs sampler to simulate a SON, using power control and user association to minimize the system energy cost as the sum of the inverse of user's experienced SINR. The main difference with this approach is that they assume a predefined discrete set of power levels and a nested user association optimization with the premiere objective of minimizing the system energy cost. Stolyar and Viswanathan scheme [110] is related to ours in the sense that they use a shadow algorithm to solve a linear program in order to find a reallocation of users on sub-bands that minimizes the power utilization, given a current power allocation configuration. Then they prove that for any user allocation, there exists a corresponding power allocation which corresponds to a Nash equilibrium. But their theoretical work based on fluid approximation does not describe a clear implementation directive with respect to the current mobile network architecture and resources which is the main purpose of our approach. Other techniques such as [42, 112] focus on the bandwidth allocation, channel resources management and compute the spectral efficiency. A constrained geometrical and geographical topology of the network is usually assumed and the resources are assigned considering inner and edge cell users while the powers levels are generally assumed constant and uniformly distributed [9]. Our approach is to consider a reuse 1 scheme and propose a mechanism independent of any predefined spatial repartition of users in the cells. The work of authors in [111] is similar to ours but they use differently the implementation of the gradient algorithm and their virtual scheduling algorithm does not clearly address the users association. The proposed algorithm here works at two distinct levels. At the higher level, we use a gradient descent based algorithm to set a power control mechanism at the eNodeBs. User equipments at the lower level solve the attachment problem using the pursuit distributed learning algorithm that will be defined in part I. An extensive analysis of the behavior of our approach under a perturbed environment has allowed us to conclude on the expected performances of our algorithms in a nearly realistic environment. We then follow on with implementation aspects of integration and simulation of our self-organizing algorithm. In this perspective, we contributed to the evolution of the LTE simulator from university of Vienna and presented the results obtained through simulations of our approach in a nearly realistic environment.

Routing configuration game in mobile ad hoc networks : We propose an autonomous routing configuration scheme for relays in DTNs. Our scheme allows mobile users to configure dynamically and autonomously their routing policies according to network

reward and other users configurations. The idea has been to define a decentralized non cooperative and energy efficient game which leads to an optimal trade-off between the successful delivery probability and the number of infected relays at the equilibrium. The aim is to provide a scheme which maximizes the expected delivery rate while satisfying a certain constant on the number of forwarding per message. We assume that each mobile may decide which routing protocol it wants to use for delivering packets. We also restrict to the case that only two routing protocols are available to mobiles: epidemic routing and two-hops. Epidemic routing on one hand, tends to maximize the probability of successful transmission of the message by flooding the network with several copies of the message originated at the source. On the other hand two-hops routing tends to limit energy consumption in the network using a hard control of the message forwarding. In two-hops routing relay nodes deliver a message, received from the source node only to the destination. We then try to define an optimal proportion of epidemic versus two-hop nodes in the population of mobiles. In our model we estimate the probability of successful transmission by modeling the propagation of the message from the source to the destination as a fluid model using mean field approximation. By considering two different overlapping regions, each containing respectively the source and destination nodes, we define a migration pattern for the messages trough a region of interaction from the source to the destination. The proposed scheme requires the introduction of a notion of a rewarding mechanism, in order to incite the mobile agents to find autonomously the appropriate stable policy to adopt which corresponds to the defined equilibrium of our game. By comparison of our mechanism performances with the utility achieved by the global optimum we also provide an insight about the worst possible performance degradation, using our approach. This is the notion of the price of anarchy.

Mechanism design and stochastic approximation for forwarding control in DTN : In this point, we provide a so-called *mechanism design* for controlling the evolutionary dynamics of the activation profile of the population of relays through the choice of appropriate forwarding control at the source of a DTN. This is a particular way of governing the replicator dynamics: such an approach provides very interesting insight into the feasibility of optimal mechanism design and, the technique appears novel compared to known results in literature. This approach operates at two different stages and bases on a two-hops routing technique. In fact, this routing protocol has two major advantages: first, compared to epidemic routing, it performs natively a better trade-off between the number of released copies and the delivery probability. Second, forwarding control can be implemented on board of the source node. Under two hop routing, the source transmits a message copy to mobiles devices it encounters. Relays, conversely, forward to the destination only. In this context, the higher the number of relays joining the forwarding process, the higher the success probability. However, battery lifetime of mobile devices may deplete due to continuous beaconing operations, which may be a critical factor discouraging the usage of mobile devices as relays for DTN-based applications. A solution is to design reward-based forwarding mechanisms where the probability of forwarding becomes function of the competition within a population of mobiles: a relay receives a unit of reward if it is the first to deliver the message to the destination. For each message generated by the source, a relay may choose two different actions

that affect message relaying: full activation or partial activation, i.e., being active for a shorter time period and then go back to low power mode, thus saving batteries. The population profile thus define the proportion of the population that decides to be fully active and conversely to not be active. The study of the evolutionary dynamics of the population dynamic give some insight about the consistency of the network efficiency. Furthermore, we propose a hierarchical algorithm that allows the source to achieve the optimal forwarding control by iterative learning procedure, with the objective of maximizing the probability of success. Then we extend our approach in a new framework of population game in DTN where diversity of sources and destinations give birth the occurrence of several local interactions for which we study a competition for good between the mobile agents. The competition is ruled by and activation control game, in which agents need to find the appropriate strategy to adopt. The novelty here is that the strategy of a mobile relay determines not only the immediate reward but also the transition probability to its next battery energy state. Compared to the previous approach, we introduce here the constraint of battery life-time. Indeed the more the node is active the more it has a chance to drain out its battery. The decision of being active thus depends by the battery state that the mobile is experiencing at a given moment. The problem is formulated as a Markov Decision Evolutionary Game (MDEG), where each relay wishes to maximize the expected utility. We provide a characterization of the Evolutionary Stable Strategies (ESS) for these games and show a method to compute them. We eventually highlight a paradox by showing that the success probability is not always increasing with the number of message copies, and may well *decrease* under some conditions, which is adding an intriguing novel facet to the control of forwarding in DTNs.

A framework for coordination without cooperation in DTNs Eventually, we remain in the context of activation control game and propose a framework based on the principle of coordination without cooperation in DTNs. This framework relies on a Minority Game(MG) setting. Our approach is pivoted around the following idea: by modeling competition of relay nodes as a coordination game we show that it is possible to enforce a behavior of cooperation within a population of relays through competition. The relay activation control in this context is fully decentralized and does not require additional control messages. In this purpose, we use a novel and specific utility structure. Such utility is rooted on the following trade off: the success of a tagged relay depends explicitly on the number of opponents met, namely, nodes adopting the same strategy. In fact, the bigger the number of relays participating to the message delivery, the higher the delivery probability for the message, but indeed the less the chance for the tagged relay to receive a reward from the system. There comes the minority rule of our activation control game. Indeed, the global activation target settles the number of opponents of a randomly tagged relay, i.e., the active fraction of the population so that nodes always try to avoid non-positive utility. The objective of our approach in DTN is the global one : finding a trade-off between the energy consumption and successful forwarding of messages originated at source nodes. Since the MG scheme rules the number of active relays, the message source can achieve a target performance figure, e.g., the probability of successful message delivery, by setting the rewarding mecha-

nism appropriately. Conversely, the source can reduce the quality of service in order to reduce the relays energy consumption. Thus, our incentive mechanism can match quality of service metrics such as delivery probability to the available resources. Compared to existing literature, the novelty of this approach stands in the way the activation and forwarding process is jointly controlled by the operator of the network acting on a distributed mechanism which takes place among the competing relays based on the MG. We decline our approach for homogeneous population where all nodes present similar technical characteristics such as communication range and battery profile. Then we extend to heterogeneous population of nodes in the DTN. Eventually we design a distributed algorithm to attain the defined equilibriums of the game.

The organization of the thesis is the following. As announced previously, in a first part we will give a survey on leaning algorithms and mainly, reinforcement learning algorithms. The second part discusses in two chapters, the contributions presented for mobile cellular networks. In a third part, we address different sort of games for performances optimization in DTNs. The three chapters in this part, focus on the contributions related to routing configuration game in mobile ad-hoc networks, mechanism design and stochastic approximation for forwarding control in DTNs and framework design for coordination without cooperations in DTNs. The last part of the thesis is dedicated to implementations and integrations for evolution of a simulation platform for LTE system level simulations.

Part I

State of Art of Learning Algorithms : Focus on reinforcement learning algorithms and application in games

Introduction

In this part we will present a general theoretical background behind learning mechanisms. Note that the majority of the developments here are drawn from one of the deliverables for the ECOSCELLS project to which we have participated. We essentially focus here on the notion of reinforcement learning. Reinforcement learning is defined not by characterizing learning algorithms, but by characterizing a learning problem. In many situations, the reinforcement that can be supplied as an evaluation, is itself a random variable. Hence all actions have to be tried a number of times to evaluate the mean reinforcement associated with them in order to find the best action. For successful learning, there should be an appropriate mechanism for action selection and reinforcement that can utilize the experience in the form of action selected during learning. However, each action should be tried many time in order to evaluate each action in view of the randomness of the environment. Such a learning model is useful in many scenarios in small cells network for example and networking application in general when often information available to agent is a set of training examples. Hence it would be attractive to have an algorithm that can learn to make good choices based on some 'noisy' feedback regarding the 'goodness' of the choices.

Stochastic optimization

Optimization with noisy corrupted measurements is a common problem in many areas of engineering. Many efficient methods like the steepest descent method and Newton method are available when gradient is explicitly available. But usually due to the lack of sufficient information concerning the structure of the system, only the noise corrupted value of function at any chosen point can be observed. Several important classes of algorithm are available for solving the optimization problem when only the noise corrupted observations are available : Stochastic approximation [65], learning automata [81, 80], and reinforcement learning with function approximation.

Stochastic approximation

The stochastic approximation algorithms, introduced by Robbins and Monro in [95] and by Kiefer and Wolfowitz in [61], have been the subject of large number of applications. In general a stochastic approximation algorithm is a discrete time stochastic process whose general form can be written as

$$X_{n+1} = X_n + \epsilon_n Y_n \tag{1}$$

where X_n takes its values in some Euclidean space, Y_n is a random variable function of "noise-corrupted" observations taken on the system when the parameter is set to X_n, and ϵ_n is the step size with $\epsilon_n > 0$. Typically X_n represents the parameter of a system which is adapted over time and $Y_n = f(X_n, \xi_n)$. At each time step the system receives a new information ξ_n that causes X_n to be updated according to an algorithm characterized by the function f. For example, the function f is used so that some goal

(estimation, optimization, cooperation, etc) is achieved. Robbin and Monro in [95], developed a stochastic algorithm for finding the value of X, X_0, satisfying $f(X) = \alpha$ where f is unknown. In that case the random sequences Y_n is a "noisy" estimate of value of $f(X_n)$ obtained by averaging many observations, i.e, $E[Y_n(X)|X] = \alpha - f(X)$. Moreover, they considered that the step sizes in the parameter updates of the algorithm, go to zero in order to obtain an implicit averaging and eliminate the effect of the noise. In particular, they assumed that the sequences ϵ_n satisfy

$$\epsilon_n \to 0 \text{ and } \sum_n \epsilon_n = \infty \tag{2}$$

Under some conditions on the function f and random variable Y_n, then the sequence X_n converges to X_0 with probability one. Kiefer-Wolfowitz [61] introduced an algorithm based on stochastic approximation in which the gradient of function f is approximated by a finite difference method and using the evaluation function obtained at points which are chosen close to each other. This algorithm is useful to find the maximum of the unknown $f(x)$. The theory of stochastic approximation has been extensively used in problems of signal process, adaptive control [65, 73, 7] and recursive estimation [60]. Much of the classical works in stochastic approximation dealt with the situation where it is convenient to rewrite the noise term as

$$Y_n = F(X_n) + M_n \tag{3}$$

where $F : \mathbb{R}^m \to \mathbb{R}^m$ is a deterministic vector field obtained by suitable averaging. In some situation the noise in each observation M_n is martingale difference, i.e.,

$$E[M_n|M_i, i < n] = 0$$

The asymptotic behavior of the algorithm can be approximated by the asymptotic behavior of the solution to the Ordinary Differential Equation (ODE)

$$\frac{dx}{dt} = F(x) \tag{4}$$

This method called (ODE) was introduced by Ljung [72]. In the next simple example, we present the connection between the asymptotic behavior of the algorithm and that of the mean ODE. Let the stochastic approximation in (1) and suppose that $E[|Y_n|^2] < \infty$ and the function F is continuous. We have

$$X_{n+m+1} - X_n = \sum_{i=n}^{m} \epsilon_i F(X_i) + \sum_{i=n}^{m} \epsilon_i M_i$$

Since expected of the second term is of the order $\sum_{i=n}^{m} O(\epsilon_i^2)$, the noise will go to zero, and the sequence X_n will follow the mean trajectory of $E[X_{n+1}] = E[X_n] + \epsilon_n F(E[X_n])$. Then when $n \to \infty$, the step size go to zero and then the mean trajectory can be approximated by the solution to the mean ODE $\frac{dx}{dt} = F(x)$.

In the sequel, we present different learning and machine learning technic developed in stochastic approximation

- **Stochastic approximation to gradient descent** : The stochastic gradient descent (SGD) method [90] is a particular case of stochastic approximation, and is of particular interest in the context of distributed stochastic optimization. It is an algorithm to maximize a sum of convex function in a distributed manner. In particular, we consider a utility function U, that aggregates the utility of I entities,

$$U(x) = \sum_{i=1}^{I} U_i(x)$$

each of which depends on a parameter vector x, $x = (x_1, .., x_L)$. A typical SGD update equation of the parameter α can be written as

$$X_{n+1} = X_n - \alpha \sum_{i=1}^{I} \Delta U_i(X_n) \tag{5}$$

- **Q-Learning :** Q-learning algorithm is a reinforcement learning method that applies to Markov decision problems with unknown costs and transition probabilities; it may also be viewed as a direct adaptive control mechanism for controlled Markov chains (see [13]). Q-Learning might be the most often implemented among standard reinforcement learning methods. The main idea of the algorithm consists in evaluating the current policy $\pi(a|s)$ through a value function $(Q_\pi(s,a))$ which is updated from observed reward and transited state. This step is referred to as "policy evaluation" or "critic". A policy is inferred from this value function and re-injected in the algorithm in order to update the value function. This step is referred to as "policy improvement" or "actor". Such kind of algorithm is thus often called "actor-critic architecture". The value function Q tends to approximate the optimal value function (fixed-point technique) for which the associated policy π is the seek optimal policy.

  In this algorithm, exploration/exploitation compromise can be implemented through the use of greedy policies to explore the state-action space. The choice of the learning rate is not really sketchy for most of applications: if one wants to track non-stationarity of the optimal policy, a fixed value will be set, otherwise if one wish to obtain accurate estimations of the optimal policy (in case of probabilities computations for a random optimal policy for instance), a sequence of small values converging towards 0 will be chosen. Standard version of the Q-Learning algorithm can be improved by including the current action chosen in the updating computation of the value function (this kind of algorithm is referred to as SARSA). Another possible improvement is to use memory signals (called eligibility traces, often with exponential decay) for every state-action pair in order to update the value function for several or all pairs at every iteration, instead of just one. This result is an acceleration of the learning phase for the optimal policy which can be significant for some applications, as well as value function estimates more robust.

- **Stochastic Recursive Inclusion** The stochastic recursive inclusion is strongly motivated by certain problems in economics and games theory. This type of algorithms is used as approximation algorithms to analyze non-continuous stochastic

process, such as the best-response process in zero-sum games. This algorithm represents an important generalization of the previous stochastic approximation. The idea is to replace the ordinary differential equation by the stochastic recursive inclusion which defined as

$$\frac{dx}{dt} \in \tilde{F}(x) \tag{6}$$

where $\tilde{F}$ is a set-valued map given by $x \rightarrow \bar{co}(F(x))$ with $\bar{co}(A)$ is the closed convex hull of A. However, the stochastic recursive inclusion refers to the scheme

$$X_{n+1} = X_n + \epsilon_n[z_n + M_n] \tag{7}$$

where $z_n \in \tilde{F}(X_n)$. Note that differential inclusion limit in this case is one of the standard solution concepts for differential equations with discontinuous right hand sides [41].

Learning automata

Learning automata is a simple model for adaptive decision making devices that operates in unknown random environment and progressively improve their performance via a learning process. Learning automata is very useful for optimization of multimodal function when the function is unknown and only noise-corrupted evaluations are available. Such a learning model is useful in many applications involving adaptive decision making. Learning automata methods have two distinct advantages over the stochastic approximation algorithms. The first advantage is that the action space need not be a metric space because as in stochastic approximation algorithms the new value of the parameter is to be chosen close to the previous value. The second advantage is that the methods based on learning automata lead to global optimization because at every stage any element of the action set can be chosen. In general, the learning automata proceed as follows : Each time the agent randomly chooses an action from action space based on some probability distribution. Based on the feedback of this action, the agent updates its action probability distribution. This algorithm is able to choose the best action if the probability to choose this action converge to unit while the other action probability go to zero. As discussed before, a good advantage of learning automata approach is the action space need not be a metric and allows to consider complete generality of the concept of action.

In the sequel, we will present some learning automata

- **Finite action-set learning automata** In this part we present an automata with finitely many actions. Let $\mathbf{q}(n) = [q_1(n), q_2(n).., q_K(n)]^T$ where $q_i(n)$ is the probability to choose action i at step n. We have $q_i(n) \geq 0$ and $\sum_{i=1}^{K} q_i(n) = 1$. The learning algorithm for updating $\mathbf{q}(n)$ is of the form

$$\mathbf{q}(n+1) = \Gamma(\mathbf{q}(n), i(n), \beta(n)) \tag{8}$$

where Γ is some function, $i(n)$ is the action selected at step n and $\beta(k)$ is the feedback of the environment to the selected action $i(n)$. The set of possible random variable $\beta(k)$ may be discrete or continuous. However, since each action is tried many time, the learning automata may estimate the utility using this action given by

$$U_i = E[\beta(k)|a(k) = i]$$

Hence the goal of learning is to find the best action which is the one that gives maximum expected reinforcement, i.e $U^* = U_l = \max_i\{U_i\}$.

There are different types of learning algorithms that can be used by finite action-set learning automata (8). One interesting of those algorithms is the Linear Reward algorithm where the action probability vector $\mathbf{q}(n)$ is updated when the action selected at step n is a_i

$$q_i(n+1) = q_i(n) + \epsilon_n\beta(k)(1 - p_i(k)) \quad (9)$$
$$q_j(n+1) = q_j(n) - \epsilon_n\beta(k)p_j(k) \quad (10)$$

where $\epsilon_n > 0$. This learning algorithm has been extensively used as it is simple and has several nice proprieties (see [82]). The Markov process $\mathbf{q}(n)$ generated by the updating in (9)-(10) has K unit vectors as absorbing state. Although $\mathbf{q}(n)$ converges to any unit vector, but by choosing $\epsilon_n \to 0$ when n go to infinity, the Markov process $\mathbf{q}(n)$ converge to the optimal solution, i.e the unit vector e_l.

In attempting to design faster converging learning algorithms, Thathachar and Sastry [117] opened another avenue by introducing a new class of algorithms, called pursuit algorithms. As their names suggest, these algorithms are characterized by the fact that the action probability vector pursues the action that is currently estimated to be the optimal action. The estimator algorithms presented thus far update the probability vector based on the long-term properties of the environment, and no consideration is given to a short-term perspective. In contrast, the linear reward algorithm (9)-(10) relies only on the short-term (most recent responses) properties of the environment for updating the probability vector. The first step consists of choosing an action based on the probability distribution. The second step is to increase the component of whose reward estimate is maximal (the current optimal action), and to decrease the probability of all the other actions. Vectorially, the probability updating rules can be expressed as follows:

$$q_i(n+1) = q_i(n) + \epsilon_n(1 - p_i(k)) \quad (11)$$
$$q_j(n+1) = q_j(n) - \epsilon_n p_j(k) \text{ for } j \neq i \quad (12)$$

where i is the action that maximizes the total reward obtained in response to action i till n. An interesting aspect of the pursuit algorithm is that the updating of the probability actions does not directly involve the environment response and hence it is an order of magnitude faster than linear reward algorithms. The convergence property of these algorithms is given in [117]. Although most learning automata methods deal with the case of finitely many actions for the automaton,

there are also models of continuous-action-set learning automata (see [118]). If the set of action is the real line, in [118] authors considered the action probability distribution is a normal distribution characterized by its mean and variance. For this algorithm, it is shown [118] that, if the reward function is well behaved (specifically, f is continuously differentiable and the derivative function is Lipschitz), then under some assumptions in parameters and for small ϵ, $q(k)$ converges to a point arbitrarily close to a local maximum of the reward function.

Fast-learning through function approximation

An interesting approach which has been developed lately is to estimate the real value function and/or policy of the system through parametric approximations. It is thus possible to store only representation parameters throughout the optimization process instead of the whole state-action space which are often of higher dimension or cardinality. One can then consider standard optimization techniques to find the representation parameters of optimal value functions and policies. Such a method is for instance the Least-Squares Policy Iteration algorithm [71] where linear architectures are considered to approximate the value function and the computation of optimal parameters is implemented via a subspace projection technique. Another approach, Policy Gradient [15] and [63], consists in estimating the gradient of the reward function with respect to the policy representation parameter.

Learning algorithms in Games

In the previous section, we briefly indicate that the learning algorithms are useful in application that operate in unknown random environment and only noise-corrupted evaluations are available. But in the presence of multi agents, the learning problem consists of devising a learning algorithm for single agent to learn a policy in the presence of other learning agents. However the learning algorithm in the presence of multiple agents, can be viewed as a problem of a "moving target". In typical multi-agent systems, agents lack full information about their counterparts, and thus the multi-agent environment constantly changes as agents learn about each other and adapt their behaviors accordingly. The Stochastic games (SGs) has been used as a framework to study the multi-agent learning problem, where an agent tries to learn a policy in the presence of other agents. Each play of the game consists of each of the learners get a payoff from the environment and payoffs (which may be different for different learners) are random.

Game Theory traditionally studies equilibrium states in settings of full information and connectivity. However, in a large distributed network with multiple entities having limited information, it is crucial to understand both how network structure affects equilibria, and what can one expect in terms of dynamics when players are using learning to adapt their behavior. In the non-cooperative games with finite number of agents, the solution concept is the Nash equilibrium (Nash, 1951). In a Nash equilibrium, each

player effectively holds a correct expectation about the other players' behaviors, and acts rationally with respect to this expectation. Acting rationally means the agent's strategy is a best response to the others' strategies. Any deviation would make that agent worse off. Despite its limitations, such as non-uniqueness, Nash equilibrium serves as the fundamental solution concept for non-cooperative games.

There are mainly two kind of strategies for solving reinforcement learning problems. First, finding behaviors that performs well in the environment. The second is to use statistical techniques to estimate behavior of other players. Different learning algorithms are available for solving the game problem when only the noise corrupted observations are available.

Stochastic automata games

Sastry et al [91] presents a distributed algorithm which considers finite number of players having finitely possible actions and playing one every stage of the game. In fact, each player updates his strategy basing only on his current action and his received payoff after each stage of the game. The assumption that players haven't any knowledge about neither payoffs' distribution nor strategies or actual actions of others players is token. They considered learning automata models (9) for game problem and showed that the algorithm (9)-(10) converge in general to the set of unit vector with probability one. When the game has one pure Nash equilibrium, then the learning algorithm converges to that Nash Equilibrium. But it is well known that in some games, the pure Nash equilibrium does not exist but there exist always a mixed Nash equilibrium. In such situation, it converges to one of the absorbing state, i.e., unit vector. Yiping et all [129], proposed a new algorithm, called the linear reward-penalty algorithm, which is useful especially when a pure Nash equilibrium does not exist. In this algorithm, the probability updating rules can be expressed as follows:
if the action selected at step n, is i, the $\mathbf{q}(n)$ is update as

$$q_i(n+1) = q_i(n) + \bar{\epsilon}_n\beta(n)(1-q_i(n)) - \tilde{\epsilon}_n(1-\beta(n))q_i(n) \quad (13)$$

$$q_j(n+1) = q_j(n) - \bar{\epsilon}_n\beta(n)q_j(n) - \tilde{\epsilon}_n(1-\beta(n))(\frac{1}{K-1} - q_j(n)),\ j \neq i \quad (14)$$

With linear reward algorithm (9)-(10), the Markov process $q(k)$ generated by the updating in (9)-(10), converge to a unit vector. In the reward-penalty algorithm, there is always a chance that the probability of the selected action is decreased. Yiping et all [129] showed that the limiting value probability of action i, would essentially be proportional to $\frac{1}{d_i}$ where d_i is the reward probability of action i. A recent work in [129], used the reward penalty algorithm to design a stochastic power control algorithm for cellular network. Conditions when more than one stable Nash equilibrium or even only mixed equilibrium may exist are also studied. Experimental results are presented for several cases and compared with the continuous power level adaptation solutions.

Erev and Roth

The basic Erev and Roth model [39] is a reinforcement algorithm assuming that m possible actions would be taken. The algorithm associate at each action a reinforcement level $A_i(n)$ which is updated every play if the action is chosen by adding the payoff $\pi_i(n)$ of this action at this stage. We can therefore write:

$$A_i(n+1) = \{^{A_i(n)+\pi_i(n+1)}_{A_i(n)} \tag{15}$$

This reinforcement level is the main factor in the choice of the action to be played. In fact, the Erev and Roth player chooses an action with the probability:

$$P_i(n+1) = \frac{A_i(n)}{\sum_j A_j(n)} \tag{16}$$

We can easily follow that if an action is chosen, then it's reinforced and will be more likely to be chosen in future. The basic Erev and Roth model has manly two limits. The first is the experimentation issue that can be viewed as an extension of the law of effect (see [105, 22]). In fact, it's true that choices which were successful would be employed in the future, but also similar choices will be employed often as well and then the player will not rapidly converge to one choice in exclusion of all others. The second, called recency, can be viewed as interaction between the law of effect and the power low of practice. Recency means that recent experiences may be more important in determining behavior than past ones. The basic Erev and Roth algorithm use equal weight on all experiences however Recency is a robust effect as considered and observed by John B. Watson [123] and Edwin R. Guthrie [51]. These two issues can be dealt with the introduction of "forgetting" into the Erev and Roth model. This can be resumed in the modification of the following equations:

$$p_{nj}(t+1) = (1-\phi)p_{nj}(t) + E_k(j, R(x)) \tag{17}$$

where ϕ is a recency parameter and $E_k(j, R(x))$ is an experimentation function. Erev and Roth [39] studied reinforcement learning in experimental games with unique, mixed strategy equilibrium. Borgers and Sarin [20] analyzed a reinforcement learning model discussing its relationship to the replicator dynamics used in biology and gave good result on some type of economic agents. There are many studies which have focused on fictitious play and it's variant. Benaim and Hirsch [17] have been interested by convergence of strategies in games with randomly perturbed games. Fudenberg and Levine [44] summarized studies of proprieties of smoothed version of fictitious play which is quite optimal than fictitious play itself. Hart and Mas-Collel [54] studied models based on "regret" which share these proprieties. It's well known that fictitious play and regret based strategies still require greater knowledge about the game and sophistication.

Q-Learning algorithms in non-cooperative games

We can classify multi-agent reinforcement learning into two categories. In the first one, we consider heterogeneous environment where agents don't know how others agents

react. The second category assumes homogeneous environment where agents use the same algorithm, apply some equilibrium from games theory to calculate agents' policy. The multi-agent Q-Learning algorithms still simple and haven't a concrete orientation to the real world problems. In a Q-Learning algorithm, we consider a finite number of agents having a finite number of possible states. Each agent has a set of action choosing one every play. We consider also a transition function from current state basing on some actions to resultant state, and an immediate reward function for each agent. Let $Q^*(s,a)$ be the expected discounted reinforcement when choosing action a in the state s. The expected Q-Learning value is described by the following equation:

$$Q^*(s,a) = R(s,a) + \gamma \sum_{s' \in S} T(s,a,s') \max_{a'} Q^*(s',a') \tag{18}$$

The Q-value $V^*(s) = \max_a Q^*(s,a)$ is the value assuming we are choosing the action which optimize our utility function and $\pi^*(s) = \arg\max_a Q^*(s,a)$ is the optimal policy. The Q-Learning rule is :

$$Q(s,a) = Q(s,a) + \alpha(r + \gamma \max_{a'} Q(s',a') - Q(s,a)) \tag{19}$$

where α is a parameter, s is the current state, s' is the resultant state, a is the chosen action and r is the received payoff.

No-regret algorithms

One of the most popular reinforcement algorithms is the regret minimization known as the no-regret algorithm. These kinds of algorithms take a sequence of loss function or regret function l_t as input and produce as output a sequence of action a_t to play in the stage t. The main learner object is to minimize its cumulative loss L_t.

A general class of no-regret algorithms is called Φ-no-regret learning, which span the spectrum into no-internal-regret learning and no-external-regret learning. The simplest regret type is the external regret which is the difference between the actual achieved loss and the smallest possible loss, and is defined as follows: Agent can replace its sequence of action $a_1 \dots a_t$ with $\phi(a_1) \dots \phi(a_t)$, where ϕ is some action transformation that maps A into itself. We note Φ the set of these action transformations. In designing A, our Φ-regret minimizing algorithm, we assume that we have access to subroutines A' and A'' and we define the Φ-regret algorithm as:

For $t = 1, ..., T$:

1. Send transformation ϕ_t to the fixed-point algorithm A', along with accuracy parameter $\epsilon_t = \dfrac{1}{\sqrt{t}}$. Receive action a_t satisfying $\|\phi_t(a_t) - a_t\|_A \leq \epsilon_t$.
2. Play a_t; observe loss function l_t and incur loss $l_t(a_t)$.
3. Define $m_t : \Phi \to R$ by $m_t(\phi) = l_t(\phi(a_t))$.

4. Send m_t to the no-external-regret algorithm A''. Receive transformation $\phi_{t+1} \in \Phi$.

Fictitious Play

From the most used learning algorithms, we can cite fictitious play and its variants. Agents using this model behave as they are facing unknown stationary distribution of opponents' strategies. In the fictitious play mode, with a finite strategy space and payoff function, each user chooses the best response to its beliefs about its opponents, which is given by the time average of past play. Each user must know the empirical frequency vector of its past actions. This distributed algorithm has been applied to a power control game in [14]. Note that this algorithm can be used for continuous decision variables. Basing on that strategy, the player assigns to others players the follow probability that they will choose a strategy s^{-i}:

$$\gamma_t^i(s^{-i}) = \frac{K_t^i(s^{-i})}{\sum_{\widetilde{s}^{-i} \in S^{-i}} K_t^i(\widetilde{s}^{-i})} \tag{20}$$

Then, the player chooses a strategy basing on $\gamma_t^i(s^{-i})$, probability that it assigns to player $-i$ playing s^{-i}. We must remember that many fictitious play rule exist and consequently there are maybe more than one best response to a particular assessment.

Stochastic Fictitious Play

This small variation of the Fictitious play has been introduced in [44]. In order to contend discrete changes in behavior because of small changes of beliefs, authors propose that each player maximizes a perturbed utility. A typical example of perturbation is the entropy function.

Stochastic fictitious play is a variant of fictitious play in which players chooses their actions basing on a stochastic best response function. Here the payoff to each player or agent is perturbed by an i.i.d. noise that private information to that agent and at each play agents chooses a rule mapping their payoff to their strategies. Stochastic fictitious play avoids inherent discontinuity of fictitious play process in which behavior can totally change face to small mutation in data.

Replicator dynamics and evolutionary dynamics

Evolutionary game dynamics has been originated in the field of evolutionary biology providing a population dynamical method to game theory. There are many reasons for the interest on the evolutionary dynamics by the game theory. First, the replicator dynamics was originally motivated by biological evolution and some type of economic agents. Second, the gradient evolutionary dynamic which is a cultural model representing characters that are determined largely by learning rather than genetics. Both the replicator dynamics and the gradient evolutionary dynamics are composed of a

Algorithm	Info required	Equilibrium	Convergence
Linear reward automata	only his payoff	Pure	Slow convergence
Pursuit algorithm	Only his payoff	Pure	Fast convergence
Reward-penalty	Only his payoff	Mixed	Slow convergence
Erev and Roth	Only his payoff	Pure/mixed	If Nash equilibrium exists, then the play will converge with positive probability
Q-Learning	The reward matrix	Pure/mixed	If Nash equilibrium exists, then the play will converge with positive probability
No-regret	More information generally about itself	Pure/mixed	It has been proven to converge to a Nash equilibrium in any two-player, two-action game.
FP/SFP	More information about others players	Pure/mixed	Proven to converge in some type of games

Figure 2: *Reinforcement learning algorithms characteristics*

selection and mutation components. It was proved that there is also a mathematical connection between these two evolutionary models. P. Taylor and T. Day in [115] have proved in their stability study of replicator dynamics and gradient dynamics that the mean and the variance dynamics are essentially identical in both models under the assumption that the population distribution is normal.

Conclusion

Throughout this chapter we discussed over different aspects, how learning processes can be useful to automation of distributed learning agents and presented how their are developed. This theoretical basis also set the mathematical background necessary for the validation of the different approaches that we are going to develop throughout this document. Note that each algorithm obeys to some specific constraints and assumptions which makes it relevant for a particular network scenario. The major downside with the reinforcement learning algorithms and learning algorithms in general as we presented in the chapter resides in the calibration of the different parameters mainly the learning rate. This aspect is important when it comes to the design of learning algorithms for implementation on real devices.

Part II

Distributed Reinforcement learning in cellular mobile networks

Chapter 1

Learning and population game for Nash Equilibrium : (Interoperability and coexistence of two technologies)

1.1 Introduction

In the world of wireless networks, the emergence of new technologies induce that mobiles will have the possibility to have access to different technologies at the same location. Indeed, with the growing offers for 3G+ access everywhere from telephony providers, more and more users equipped with laptops, netbooks, new generation phones have access simultaneously to WiFi and 3G. Moreover, the emergence of wide area wireless networks based on WiMAX or LTE will complicate the allocation problem for mobiles. Indeed those technologies have their own advantage in terms of throughput for the mobiles.

The goals and benefit of multiple possible connectivities should satisfy the following requirement :

- Load Balancing: In current wireless networks, each mobile scans the wireless channel to detect BSs and associate itself with BS that has the strongest received signal (SNR), while ignoring its load. As user are not uniformly distributed, most of them may be associated with a few BSs while adjacent AP, may carry only light load or idle. This motivates for more efficient method to select a BS.
- Permanent and ubiquitous access: To provide an extended area via distinct access technologies. For instance, IEEE 802.11 has a typical coverage of 100 m, whereas a WiMAX or UMTS can usually a radius of over 1 km. Thus it is possible for a node to use different access technologies at different time to assure a permanent connectivity.
- Fairness: An important problem in wireless networks is rate allocation, i.e ensuring that the available network bandwidth is shared among user in a fair manner.

This type of fairness is know as max-min fairness as discussed by Bertsekas and Gallager [18].

Mechanisms for optimizing and controlling mobiles in an area where there is coexistence of wireless technologies is studied in [126] with cognitive radios WiFi cells inside a WiMAX cell. Another approach is described in [67] with the technology 802.11(e) and a central controller. Those studies propose a centralized approach. In this chapter we deal with a fully distributed algorithm for that kind of problem but in which the environment is dynamic. Actually, the dynamic is about the number of mobiles that are looking, through reinforcement learning for their best associated technology.

Reinforcement learning techniques have been first applied in a wireless network for studying optimal power control mechanism in [62]. Their distributed algorithm is based on [91] in which authors propose a decentralized learning algorithm for Nash equilibrium. The advantage of that mechanism is twofold. First, the algorithm is fully distributed. Each agent doesn't need a lot of information to update his decision, he needs only his perceived utility which depends on other players actions but which can be easily obtained (for example if we consider that the utility depends on QoS metrics like the throughput and/or the delay). Second, it has been theoretically proved that this decentralized mechanism, if it converges, converges to a Nash equilibrium. Moreover, if the non-cooperative game has some properties like having a potential (like in [89]), then this algorithm necessary converges to a Nash equilibrium. Numerous applications of this algorithm or small variants have been proposed in the literature: spectrum sharing in a cognitive network [128], routing protocols in an ad hoc network [93], repartition of traffic between operators [28], pricing [86]... Mainly, those studies consider fixed number of players. But, the algorithm takes a certain amount of time and depending on the system, the number of users playing the game can evolve very quickly. This is the case with our user-network association problem in which mobiles are moving through different wireless cells. Then the number of mobiles competing for wireless access to different technologies in an overlapping area is always evolving. This is the novelty of our approach compared to all other studies that consider a reinforcement learning for converging to a nash equilibrium.

Throughout this chapter, we consider a cell with two co-localized radio access technologies (RAT). Since in most practical scenarios, distributed algorithms are preferred over centralized ones (centralized algorithms tend to be complex and not easily scalable), we address the association problem through a fully distributed algorithm. We assume that the mobiles take alone the decision about which technology to be connected to. The utility of a user is given by the throughput perceived which depends on the number of users in the system as well as the channel condition. We model the problem as a non-cooperative game where the players are the mobiles and the strategy of a player is the choice of the technology. Whenever the system state changes (new arrivals or departures of players), every player remaining in the system, try to maximize his utility function.

1.1.1 Main contributions and organization

We develop all along this chapter a novel framework to achieve an efficient dynamic and decentralized radio resource management in heterogeneous wireless network scenarios. A description of the mobile network architecture for which our approach holds is given in section 1.2. A detailed equilibrium analysis under our game theoretical model is provided in section 1.3. Our analysis shows that there exists at most two Nash equilibria and we characterize sufficient conditions to have uniqueness of Nash equilibrium. In section 1.4, we design a fully distributed algorithm that can be employed for convergence to a pure Nash equilibrium. The major contributions in this chapter are listed as follow:

- The novelty of the proposed algorithm is that under a dynamic environment (variable number of players), this algorithm is still robust and provides fast convergence to pure Nash equilibrium.
- Under the assumptions set in this work for a seamless dynamic (described in section 1.5.1), if there is a new arrival, each mobile in the system always stays connected to a single base station which avoids repeated vertical handovers.
- We eventually study the impact of partial overlapping area on the performance of this algorithm.

We present in section 1.5 an application of the proposed mechanism in a context of Poisson games and describe how physical rates capacities can be used to design a suitable utility function. The performances of our mechanism in a dynamic environment are described with several simulations and statistics in section 1.6.2. Finally, we conclude the chapter in section 1.7.

1.2 Problem Statement

In a context of heterogeneous networks, we consider the presence of two different technologies in a common network area. This network structure has been frequently issued in the literature, and comes to meet the increasing demand of mobile communications with high data rates. We can differentiate two main scenarios of coexistence.

First, we consider a geographical area divided into two neighboring cells with different technologies as depicted in Figure 1.1. This scenario is an application of a heterogeneous system made of a wide UMTS cell and several small Wifi cells. In this scenario, the two technologies are not totally overlapped. That kind of scenario is called complementary network architecture because the two technologies are increasing the total coverage (the Wifi cells extend the coverage of wide UMTS or WiMAX cells). Users are spread into the cells and some of them are located out of the area of coexistence. Those later are connected to a single base station(technology) and have a heavy influence on the dynamic game of users inside the overlapping area.

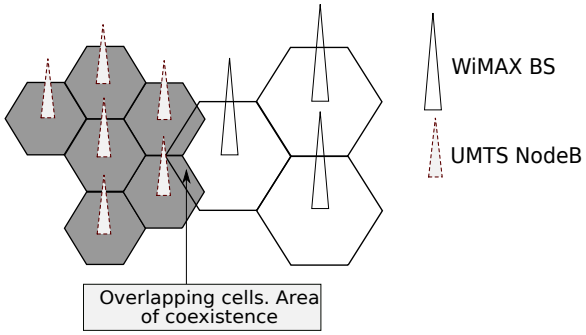


Figure 1.1: *Complementary networks architecture*

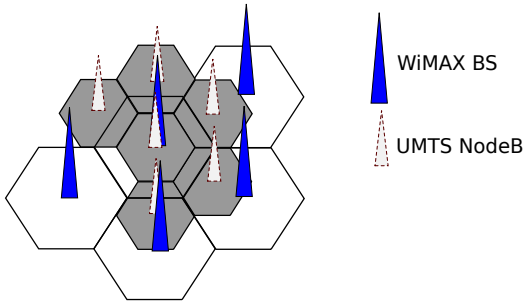


Figure 1.2: *Concurrent networks architecture*

The second scenario introduces a concurrent networks architecture: for example, the inter-networking between mobile WiMAX and HSxPA systems, with co-localized base stations as shown in Figure 1.2. Here, the mobiles are always in presence of the two coexisting technologies. Many of the radio access technologies in next generation heterogeneous networks are characterized by higher rates and frequency range coupled with a relatively low coverage in order to ensure the optimization of distribution and reuse of radio resources. This scenario is thus an application of network tight coupling architecture and can be applied in a context of small cells in restricted areas or city centers.

In both scenarios, a mobile can be faced to a decision problem. A mobile may decide on which base station or technology to be connected to. This decision can be based on individual performance. Mobile performance at each base station/technology is determined by a QoS metric which depends on the number of the mobiles connected as well as the physical rate being used by the technology chosen. Every mobile would like to find the technology that maximize his individual performance. But, as this performance depends on the actions of the other mobiles, the system can be described as a non-cooperative game.

1.3 Game theoretic Model

1.3.1 Static environment

We consider two systems : system 1 and system 2. At each slot, every mobile decides on which system to be connected to. Let N be the total number of users in the area of coexistence system. We define by n_1 (resp. n_2) the number of users connected to system 1 (resp. 2). For every user, we consider that the utility function is equal to the throughput perceived by the user. The throughput is determined by the number of users as well as physical rate being used by the technology chosen. We assume also that the mobiles connected to the same system will receive the same throughput (the game is symmetric). This means that the utility function of any mobile depends only

on the number of user in the system. This type of non-cooperative game is a congestion game [97]. Let U_i be the utility function of a user connected to the system i. As the game is symmetric, a Nash equilibrium (NE) (n_1^*, n_2^*) is given by the two conditions:

$$U_1(n_1^*) \geq U_2(n_2^* + 1) \text{ and } U_2(n_2^*) \geq U_1(n_1^* + 1) \tag{1.1}$$

The previous definition means that no user connected to a system has an incentive to move to the other system. Now we consider the following assumptions:
Assumption

- U_i is bounded and decreasing,
- there exists an integer n_i^{th} such that U_i is strictly decreasing for all $n \geq n_i^{th}$.

The last assumption immediately implies that, in any NE (n_1^*, n_2^*) such that $n_i^* \geq n_i^{th}$, at least one inequality in (1.1) is a strict inequality. We have the following proposition saying that there exists always a Nash equilibrium
Proposition 1.3.2. *For every N, there exist a Nash equilibrium.*

Proof. Without loss of generality, we assume that $U_1(1) \geq U_2(1)$. Let $n_{th} = max\{n : U_1(n) \geq U_2(1)\}$. If $n \leq n_{th}$ then the partition $(n,\ 0)$ is a NE. We first show that there exists a NE for $N = 2$. Indeed, if $U1_(2) \geq U_2(1)$, then $(2,\ 0)$ is a NE, otherwise $(1,\ 1)$ is a NE. Using the method of mathematical induction, we assume that there exists a NE (n_1^*, n_2^*) for N, and show that the above proposition is true for $N + 1$. To this end, we prove that the following relations hold :

$$\text{If } U_1(n_1^* + 1) \geq U_2(n_2^* + 1) \text{ then } (n_1^* + 1, n_2^*) \text{ is a NE} \tag{1.2}$$
$$\text{If } U_1(n_1^* + 1) \leq U_2(n_2^* + 1) \text{ then } (n_1^*, n_2^* + 1) \text{ is a NE} \tag{1.3}$$

We shall only prove (1.2), since (1.3) is symmetric. Given that (n_1^*, n_2^*) is a NE for N, then

$$U_2(n_2^*) \geq U_1(n_1^* + 1) \geq U_1(n_1^* + 2) \tag{1.4}$$

where the last inequality follows from the monotonicity of U_1. With the assumption in (1.2) it has been established $(n_1^* + 1, n_2^*)$ is a NE. □

The following proposition characterizes the number of equilibria and sufficient conditions to have uniqueness of NE.
Proposition 1.3.3. *For each N, the non-cooperative game has one or two Nash equilibrium. Furthermore,*

1 *If* $U_1(n_1^*) > U_2(n_2^* + 1)$ *and* $U_2(n_2^*) > U_1(n_1^* + 1)$, *then* (n_1^*, n_2^*) *is the unique N.E.*
2 *If* $U_1(n_1^*) = U_2(n_2^* + 1)$, *then there are two Nash equilibria :* (n_1^*, n_2^*) *and* $(n_1^* - 1, n_2^* - 1)$.
3 *If* $U_1(n_1^* + 1) = U_2(n_2^*)$, *then there are two Nash equilibria :* (n_1^*, n_2^*) *and* $(n_1^* + 1, n_2^* - 1)$.

Proof. First we start to show that for each $k \geq 2$ the partition $(n_1^* - k, n_2^* + k)$ is not a NE. Let us assume that is not true, i.e., there exists a $k \geq 2$ such that $(n_1^* - k, n_2^* + k)$ is a Nash equilibrium. Thus we have

$$U_2(n_2^* + k) \geq U_1(n_1^* - k - 1) > U_1(n_1^*) \geq U_2(n_2^* + 1) \tag{1.5}$$

This therefore contradicts the fact that the function U_2 is a decreasing function in n. In the same way, we may show that $(n_1^* + k, n_2^* - k)$ is not a NE.

(1) To this end, we shall only prove that $(n_1^* + 1, n_2^* - 1)$ is not a Nash equilibrium. Assume that $(n_1^* + 1, n_2^* - 1)$ is a NE. Thus,

$$U_2(n_2^* + 1) \geq U_1(n_1^*) > U_2(n_2^* + 1)$$

This therefore contradicts the fact that the decreasing function U_2 is a decreasing in n.

(2) We have

$$U_2(n_2^* + 1) = U_1(n_1^*) \text{ and} \tag{1.6}$$
$$U_1(n_1^* - 1) \geq U_1(n_1^*) \geq U_2(n_2^* + 1) \geq U_2(n_2^* + 2) \tag{1.7}$$

which prove that $(n_1^* - 1, n_2^* + 1)$ is a NE. It is easy to show that from equality (1.6), $(n_1^* + 1, n_2^* - 1)$ is not a NE.

(2) The proof is similar to that of (2).

□

The main focus of our analysis is to adapt a totally decentralized algorithm to a stochastic environment of players. This point is very important for our networking scenario and architecture. In such system, new users arrive according to a stochastic arrival process, and each user has a finite sized file to transmit. A user leaves the system when the entire file is transmitted. Then every mobile stays in the area of coexistence following a random amount of time. This sojourn time will depend on the throughput assigned to him. We mention that the non-cooperative game is played as a sequence of one stage static games where each one, we have proved, have at least a Nash equilibrium (see proposition 1.3.3).

1.4 Stochastic approximation algorithm

The algorithm we have used is based on a reinforcement of mixed strategies. The players are synchronized such that the decision of all players (playing a pure strategy) induce the utility perceived for each one.

In [91] we can find the original algorithm on which we based our work. It has been proved that for a fixed number of players if this algorithm converges, it will always do to a Nash Equilibrium. But in mobile telecommunications systems, mobility and

user's activity is such that the number of users evolves rapidly. Nonetheless, with two objectives in mind, we will try to apply this algorithm, with as few modifications as possible, when the number of users in the system (its state) is dynamic. The first one is to confirm whether this modified algorithm can be used to induce the system to be at the Nash equilibrium as frequently as possible. The second one is to make the algorithm as much distributed as possible, meaning that we would like to get from the base stations just the essential information on the state of the system or even no information at all.

The result is presented in Algorithm 1, in which we use the idea of individual convergence, taken from [89]. The algorithm is based on an reinforcement of mixed strategies with utility obtained by playing pure strategies.

Let N_t be the total number of users in the system at time t. Given a set of strategies $C = \{1, 2\}$, each player $p \in \{1, \ldots, N_t\}$ chooses the pure strategy $c = 1$ with probability $\beta_t^{(p)}$ (and conversely chooses the strategy $c = 2$ with probability $1 - \beta_t^{(p)}$).

As the utility perceived by user p at time t depends on his strategy as well as that of the other users, his utility function can be expressed as

$$u_t^{(p)} = \mathbb{1}_{\left\{c_t^{(p)}=1\right\}} \cdot U_1(n_1^t) + \mathbb{1}_{\left\{c_t^{(p)}=2\right\}} \cdot U_2(N_t - n_1^t) \tag{1.8}$$

where n_1^t is the number of users that chose $c = 1$ (conversely, $N_t - n_1^t$ is the number of users that chose $c = 2$)

A reinforcement learning approach is used to update the probability $\beta_t^{(p)}$ according to $u_t^{(p)} \in [0, 1]$, making:

$$\beta_t^{(p)} = \beta_{t-1}^{(p)} + b \cdot \left(\mathbb{1}_{\left\{c_t^{(p)}=1\right\}} - \beta_{t-1}^{(p)} \right) \cdot u_t^{(p)}. \tag{1.9}$$

The reinforcement learning calculations are carried out by each user until a threshold ϵ on consecutive results of probability β is not surpassed. Reaching individual convergence means that each user has no incentive on changing strategies, and there is no need to keep expending energy on calculations.

However, there are some cases where a user that has already converged, needs to restart his calculation process. In [89], even though it is not explicitly stated, it can be inferred that information about arrivals and departures is distributed to every user from a centralized entity. When this information reaches a user that has converged, his probability is reset.

We will keep the idea of having users restart calculations from a point α (conversely, $1 - \alpha$) close to zero (conversely, one) for users that have already converged to strategy $c = 1$ (conversely, $c = 2$). Instead of broadcasting the information about the state of the system to all users, we define different methods that users who already converged use to estimate when an event occurs and, accordingly, restart calculations. In other words, our algorithm is defined such that users can adapt themselves to changes in the system.

The first method involves comparing at time t, the utility obtained by a user in a window of the last n iterations, against a distinctive pattern that indicates the occurrence of an event. This pattern is composed of $n-2$ points with the same utility value and the remaining 2 points with a different one. If the latest two points in the window have a smaller value than the remaining have, we can say there has been an arrival to the technology selected by the user. If, however, these points have a bigger value, we can say there has been a departure on the technology selected by the user. Depending on the size of the window used, false positives are more or less frequently signaled.

The method for detecting changes will be applied to infer changes in both, the technology to which the user is connected, as well as the other technology. In the former, the user will detect a reduction of the available throughput that induce a possible change in strategy. In the latter, a reduction of the noise level signals a rise in available throughput and an evaluation of a change in strategy. The user's decision about making a handover will be taken using this information.

Detecting changes in noise, on the technology the user is not connected to, requires probing it, increasing energy consumption. It is for this reason that our second method tries to capture the departure rate of the system, making users that have converged to restart calculations periodically.

This policy might be costly in terms of energy consumption, as users in both technologies are going to restarting frequently, so the third policy we use does not require that a user who already converged to restart calculations, leaving the choice of technology to users arriving to the system.

After a departure, the system might be left in a different point than the Nash equilibrium, until the arrival of new users returns the system to equilibrium. We will evaluate the impact of this policy in the system's performance, but we expect faster convergence as well as a drastic reduction in the number of vertical handovers.

Algorithm 1 Dynamic Distributed Algorithm.

1. Initialize $\beta_{t-1}^{(p)}$ as starting probability for new players in P.
2. For each player p:
   (a) If player p has converged and restarting conditions are met, then:
      i. If $\beta_{t-1}^{(p)} \approx 1$ set $\beta_{t-1}^{(p)} = 1 - \alpha$.
      ii. If $\beta_{t-1}^{(p)} \approx 0$ set $\beta_{t-1}^{(p)} = \alpha$.
   (b) If player p has converged and restarting conditions are not met, move to player $p+1$
   (c) Player p performs a choice, over C, according to $\beta_{t-1}^{(p)}$.
   (d) Player p updates his probability $\beta_t^{(p)}$ according to his choice using (3.13).
   (e) If $\left|\beta_t^{(p)} - \beta_{t-1}^{(p)}\right| < \epsilon$ then player p has converged.
3. Remove players that departed, make $t = t+1$ and go to step 1.

In the next section we will apply and recall the results obtained in the general setting

to the concept of Poisson games and in a realistic setting of coexistence between two wireless technologies.

1.5 Applications

1.5.1 Poisson game

As before let our system modeled as a non-cooperative game between finite number of players. At each slot, every user decides on which base station/technology to be connected with. The throughput received by each user depends then on the number of mobile connected to his base station. We assume that the total number of user in the area of coexistence at time t is N. We define by n^* the number of users connected to WiMAX at that instant. For every user i, we consider that the utility function is equal to the throughput perceived by the user. For simplicity of analysis we assume that the throughput in one technology is equal to a constant (s_1 for WiMAX and s_2 for UMTS) divided by the number of users connected to. As a consequence, let's denote by $N_t = \{n_t^{c_0}, n_t^{c_1}\}$ players performing different actions, over the set $C = \{c_0, c_1\}$ at time t. Considering for each technology a constant total throughput in the set $S = \{s_{c_0}, s_{c_1}\}$, the choice performed by user p at step t will result in an individual utility given by:

$$u_t^{(p)} = s_{c_t^{(p)}} / \left(n_t^{c_t^{(p)}} \cdot \max(S) \right). \tag{1.10}$$

This type of non-cooperative game is a congestion game. It has been firstly studied in [97] where it is proved that this game has almost one nash equilibrium. A repartition $(n^*, N - n^*)$ is a nash equilibrium of this game if any user has no interest to change unilaterally his decision. As the game is symmetric, no distinction is made between users, the two necessary and sufficient conditions for a repartition n^* to be a nash equilibrium are

$$\frac{s_1}{n^*} \geq \frac{s_2}{N - n^* + 1}, \tag{1.11}$$

(none users from the first set, connected to the base station 1, has an incentive to move to the other base station) and

$$\frac{s_2}{N - n^*} \geq \frac{s_1}{n^* + 1}, \tag{1.12}$$

(conversely, none users from the second set has an incentive to move to the first base station). From [97], we know that there exists almost one Nash equilibrium in our congestion game. We have the following result saying that there are at most 2 nash equilibria.

Proposition 1.5.2. *Our congestion game has almost one Nash equilibrium and at most two Nash equilibria.*

Proof. Combining the two necessary and sufficient conditions 1.11 and 1.12 we obtain:

$$\begin{cases} \frac{s_1}{n^*} \geq \frac{s_2}{N-n^*+1} \\ \frac{s_2}{N-n^*} \geq \frac{s_1}{n^*+1} \end{cases} = \begin{cases} s_1\left(N-n^*+1\right) \geq s_2 n^* \\ s_2\left(n^*+1\right) \geq s_2\left(N-n^*\right) \end{cases}$$
$$= \begin{cases} n^*\left(s_1+s_2\right) \leq s_1 N + s_1 \\ n^*\left(s_1+s_2\right) \geq s_1 N - s_2 \end{cases}$$

which will mean that:

$$\frac{s_1 N - s_2}{s_1+s_2} \leq n^* \leq \frac{s_1 N + s_2}{s_1+s_2}.$$

If we make $\alpha = \frac{s_1 N - s_2}{s_1+s_2}$ and $\beta = \frac{s_1 N + s_2}{s_1+s_2}$, then $\beta - \alpha = 1$. If α is not an integer, there is always an integer between α and β, which corresponds to n^*. If α is an integer, then α and β are both Nash equilibria. □

Then we have proved the existence and the non-uniqueness of Nash equilibrium in our system. Our aim is to propose a totally decentralized mechanism that converges to these nash equilibrium, in a stochastic environment. First of all, we define the metrics which will be used for evaluate the performance of our algorithm.

Description of environment dynamics The main focus here is to adapt a totally decentralized algorithm to a stochastic environment of players. This point is very important for our networking scenario and architecture. In particular we consider that mobiles are moving and enter in the coexisting area following a Poisson process with rate λ. Moreover, their sojourn time in this area of coexistence technologies is assumed to be exponentially distributed with average $1/\mu$. Note that the sojourn time of a mobile in the system does not depend on the throughput. Our system can be modeled as a $M/M/\infty$ queue and then the number of mobiles in the system follows a Poisson process with average $\rho = \lambda/\mu$.

That kind of stochastic environment has been considered for auction mechanism in [75, 109] where users come into the system following a Poisson process and leave it after an exponentially distributed sojourn time. They have shown that this stochastic environment induces very different results in the auction process.

In Poisson games [79] the number of players is a random variable following a Poisson process. Those games have at least one Nash equilibrium by applying Kakutani fixed-point theorem when actions set and types of players are finite and utility functions are bounded. The main difference with our analysis is that in [79], the number of players, which is a Poisson random variable, is a common knowledge for every player.

1.5.3 Utility function for WiMAX and UMTS physical rates

In this section we extend our analysis to the use of WiMAX and UMTS, which at the end will only help for setting the network model to which the theory developed here will apply. We now consider a fixed capacity s_w (resp. s_h) for the WiMAX (resp. HSDPA) system. In WiMAX systems, the available throughput is gradually shared between users, depending on the number of available sub-carriers. Considering there is no inter-cell interference, we assume that the global system capacity s_w is constant. The utility $u_t^{(w)}$ perceived by every user in WiMAX at time t is given by:

$$u_t^{(w)} = s_w / \left(n_w^t \cdot \max\left(s_w, s_h\right)\right). \tag{1.13}$$

On the other hand, in HSDPA systems, the throughput per user is mainly affected by intra-cell interferences, created by the increasing number of users connected to the system. The type of modulation and coding scheme used, inter-cells interference and the distance factor are also taken into consideration. The throughput allocated to each user p, is computed for a given value of the signal to interference-and-noise ratio (SINR), estimated by:

$$SINR_p = \frac{g_p * P_p}{\sigma^2 + \sum\limits_{j \neq p} g_j P_j \delta_{jp}} \tag{1.14}$$

where inter-cells interferences are not considered, g_p, P_p respectively the channel gain and transmission power of user p, δ_{jp} the orthogonality factor and σ^2 the additive background noise. The throughput for user p is then,

$$T_p = R_p f(SINR_p), \tag{1.15}$$

with R_p the user's transmission rate and $f(.)$ the cumulative distribution function of the efficiency [23] estimated by $f(SINR_p) \simeq (1 - \exp(-SINR_p))^M$, where M is the packet length.

We will assume that the users are identical and the global throughput is a decreasing function of the number of users present in the system, that will be shared between those users. The utility perceived by each user connected to the HSDPA system at time t is $u_t^{(h)} = s_h f(SINR_{n_h^t}) / \left(n_h^t \cdot \max\left(s_w, s_h\right)\right)$.

In the next section, we then observe by simulation how behaves our distributed algorithm in a dynamic environment where there exists a coexistence area between two wireless wireless technologies.

1.6 Simulations

As presented so far in the previous sections we will study in the simulations and numerical analysis the coexistence between the WiMAX and HSDPA technologies. Given this configuration of the simulations, we will look at the impact of the variation of the

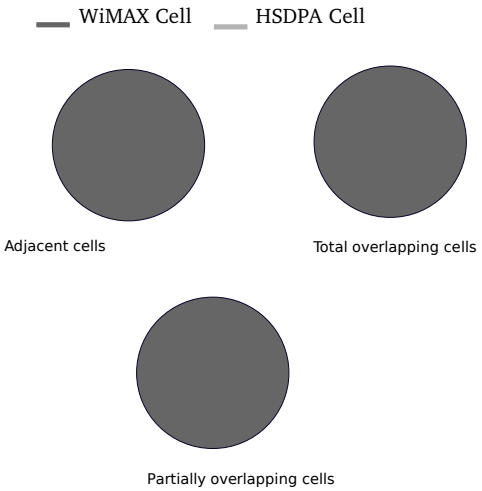


Figure 1.3: *Joint coverage area*

overlapping region between the different technologies on the convergence of the algorithm. Moreover, we will consider different proportions of the coexistence area, from the superposition of the cells to near adjacent cells. Another consideration will be to observe the impact of the arrival rate on the performance in term of convergence to Nash equilibrium, when the load of the system is constant. First of all, we define the metrics which will be used for evaluate the performance of our algorithm.

1.6.1 Metrics

For the performance evaluation of the algorithm we will use some metrics that will allow us to compare the different mechanisms inside the algorithm in terms of both, convergence and computational cost. To this purpose we calculate, over a sliding non overlapping window h of 900 iterations of each simulation, a weighted average of the samples obtained using the kernel smoothing method:

$$KS(t) = \frac{\sum_i K\left(\frac{t-i}{h/2}\right) f(i)}{\sum_i K\left(\frac{t-i}{h/2}\right)}; K(u) = \frac{3}{4}\left(1-u^2\right) \cdot \mathbb{1}_{\{|u|\leq 1\}}, \tag{1.16}$$

where t is the point in the middle of the window and $i \in h$ are the iterations in the window.

Convergence will be evaluated through the cumulative convergence to Nash equilibrium, defined by:

$$CC(t) = KS(t) : f(i) = \mathbb{1}_{\{NE\}}$$

i.e. the average proportion of iterations the system is at the Nash equilibrium in the window whose middle point is the iteration t.tells in the long run the proportion of time the repartition is a nash equilibrium.

Computational cost will be evaluated through the percentage of users in each simulation that, at iteration t, have not reached individual convergence and, hence, are performing calculations:

$$UC(t) = KS(t) : f(i) = \frac{\sum \mathbf{1}_{\{p \in \mathfrak{C}_i\}}}{N_i}, \tag{1.17}$$

where $\mathfrak{C}_t$ is the set of users running the reinforcement learning algorithm at each iteration $i \in h$, and N_i is the number of users in the system at each iteration $i \in h$. Therefore, UC measures the average proportion of users running the reinforcement learning algorithm in the window whose middle point is the iteration t. This metric allows to compare how often the policies make the users perform the reinforcement learning algorithm. A good balance between both metrics is desired. Finally, a normalized entropy metric will be calculated to see how the changes of state affect the performance of player p at iteration t:

$$Q_t^{(p)} = \frac{\sum_{m=0}^{w-1} \frac{\sum_{n=m+1}^{w} (m-n) \cdot \text{sign}\left(u_{t-m}^{(p)} - u_{t-n}^{(p)}\right)}{\sum_{j=m+1}^{w} (m-u)}}{w-1}, \tag{1.18}$$

where w is the size of the window (number of slots) to evaluate the entropy for player p, u_i^p is the normalized utility for player p at the beginning of slot t and:

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0\,. \end{cases} \tag{1.19}$$

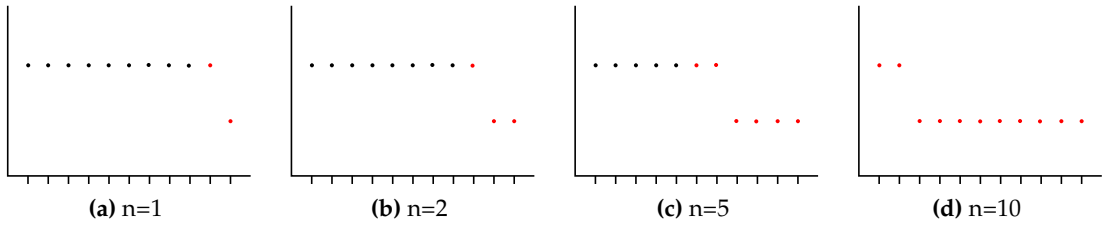
1.6.2 Simulation scenario

In our model, we have considered that users who are trying to send files with exponentially distributed sizes, arrive following an exponentially distributed inter-arrival time rounded to the closest iteration. There are no simultaneous arrivals, since the iteration when the next arrival happens, is calculated with each new arrival. This first set of simulations scenarios has the following structure:

1. Change of state detection (Case):

   Case 1 - Original algorithm: as control case we used the original algorithm from [91], in which changes of state are not detected and individual users never stop calculating because convergence is taken globally.

   Cases 2, 3, 4 and 5 - Pattern of n-iteration memory: as shown in Figure 1.4, we have used predefined patterns to detect the case of arrivals. We will take the set of $n+1$ red dots, where the rightmost dot is the most recent value of utility a user has. Starting from that point we would like to know if there has been a drop on the performance obtained n slots before, and if this drop in performance has been recurrent. For departures we have used the same strategy, but with the mirror



(a) n=1 **(b)** n=2 **(c)** n=5 **(d)** n=10

Figure 1.4: *Patterns for n-iteration memory (arrivals).*

pattern, i.e., "lows" become "highs" and viceversa. We evaluated patterns for 1, 2, 5 and 10 iteration memory.

Case 6 - No restarting after individual convergence: there is no restarting for changes of state.

Case 7 - Restarting on changes of state: information about changes of state is broadcasted by the base station to each user that has converged so they can restart calculations.

Case 8 - Entropy detection with $w = 5$: players that have already converged evaluate changes in the system through (1.18). If $\left|Q_t^{(p)}\right| > \tau, \forall \tau \in [0, 1)$ a change of the state has been detected an the user should restart. In our simulation we have used $\tau = 0.8$.

2. Starting probabilities for new users were set according to two strategies: a) using random values from a uniform distribution ; and b) using always 0.5.
3. Users that needed to restart calculations followed one of two strategies: a) restarting all of them; and b) restarting them with probability $1/P_{jt}$.

This simulation scenario is composed of 120 simulations, one for each set of conditions. For each simulation, 25 independent runs were made. Each independent run was composed of 2500 iterations of the algorithm.

The arrival and departure process can be modeled as an M/M/∞ queue, where ρ will be the average number of users in the system. For all simulations we will use $\rho = 5 = 3000/600$.

We will start simulations with 5 users. The maximum available througputs for each technology will be fixed at $S = \{s_u = 1, s_w = 3\}$, where s_u is the maximum available thorughput for UMTS users and, conversely, s_w will be the one for WiMAX users. We have picked an acceleration parameter $b = 0.3$, a convergence threshold $\epsilon = 10^{-4}$ and restarting probability $\alpha = 0.3$.

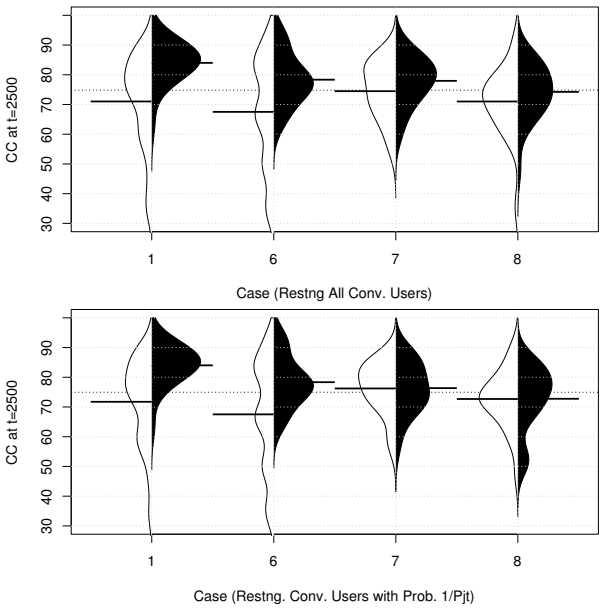


Figure 1.5: *Distribution of CC_{2500} for simulations of Cases 1-6-7-8, $\rho = 3000/600$, split by number of users that restart after a change of state (with probability $1/P_{jt}$ on bottom), and starting probability: random on the left (white) side, 0.5 on the right (black) side).*

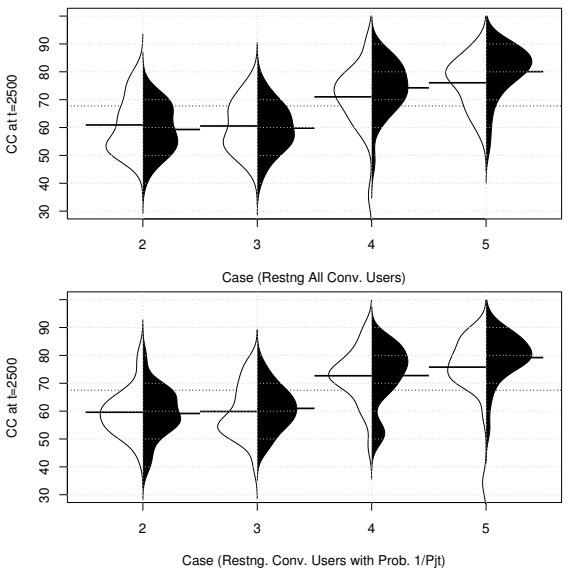


Figure 1.6: *Distribution of CC_{2500} for simulations of Cases 2-3-4-5, $\rho = 3000/600$, split by number of users that restart after a change of state (with probability $1/P_{jt}$ on bottom), and starting probability: random on the left (white) side, 0.5 on the right (black) side.*

1.6.3 Firsts results and analysis

In Figure 1.6 we plotted the distribution of cumulative convergence at $t = 2500$ (CC_{2500}) for the independent runs made of simulation cases 2, 3, 4 and 5, with $\rho = 300/310$, being this the mid-level rate we studied. On the top plot we have simulations when all the users that converged restart after a change of state detection, and on the bottom plot we have those where users only restart with probability $1/P_{jt}$. The left (white) side of each case shows the results for random starting probability used for new players and the right (black) side those when new players use 0.5 as starting probability. The black dotted line shows the global average, while the thick black lines show the average for that particular simulation. As we can see, averages on the left (white) side of the plot are always smaller than those on the right (black) side, meaning that fixing starting probability for new users with a value of 0.5 leads to better convergence rates. Also, we can see that the bigger (case 2 is the smallest, case 5 is the biggest) the pattern used for a change of state, the better cumulative convergence levels we achieve. This plot leads us to use only case 5 for further analysis.

Figure 1.5 shows us the same plot but with cases 1, 6, 7 and 8. Again, performance obtained when assigning new players a random starting probability is always worst than when they are assigned a fixed starting probability. This will lead us to remove the case of random starting probabilities from further consideration. On the other hand,

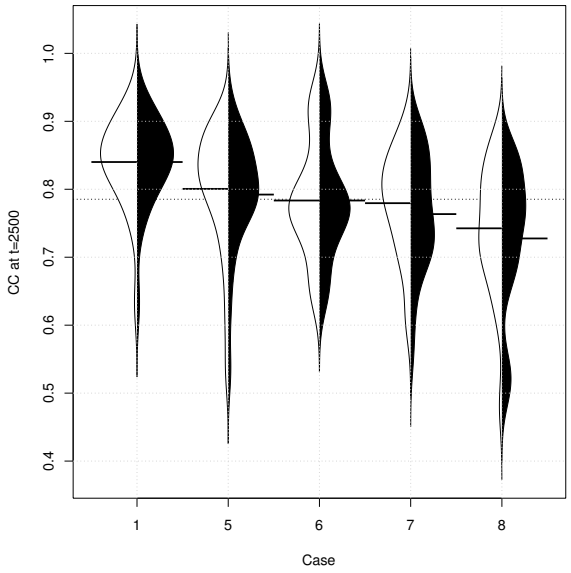
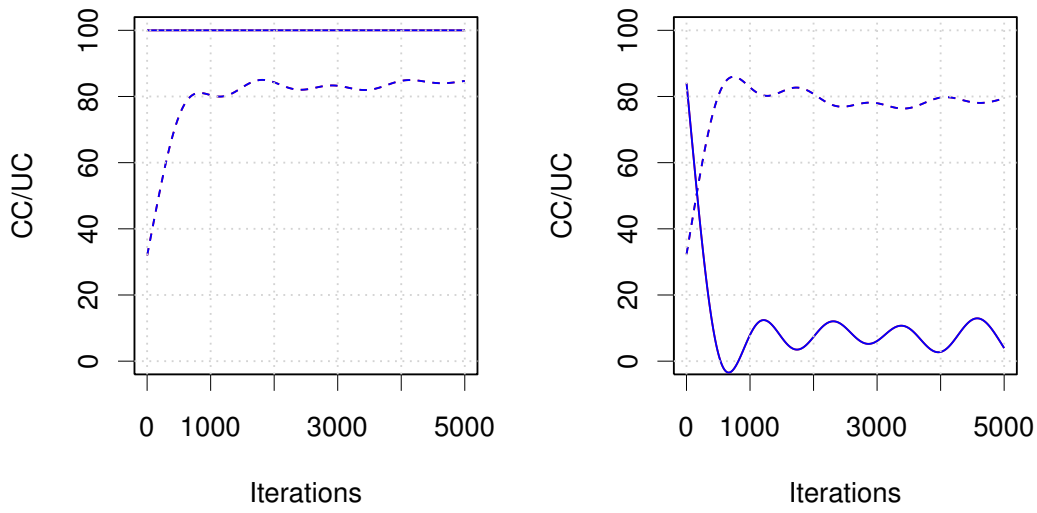


Figure 1.7: *Distribution of CC_{2500} for simulations of Cases 1-5-6-7-8, $\rho = 3000/600$ and fixed starting probability, split by number of users that restart after a change of state (all, on the left (white) side, with probability $1/P_{jt}$ on the right (black) side).*

this plot would not allow us to discard any other simulation case.

In Figure 1.7 we can see, for the remaining cases, the cumulative convergence fixing the starting probability at 0.5. Simulations when all the users that converged restart after a change of state detection are on the left (white) side and those where users restart with probability $1/P_{jt}$ are on the right (black) side. There seems to be a marginal difference on CC_{2500} between both strategies within each case that use restarting criteria (cases 5, 7 and 8), being better the strategy of restarting all converging users. Nonetheless, this could lead to a bigger proportion of users making calculations at any given time, so we should explore UC_t. From this same figure we can discard case 8 as it offers lower performance that the other cases. The original algorithm (case 1) seems to be the best of the group in terms of cumulative convergence, but there is, as we can see later, a big drawback to this performance.

In Figures 1.8 and 1.9 we present the average over the 25 independent runs for both CC_t (dashed lines) and UC_t (solid lines) for cases 1, 5, 6 and 7. We have also plotted the strategies for restarting users as red lines, when we restart all converging users, or as blue lines, when they are restarted with probability $1/P_{jt}$. As we can see, there are no significant differences in cumulative convergence for the different cases, but case 6 achieves the same level of cumulative convergence with a lower proportion of users calculating at any given time. This means that after users select one of the two technologies, they should keep their choice for as long as their call last. This can pose a problem at departures when both the service and arrival rates are small, because converging users will not be restarting and they could end in a partition that is not a Nash equilibrium for a long time. Nonetheless. this might be a good deal to have if



(a) Case 1. **(b)** Case 6.

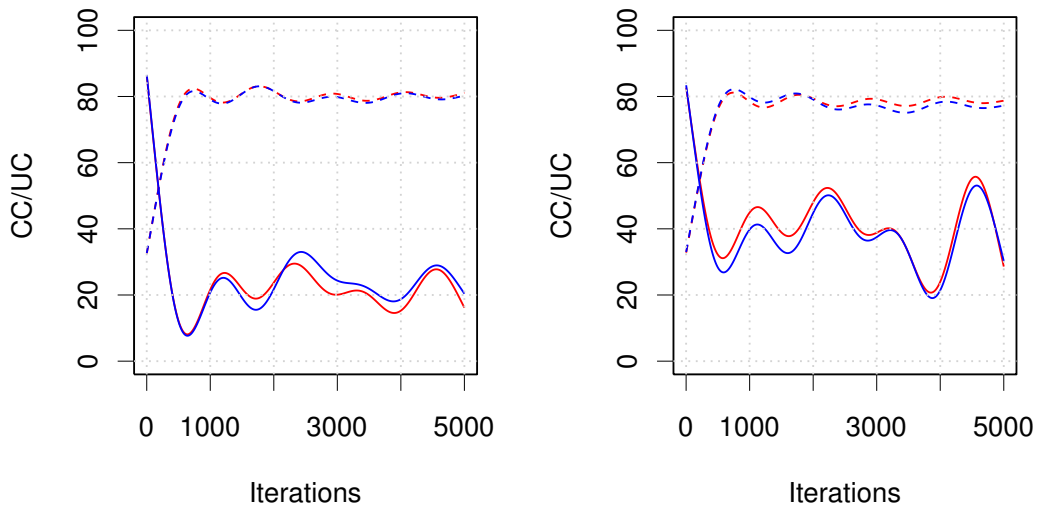
Figure 1.8: *Evolution of CC_t (dashed lines) and UC_t (solid lines) for $\rho = 3000/600$ and starting probability fixed at 0.5. No restarting strategies are considered for these cases.*

arrival rates are high, as this will keep an steady flow of new users who will always move the partition to a Nash equilibrium.

Summary From our first set of analysis we can conclude that the best strategy to follow for arrivals and departures seems to be keeping the probabilities achieved for the previous state and using them as starting probabilities for the new state, leaving players that already reached convergence in the technology originally selected. This strategy doesn't affect convergence and reduces the proportion of users calculating to achieve the Nash equilibrium.

In the following of our numerical investigations, we will go through extended analysis with the best scenario obtained in previous simulations. We then refine the scenarios and select Case 5 as Case A, an enhanced version of Case 5 as Case B as will be defined later and Case 6 as Case C. Under our new settings, the sojourn time will depend on the throughput assigned to each user, which means they are being served by a M/M/1 queue with Processor Sharing (PS) discipline. This simulation scenario has the following structure:

1. Change of state detection (Case):
    - Case A: In this case, every user that have converged is actively detecting the restarting pattern on either of the two technologies using a window of 10 iterations.
    - Case B: In this case, users actively detect the restarting pattern using a window of 10 iterations on the technology in which they are connected with a forced periodic restarting that follows the departure rate. In this case, users



(a) Case 5. **(b)** Case 7.

Figure 1.9: *Evolution of CC_t (dashed lines) and UC_t (solid lines) for $\rho = 3000/600$ and starting probability fixed at 0.5. The strategies for restarting users are shown as red lines, when we restart all converging users, or as blue lines, when they are restarted with probability $1/P_{jt}$.*

that detect the pattern will restart with probability 0.5.

- Case C: In this case users do not restart after individual convergence.

2. We used three different sets of rates with constant load $\rho = 0.95$, by taking $\lambda \in \{1/6000, 1/4500, 1/3000\}$ and setting μ accordingly.
3. The percentage of overlapping area of both technologies to the total area covered by both base stations, α, was taken in {25%, 50%, 100%}.

For each simulation, 10 independent runs were made. Each independent run was composed of 1000000 iterations of the algorithm.

We started each simulation with 5 users. The maximum available throughput for each technology is fixed at $S = \{s_h = 2.5, s_w = 3\}$. We have picked an acceleration parameter $b = 0.3$, a convergence threshold $\epsilon = 10^{-6}$ and restarting probability $\alpha = 0.5$.

1.6.4 Final results and analysis

First, we will see how the different strategies to restart calculations work. Figure 1.10 shows the behavior of users who follow the pattern of utility to detect changes in the state of the system (Case A). On the left side (iteration 18600), users in WiMAX detect a change on the system (most probably an arrival on the non-overlapped region covered by the WiMAX base station) and restart calculations, converging most of them in less than 500 iterations. On the right side (iteration 20800), there is a detection of a change in the system by the users that are using HSDPA. They restart accordingly and then one

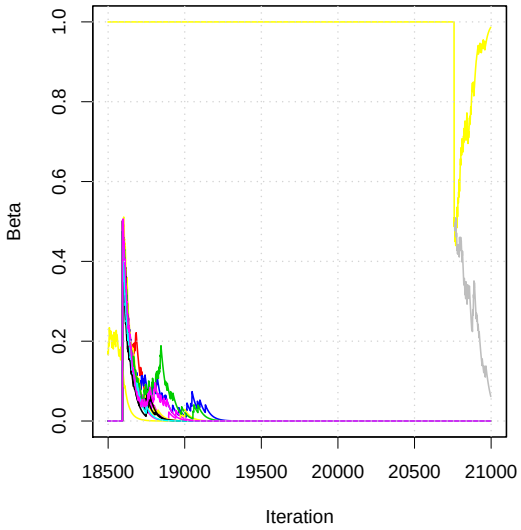


Figure 1.10: *Users in a simulation using as strategy the detection of changes on both technologies (Case "A").*

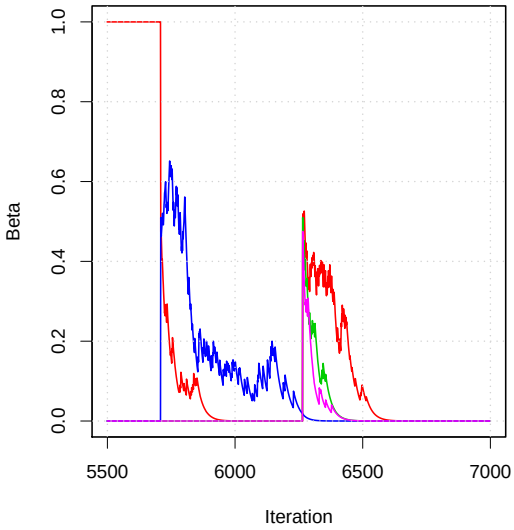
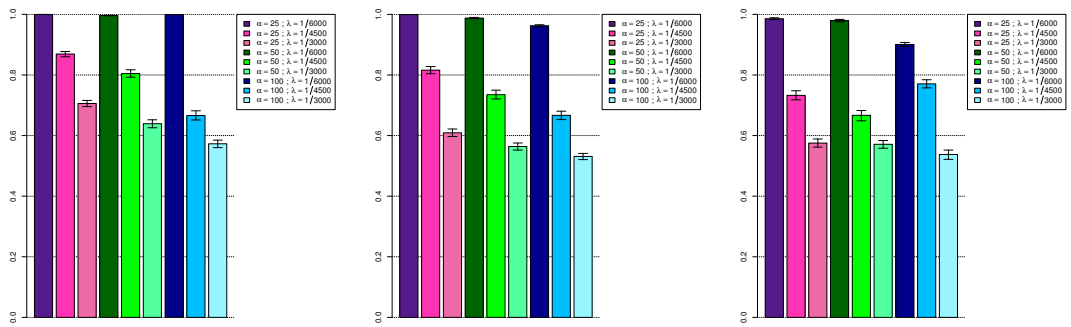


Figure 1.11: *Users in a simulation using as strategy the detection of changes on the technology they are connected to and restart periodically to follow the departure rate (Case "B").*

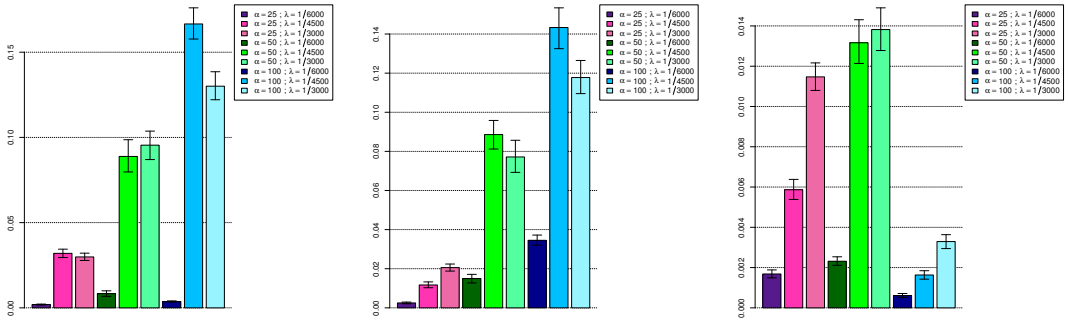
chooses to stay in HSDPA and the other picks WiMAX.

An interval of a simulation of Case B is shown on Figure 1.11. Users restart when feeling changes of utility on the technology they are using, but also some of them restart periodically following the departure rate. Here, we can see that the blue and red users restart when they reach the time mark (the period is completed), and both of them go to WiMAX. Later, between iterations 6000 and 6500, there seems to be a change in the system (again, probably an arrival on the WiMAX non-overlapped region), that triggers a restarting flag for all users that converged to WiMAX. We see that the blue user is not affected by this event.



a) Case A b) Case B c) Case C

Figure 1.12: *Average proportion of iterations that a simulation is in the Nash Equilibrium, with 95% bootstrap confidence intervals.*



a) Case A b) Case B c) Case C

Figure 1.13: *Average proportion of users performing the reinforcement learning algorithm at each iteration, with 95% bootstrap confidence intervals.*

On Figure 1.12 we can see that $CC(t)$ is inversely proportional to both α and the rates, but is not very sensitive to the policy used in the algorithm. The smallest the overlapped area, the smallest the number of users playing the game and, therefore, the better performance obtained. The same behavior is observed with the change of rates, but in this case higher rates imply faster changes in the state of the system and, therefore, worst performance is achieved. As for the policy used, the best performance is attained making users feel changes of state in both technologies (Case A), while letting users that have converged stay in the selected technology for the remainder of their call (Case C) offers the worst performance in general, although margins of difference are not quite dramatic.

On the other hand, differences in $UC(t)$ for changes in any of the parameters are big (see Figure 1.13). For policies 1 and 2, the biggest the overlapped area, the biggest the proportion of users performing calculations at any given iteration. Policy 3 does not follow this pattern of behavior, but this might be due to the difference of at least one order of magnitude with respect to the other two policies, which makes absolute differences very small. Another cause for the differences might be that the number of users in the overlapped region is small in with $\alpha = 25\%$, but they are responsible to get the system to the Nash equilibrium, making them more active and taking them longer to converge. As the size of the region increases ($\alpha = 50\%$), the number of users grows proportionally, but they are not able to easily converge. Finally, when the area is completely overlapped ($\alpha = 100\%$), all users can play the game and converge quickly, leaving the responsibility to make the system reach the Nash equilibrium to new users. Changes in rates do not create the regular patterns of changes for $UC(t)$ that were observed in $CC(t)$.

1.7 Conclusion

In this chapter we have studied the problem of access network selection under a network framework of coexistence between the WiMAX and UMTS(R-8). For the two con-

sidered RATs the nominal capacities are similar and the rate obtained by each node attached to a given cell depends not only on the channel condition but also and mainly on the load of each cell. This fact has allowed us to build a model of competition between the mobile users of the network, with the objective to select the best access. This defines a classical problem of access selection for user-network association. Along this chapter, we have modeled the problem as a game where players are the mobiles nodes and their strategies are selection of either of the two technologies. We studied the problem under a game theoretical setting and defined the possible equilibria. We then defined the policies by which mobile players can attain the different equilibria in a distributed fashion. To do so, we made use of reinforcement learning techniques. We designed a learning algorithm that would be implemented on board of mobiles and allow to reach the desired equilibria. We have also analyzed the impact of overlapping areas giving rise to an extra load of users located out of the study area of competition. We have explored several scenarios and obtained the following results under adaptation to Poisson Games:

- Frist, it is possible to follow the basic algorithm used in this work, originally developed for a fixed state (number of players in the system), and use it for a dynamic environment (variable number of players due to an $M/M/1$ queue with PS discipline) with excellent levels of convergence to the Nash equilibrium (above 90%) in scenarios with low rates of arrivals and departures ($\lambda = 1 = 6000$), independently of the size of the overlapped area.
- However, the algorithm decreases its performance with more users, due to higher motion rates or bigger overlapped area. Nonetheless, convergence to the Nash equilibrium is not very sensitive to the policy used in the algorithm.
- The policy of not doing anything in the case of arrivals clearly has the best performance with respect to the proportion of users calculating at any given iteration (one order of magnitude smaller than the other policies), which coupled with its fairly good convergence (close to the other policies) makes it the best strategy tested.

In summary, a totally decentralized algorithm with no information broadcasted by the base station, can be used by each user to collectively reach a Nash equilibrium reducing the total number of vertical handovers performed. However, note that other reinforcement learning algorithms can be used in the same framework developed in this chapter, such as fictitious play, Q-learning, no-regret learning and others.

Chapter 2

Hierarchical game and reinforcement learning without memory

2.1 Introduction

Recently, the use of Self Organizing Network (SON) features in a framework of general policy management has been suggested. In such frameworks SON entities are used as a mean to enforce high level operator policies, introduced in the management plane, and are translated into low-level objectives guiding coordinated SON entities [3]. Among the most important self-optimization mechanisms in Radio Access Networks (RAN) are interference coordination [111], mobility management, and energy saving [84]. Several such problems need further investigation to fully benefit from SON in RAN, in areas where little material has been published. Examples are autonomous cell outage management, and coverage capacity optimization [27]. It is noted that the problem of coordinating simultaneous SON processes is an open and challenging problem that needs to be addressed in order to allow the deployment of SON mechanisms.

We propose in this chapter a self-optimization framework for Inter-Cell Interference Coordination (ICIC) in an orthogonal frequency division multiple access (OFDMA) network. Inter-cell interference can dramatically degrade cell performance and perceived Quality of Service (QoS), particularly at cell edge. We are interested in distributed solutions that can be implemented in a flat architecture (e.g. LTE-Advanced architecture). To coordinate interference between neighboring cells, eNodeBs need to exchange information. In the case of LTE for example, signaling between eNodeBs can be exchanged over the *X2* interface (see Figure 2.1). Recent studies such as fractional frequency reuse [47, 107] and soft frequency reuse [2] allowing users in different channel conditions to benefit from different reuse patterns have been proposed. Still, all of these schemes mentioned above are static interference management approaches, where a specific reuse pattern is predetermined a priori by a network operator at off-line.

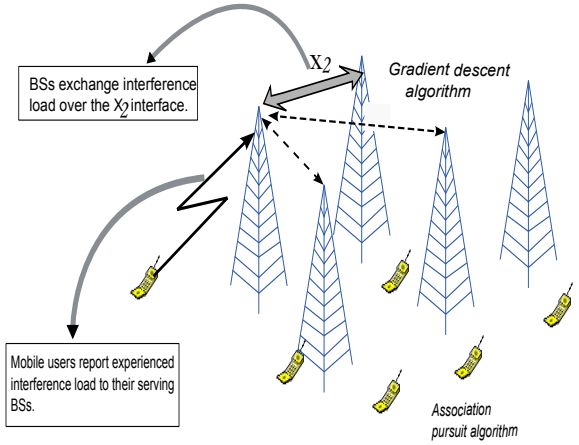


Figure 2.1: *The multi-cell network model.*

Specifically, we assume that the fractional frequency-reuse (FFR) of a cell can be configured dynamically. In that case, some base stations (BSs or eNodeBs) would be enabled to adjust their FFR in order to provide coverage/capacity for other neighboring cells. We further model the network behavior as a Stackelberg game between the network manager and the mobile users using the game theory framework [45].

At the core lies the idea that introducing a certain degree of hierarchy in non-cooperative games not only improves the individual efficiency of all users but can also be a way of reaching a desired trade-off between the global network performance at the equilibrium and the requested amount of signaling. The proposed approach can be seen as intermediate scheme between the totally centralized policy and the non-cooperative policy. It is also quite relevant for flexible networks where the trend is to split the intelligence between the network infrastructure and mobile users' equipments. In the Stackelberg game, the network manager is acting as the leader and mobile users as the followers. In the first stage, the leader chooses its strategy profile and announce it to the followers. Then, the followers decide their respective outcomes depending on the strategy profile of the leader. Under our scenario, the network manager maximizes the total network throughput by means of power control and announces its strategy profile to mobile users. Each mobile will decide *individually* to which of the available base stations it is best to connect according to its radio condition and the strategy profile broadcasted by the network.

We also propose a two-stage self optimization algorithm for both the leader and the followers. The objective is to achieve dynamically an efficient frequency-reuse pattern based on their past experience and their learning capabilities. The leader's algorithm is based on stochastic gradient descent algorithm which requires some information to be exchanged between neighboring base stations. For user association, we propose an iterative distributed algorithm based on automata learning mechanisms. Both algorithms have been shown to converge to the Stackelberg equilibrium while providing

substantial gain compared to optimal solution and fixed full reuse scheme.

We further explore the case when the network environment is dynamic. By dynamic we mean that the number of users varies in time with mobiles arriving and departing the system. Through extensive simulations based on a realistic network setting, the proposed approach is shown to be robust and scalable. In this latter setting, we also give some insight on how designing a trade-off between the global network performance at the equilibrium and the requested amount of signaling.

2.1.1 Main contributions and organization

The original contributions in this chapter are presented as follows:

- Investigating fractional frequency reuse technique for inter-cell interference coordination in an OFDMA network,
- Modeling the interaction between the network and mobiles using a Stakelberg game framework,
- Proposing a hierarchical algorithm that allows convergence towards the Stackelberg equilibrium,
- Exploring the robustness of the proposed approach with time-varying number of users
- Giving some insight on the ways of finding a desired trade-off between the desired global network performance and the amount of control feedback,
- At the equilibrium, our mechanisms achieve up to 90% of the optimal association policy,

The chapter is organized as follows. The system model is exposed in Section 2.2.1. Section 2.2.2 provides a description of the network scenario adopted throughout the chapter. In Section 2.3, we present the game theoretic framework and propose formally how the network manager and mobile users can obtain their respective equilibria by means of a Stackelberg formulation. In Section 2.4, the proposed hierarchical algorithm is investigated for both the leader and the followers. In Section 2.5, simulation results under realistic wireless network settings are shown to exhibit interesting features in terms of self-optimizing deployment for inter-cell interference coordination. Section 2.6 eventually concludes the chapter.

2.2 Scenario Description

2.2.1 The System Model

Consider the downlink of a multi-cell system, operating in an OFDMA context giving rise to an inter-cell interference phenomenon. Power control is used by the base sta-

tions in an effort to preserve power and to limit interference and fading effects. With the same goal of maximizing their payoff, mobiles users try to connect to the best serving cell. Specifically, we consider $\mathcal{M} = \{1, ..., M\}$ as the set of all possible serving base stations (or cells) within the network and $\mathcal{K} = \{1, ..., K\}$ a set of K mobile users randomly distributed over the network. Each cell operates in a multi-band context with $N \geq K$ physical resource blocks (PRB). Let $\mathcal{N} = \{1, ..., N\}$ be the set of N PRBs per cell. Mobile users strategies s_k are the choice of a PRB n at a given BS j, i.e., $s_k = (j, n)$. Hence, the signal received by a mobile user k using strategy s_k, depends not only on the BS transmit power but also on the interferences introduced by the other cells. The Signal-to-Interference plus Noise Ratio (SINR) measured at the user k associated with BS j can be expressed for all $j \in \mathcal{M}$ and $k \in \mathcal{K}$ as:

$$\text{SINR}_{j,k} = \frac{h_{j,k} \cdot P(s_k)}{\sigma^2 + \sum\limits_{l \neq k} h_{j',k} \cdot P(s_l) \cdot f(s_k, s_l)} \tag{2.1}$$

where $h_{j,k}$ is the block fading process measured at user k associated with BS j, $P(s_l)$ is the power received from BS i at PRB n' for mobile user l with $s_l = (j', n')$, and σ^2 is the noise variance. The interference function $f(s_k, s_l)$ is defined as:

$$f(s_k, s_l) = \begin{cases} 1, & \text{if } n = n' \\ 0, & otherwise \end{cases} \tag{2.2}$$

2.2.2 Network Resources

A key example of dynamic resource allocation is that of power control, which serves as means for both battery savings at the mobile, as well as interference management in the network. Formally, we assume that the network manager optimizes its global utility by means of power control optimization. Let $\mathbf{P}$ be the $(M \times N)$ power control matrix whose element $P(j, n)$ represents the power received from BS $j \in \mathcal{M}$ at PRB $n \in \mathcal{N}$. Given these optimized power levels $\mathbf{P}$, mobile users choose the association actions that optimize their individual utilities. Notice that the maximization of the total throughput by the network manager is based on information sent by mobile users on interferences experienced from neighboring cells. We further assume that each base station can allocate a PRB to only one mobile user at a given time slot.

2.3 Hierarchical game formulation

We make use of a hierarchical equilibrium solution concept, i.e., the Stackelberg game, where the network manager is acting as the leader and mobile users are the followers. In view of maximizing its utility, the leader enforces its strategy on the followers that react rationally to this enforcement. A mobile user can decide to either transmit data or

stay silent depending on its utility[1]. We assume that each mobile k has a target SINR noted by η_k which reflects its required QoS. Let $\mathcal{H}_k \subset \mathcal{M}$ be the set of base stations within a radius of r from user k such that $r \leq (\frac{P_{max}}{\eta_k \cdot \sigma^2})^{1/\beta}$ where β is the pathloss coefficient and P_{max} is the maximum power at each base station. The motivation behind doing so is that, for computation purpose, one may only consider the subset $\mathcal{H}_k$ rather than the original set $\mathcal{M}$. Let w_k be the user's strategy when the user decides to stay silent on that specific slot. Hence, the set of mobile user actions is $\Omega_k = \mathcal{H}_k \times \mathcal{N} \cup w_k$. Mobile user utility function for each choice of strategy is the following

$$v_k(s_k) = \begin{cases} (R_{j,k} + \epsilon)\, \mathbb{1}_{\{\text{SINR}_{j,k} > \eta_k\}} - \epsilon, & if\ s_k \neq w_k \\ 0, & otherwise \end{cases} \tag{2.3}$$

where ϵ is a small positive value and $R_{j,k} = \log\left(1 + \text{SINR}_{j,k}\right)$ is the throughput of user k associated to BS j. This means that if a mobile user decides to transmit, it obtains either a utility equal to its transmission rate ($R_{j,k}$) or a negative utility ($-\epsilon$) depending on its SINR. Otherwise, the mobile user decides to stay silent ($v(s_k) = 0$). As a result, this tends to lead users which do not contribute enough utility, to outweigh the interference degradation and remain silent. In order to provide a right balance between efficiency and fairness between cells, one possible remedy would be to use the so-called α-fairness [77]. This guarantees that any point in which one BS is shut down cannot be a local maxima. The global utility can be expressed as:

$$U = \begin{cases} \dfrac{1}{1-\alpha} \sum\limits_j U_j^{1-\alpha}, & \text{if } \alpha \neq 1 \\ \sum\limits_j \log\left(U_j\right), & \text{if } \alpha = 1 \end{cases} \tag{2.4}$$

where $U_j = \sum\limits_k R_{j,k}$ and α is the fairness parameter. The network manager is assumed to perfectly know the set of strategies and the utilities of the K mobile users. Similarly, it is guaranteed under this setting that the followers can observe the actions of the leader through the broadcast channel. Accordingly, the Stackelberg game can be formulated as :

$$\begin{aligned} &\mathbf{P}^{SE} = \text{argmax}\, PU(\mathbf{P}(\mathbf{s}^{NE})) \\ &s.t. \qquad \sum_{n=1}^{N} P(j,n) \leq P_{max} \ \ ; \forall j \in \mathcal{M} \end{aligned} \tag{2.5}$$

where $\mathbf{s}^{NE}$ is a Nash equilibrium among K mobiles considering the strategy of the leader. Let $\mathcal{S} = \{\Omega_1 \times \cdots \times \Omega_K\}$ the strategy space of our one shot game and $s = (s_k, s_{-k})$ a strategy profile in the game. Mathematically, the Nash Equilibrium can be expressed by the following inequality for all association strategies $\mathbf{s} \in \mathcal{S}$:

$$v_k(s_k, s_{-k}) \geq v_k(r_k, s_{-k}); \qquad \forall\ k = 1, ..., K \tag{2.6}$$

[1]Though some users stay silent, they may be active during the next scheduling period.

for every $r_k \in \Omega_k$ and $s_{-k} \in \Omega_{-k}$ where $\Omega_{-k} = \{\Omega_1 \times \cdots \times \Omega_{k-1} \times \Omega_{k+1} \times \cdots \times \Omega_K\}$ is the joint feasible strategy space of all users but the k-th one.

2.4 Learning for optimal decision

The interaction between the leader and the followers provides a potential incentive for both agents to make a decision process based on their respective perceived payoff. This section focuses on how to reach the Stackelberg equilibrium for both the leader and the followers. To accomplish the task of global optimization problem, a two-stage optimization algorithm is proposed. One difficulty in our context is that mobile users do not know the payoffs (thus the strategy) of each other at each stage. Thus, the environment of each mobile user, including its opponents, is dynamic and may not insure convergence of the algorithm. In [53], authors develop a a Nash-Stackelberg fuzzy Q-learning in a heterogeneous cognitive network. As an alternative way, we adopt a hierarchical algorithm. The proposed approach requires neighboring base stations to exchange load (or interference) information experienced at user level on regular intervals. Consequently, the hierarchical algorithm is performed based on a coordination on both local (user level) and global scope (network level), which could scale accordingly.

As far as the two-stage learning algorithm is concerned, this can be conducted in the following steps: First, every user reports to its serving base station the experienced interference from neighboring cells. Then, the interference information is exchanged between base stations over the X2 interface while trying to optimize the global network utility by means of power control. Based on these power levels (broadcasted by BSs), each user check *distributively* whether the serving BS is still the best choice according to its utility. Otherwise, it can perform a handover to the other RANs after checking that it could be admitted on it. As a result, this approach tends to substantially reduce signaling overhead from the base stations.

2.4.1 Leader: Gradient computation mechanism

In this section we propose an operational way of computing the derivative of the global utility in a distributive fashion. At each time epoch, consider that each mobile user k reports to its base station j the matrix $(b_k^{1j}, \ldots, b_k^{Mj})$, where $b_k^{mj} = (P(m, n') \cdot h_{m,k}, n' \in \mathcal{N})$ is the vector of interferences perceived by user k from base station m or its signal strength on sub-band n when $m = j$. In the scope of this work, and without lost of generality, we consider that only one user can interfere per base station with another user from a neighboring base station if they use corresponding channels. Hence, adjacent channel interferences are not included and we assume that users are not allocated more than one PRB at the time. The vector b_k^{mj} restrics then to a single interference value for each m. Base station j will then be able to built the hyperma-

trix $B = \begin{pmatrix} b_1^{1j} & . & . & . & b_K^{1j} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ b_1^{Mj} & . & . & . & b_K^{Mj} \end{pmatrix}$, and send the matrices $a^j = (a_k^j, k = 1, ..., K)$, with $a_k^j = \sum_m b_k^{mj}$ and $b^j = \begin{pmatrix} b_1^{j'j}, & \dots & , & b_K^{j'j} \\ b_1^{jj}, & \dots & , & b_K^{jj} \end{pmatrix}$ to each base station j'. The derivative of the utility from base station j computed here below, is obtained in base station j' for sub-band n by:

$$\begin{cases} \dfrac{\partial U_j\left(P(j',n)\right)}{\partial P(j',n)} = \sum_{i=1}^{K} \left(\dfrac{b_i^{j'j}/P(j',n)}{\sigma^2 + a_i^j} - \dfrac{b_i^{j'j}/P(j',n)}{\sigma^2 + a_i^j - b_i^{jj}} \right) \text{ and} \\ \dfrac{\partial U_{j'}\left(P(j',n)\right)}{\partial P(j',n)} = \sum_{i=1}^{K} \left(\dfrac{h_{j',k}}{\sigma^2 + a_i^{j'}} \right) \end{cases} \tag{2.7}$$

So far, $\dfrac{dU}{\partial P(j',n)} = U_{j'}^{-\alpha} \sum_j \dfrac{\partial U_j}{\partial P(j',n)}$. We then need to express $\dfrac{\partial U_j}{\partial P(j',n)}$ for all j, assuming we are considering cell j'. Because the derivative goes the same for every sub-band, we will focus only on one particular sub-band n. It can be easily shown that $\dfrac{\partial U_j}{\partial P(j',n)}$ is given by:

$$\begin{cases} \left(\dfrac{h_{j',k}}{\sigma^2 + \sum_m P(m,n) \cdot h_{j,m}} - \dfrac{h_{j',k}}{\sigma^2 + \sum\limits_{m \neq j} P(m,n) \cdot h_{j,m}} \right); & if\ j \neq j' \\ \left(\dfrac{h_{j,k}}{\sigma^2 + \sum_m P(m,n) \cdot h_{j,m}} \right); & if\ j = j' \end{cases}$$

The pseudo-code for the proposed gradient descent approach is given in *Algorithm 1*.

Note that the implementation of gradient like algorithms is familiar in optimization problems. The convergence of such algorithms have been shown in [19], under some specifics conditions such as the derivative of objective function is Lipschitz continuous which is satisfied here, and for an accurate choice of γ_t.

Proposition 2.4.2. *The derivative of our utility function is Lipshitz continuous*

Proof. If $s_k \neq w_k$, the utility function is given by

$$U = \begin{cases} -\epsilon & if\ SINR_{jk} > \eta_k \\ R_{jk} + \epsilon & otherwise \end{cases}$$

Again $SINR_{jk} > \eta_k \implies U = R_{jk} + \epsilon$. Let's show that ∇U is Lipschitz continous. We have, $\dfrac{\partial U_j}{\partial P(j',n)} =$

Algorithm1
In each base station j'

1. Find the value of $\frac{dU}{\partial P(j',n)}(P_t(j',n))$ at time t, given the current power level $P_t(j',n)$.
2. Express $P_{t+1}(j',n) = P_t(j',n) + \gamma_t \frac{\partial U}{\partial P(j',n)}(P_t(j',n))$, to have new values of the power vector at time $t+1$.
3. Allocate the powers to each sub band and let the users associate(see **Algorithm2**).
4. Receive feed-back from users and build the hypermatrix B.
5. Send $a^{j'}$ and $b^{j'}$ to each base station j.
6. If $\left(\max_n \left(P_{t+1}(j',n) - P_t(j',n)\right)\right) \leq \epsilon$ stop; else go to 1.

End algorithm.

$$\begin{cases} \frac{h_{j'k}}{\sigma^2 + \sum_m P(m,n)h_{m,k}} - \frac{h_{j'k}}{\sigma^2 + \sum_{m \neq j} P(m,n)h_{m,k}} & if\ j \neq j' \\ \frac{h_{jk}}{\sigma^2 + \sum_m P(m,n)h_{m,k}} & if\ j = j' \end{cases}$$

Then for another value q of power level on channel n, we obtain

$$\|\frac{\partial U_j}{\partial P(j',n)} - \frac{\partial U_j}{\partial P(q,n)}\| \leq \|\frac{\partial U_j}{\partial P(j',n)}\| + \|\frac{\partial U_j}{\partial P(q,n)}\|$$

Moreover, $\forall\ P(j',n)$, we have

$$\begin{aligned} \|\frac{\partial U_j}{\partial P(j',n)}\| &\leq \|\frac{h_{jk}}{\sigma^2 + \sum_m P(m,n)h_{m,k}}\| \\ &\leq \frac{H}{\sigma^2} \text{ where } H \text{ is the line of sight} \\ &\qquad \text{channel gain.} \end{aligned}$$

which implies that

$$\|\frac{\partial U_j}{\partial P(j',n)} - \frac{\partial U_j}{\partial P(q,n)}\| \leq \frac{2H}{\sigma^2}$$

yielding $\exists\ A,\ K > 0\ s.t$

$$\|\frac{\partial U_j}{\partial P(j',n)} - \frac{\partial U_j}{\partial P(q,n)}\| \leq K \times \|P(j',n) - P(q,n)\|$$

with $K = A \times \frac{2H}{\sigma^2}$. This ends the proof. □

In our computations, we use the implementation of WOLFE linear search to find an appropriate value of γ_t at each iteration.

On another hand, at each iteration, the gradient algorithm delivers the values of the powers vector for each base station, that can go out of the bounds of the allowed space. To handle this problem, we implement a computational mechanism to satisfy the power constraint in (2.5). We define the constraint $c\left(P(j,n), P_{max}\right)$ to relax the problem, where c is built as follows:
$define\ \xi = \{n \in \mathcal{N}\ s.t.\ P(j,n) > p_{th}\}$; where p_{th} is a threshold value for every PRB $n\ s.t. \sum_n p_{th} = P_{max}$. Let $\delta = \left(P_{max} - \sum_{\mathcal{N}\setminus\xi} P(j,n) - \sum_{\xi} p_{th}\right)$;
$if\ \exists k \in \xi\ and \sum_n P(j,n) \geq P_{max}$ set the values of each $P(j,n)$, using the projection $\bar{P}(j,n) = \min\left(\frac{\delta}{|\xi|} + p_{th},\ P(j,n)\right)$. It is to say that the remaining power on each BS power budget, if any, is evenly shared among the channels requiring a power level above the threshold value.

2.4.3 Followers: Pursuit algorithm

At the user level of our stakelberg framework, we use the pursuit algorithm as a tool to allow user to reach iteratively and individually a Nash Equilibrium. The pursuit algorithm is a distributed association algorithm proposed in [118] allowing each individual in a set of players to select a given strategy, among several others, that will best maximize its utility within a limited number of iterations.

Algorithm2
At each iteration t
1. Select a strategy $s_k \in \Omega_k$ according to the current powers level $\mathbf{P}$.
2. Update the vector of average utilities $u_{avg,k}$, using the chosen strategy $s_k\left(u_{avg,k}\right)$, provided utility $v_k\left(s_k\right)$.
3. Find the strategy $s_k = arg\max(u_{avg,k})$, that gives the best average.
4. Make $\begin{cases} p_{s_k}(t+1) = & p_{s_k}(t) + \delta\left(1 - p_{s_k}(t)\right) \\ p_{s_i}(t+1) = & p_{s_i}(t) - \delta p_{s_i}(t)\,,\ i \neq k \end{cases}$
5. if $max\left(|p_{s_i}(t+1) - p_{s_i}(t)| < \epsilon\right)$ stop, else go to 1.

End Algorithm

It has been proven in [118] that the pursuit algorithm always converges under some specifics conditions on the step size parameter. They show that when the step size parameter is very small, the game converges to a stable equilibrium for the learning automata game. This algorithm has the property to converge to an extremum of the game when there exist a pure equilibrium. To reach mixed equilibrium, authors in [129] present a distributed algorithm that can be used in such situations. However, mixed equilibria are not efficient in our context since it will lead mobile users to process continuously handovers between bases stations. To avoid mixed equilibria, we

Table 2.1: *Simulations settings*

Parameters descriptions	Values
Number of cells	7
Number of PRBs per cell	10
Number of users per cell	4
Outer Radius of hexagonal cells	200 meters
Distance to insure target $SINR$	300 meters
fairness parameter	$\alpha = 1$
iterations scale [2]	1 for 30

introduce a cost of handovers in the utility function to give more incentive to mobile users in reaching pure equilibria.

Discussion on cost of handover As stated in the previous paragraph, a major weakness of the learning algorithm in mobile networks is the number of handovers, especially when the algorithm converges to mixed equilibria. We try to tackle this issue by introducing a cost of handover as a reward in the utlity function to users who are not operating handovers. Users utility function is given by, $v_k(s_k) =$

$$\begin{cases} (R_{j,k} + \epsilon)\, \mathbb{1}_{\{\text{SINR}_{j,k} > \eta_k\}} - \epsilon + \alpha_h, & if\ s_k \neq w_k \\ 0, & otherwise \end{cases} \tag{2.8}$$

where $\alpha_h = \beta_h(1 - \mathbb{1}_{\{\text{handoff}\}})$ and β_h is a small positive value. In figure 2.2 we compare the number of handovers with and without the defined handover control. On the figure we can see that the handover control policy decreases considerably the number of handovers and the system remain stable after a few iterations. Interestingly, we noticed in our simulations that users are more motivated in following the handover control when the control is a reward rather than a penalty as we suggested in a first place. A trade-off on using such control can be seen at the utility side. As shown in figure 2.3 the gap in utility between the two policies can be marginal.

2.5 Implementation and Validation

To go further with the analysis, we resort to realistic network simulations. We consider a cellular radio network as described in Figure 2.1 where users are attempting to communicate during a downlink transmission, subject to mutual inter-cell interferences. Specifically, a hexagonal cellular system functioning at 1.8 GHz where cell radius is equal to $R = 200$ meters is considered. Note that this radius only stands for geographical positions in the network. It does not prevent users located out of this area to connect

[2] This is the scale of iterations between the gradient and Pursuit algorithms

with another base station in case of significant connection opportunity. Channel gains are based on the COST-231 path loss model [1] including log-normal shadowing with standard deviation of 10 dB, plus fast-fading assumed to be i.i.d. circularly symmetric with distribution $\mathcal{CN}(0,1)$. The peak power constraint is given by $P_{max} = 100$ mWatts. We evaluate under those settings the joint processing of the gradient descent algorithm with the pursuit algorithm. Without loss of generality, we assume that every cell has the same number of users randomly positioned inside the cell. We consider a cluster of 7 interfering cells, featured with 10 PRBs each. The values of the other parameters are set in Table 2.1. The iteration scale parameter in Table 2.1 traduces how frequently BSs update the gradient algorithm and set new values of powers. By tunning this parameter, one can control the amount of signalization between BSs. We consider in our simulations that users run 30 iterations of the association algorithm for 1 iteration of the gradient. We first build the framework for a fairness parameter $\alpha = 1$ which represents the proportional fairness algorithm, and then extended it to different values of α.

2.5.1 Dynamic fractional frequency reuse

In Figure 2.4, we illustrate the snapshot of the dynamic fractional frequency reuse pattern at the equilibrium. The small colored disks indicate the positions of users inside the cells and the faces colors are the frequencies used by those users. Disks are indexed with a couple of values (BS, power) where the first value represents the base station to which this user is connected and the second the power level assigned by the base station on that frequency. As expected, users close to each other are attributed different frequencies and power levels are set accordingly, to avoid high level of interferences. From the same figure, we also have an overview on user-network association. Indeed, many cases appear where users would rather associate in a neighboring cell rather than in the cell where they are positioned, due to the influence of path-loss and/or interference impairments. For instance, in Figure 2.4, user indexed $(2, 7.5)$ in cell 2 is connected to BS 2 and is assigned frequency $F2$ with a high power level. This reflects the maximization goal of the gradient algorithm, since frequency $F2$ is reused only once by a user far away in cell 7 at a low power level.

2.5.2 Utility maximization

In Figure 2.5, we compare the proposed FFR algorithm with a traditional fixed reuse patterns namely, the full reuse. The exhaustive search algorithm, in dash, considers all possible combinations of PRB selection given the power level of the gradient algorithm. This will thus serve as an *optimal* association solution for users and will demonstrate just how much gain may theoretically be exploited through the pursuit algorithm. It clearly appears that the joint gradient and pursuit algorithm performs better than the full reuse and reduce considerably the gap with the exhaustive search. As it is shown in Figure 2.5, we reach up to 90% of overall network throughput compared to the exhaustive (optimal) association search.

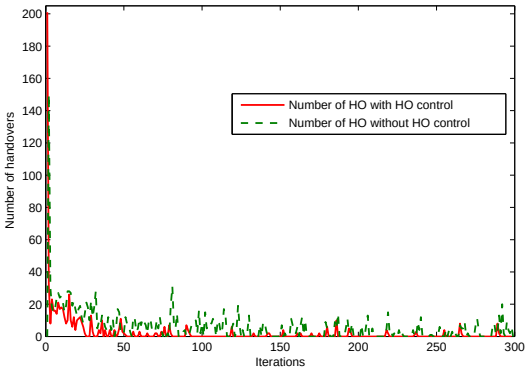


Figure 2.2: *Comparision of the number of handover.*

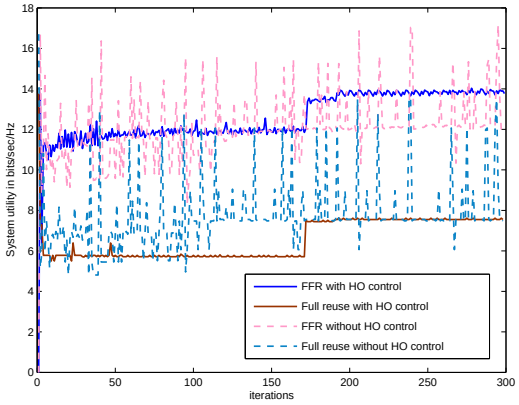


Figure 2.3: *Comparision of Network uilities.*

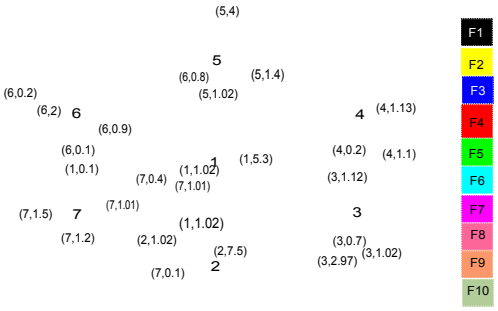


Figure 2.4: *Snapshot of the dynamic fractional frequency reuse pattern at the equilibrium for $\alpha = 1$.*

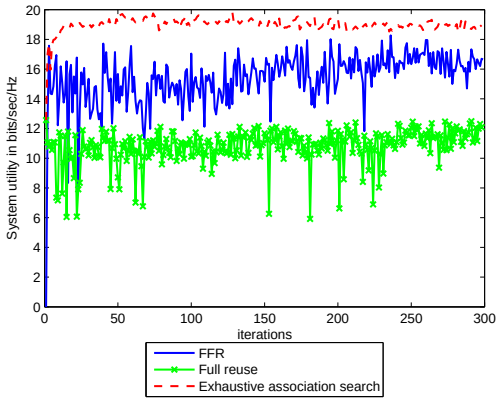


Figure 2.5: *Network utility for $\alpha = 1$.*

2.5.3 Fairness Issues

In this section, we intend to show the impact of fairness on the global utility maximization by simulating different values of the α-fairness parameter. For $\alpha = 0$ (the maximum throughput algorithm), we can see from Figure 2.6 that some BSs (BS 6 for instance) are set to idle. Several other channels in the network are also switched off while a few number of users are attributed very high levels of power. This behavior was somehow expected since using $\alpha = 0$ means that the major goal of the network is to maximize the overall network utility. Nevertheless as it is shown in Figure 2.7, this policy does not help to tighten the gap to the exhaustive association search (71%), as much as using a value of $\alpha = 1$. Further analysis of the max-min fairness policy ($\alpha \to \infty$) shows that most of BSs are set to idle, and only few channels are activated. Being too fair leads then the network to follow the policy of highly loaded BSs, thus providing an overall network utility almost null. Finally, we plot in Figure 2.8 the net-

work block call rate (BCR) for increasing number of iterations obtained when users follow the strategy corresponding to the Stackelberg equilibrium. We can observe that the BCR can be substantially reduced as the number of iteration increases. Moreover, the fairness policy has a negligible influence on the BCR which remains less than 10% for the different fairness policies.

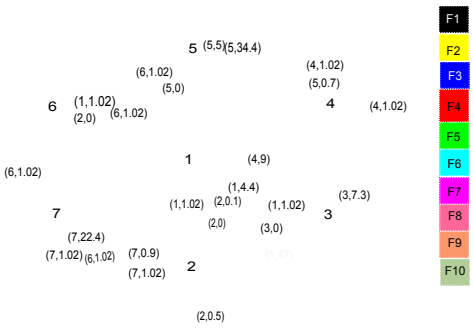


Figure 2.6: *Snapshot of the FFR system pattern* ($\alpha = 0$).

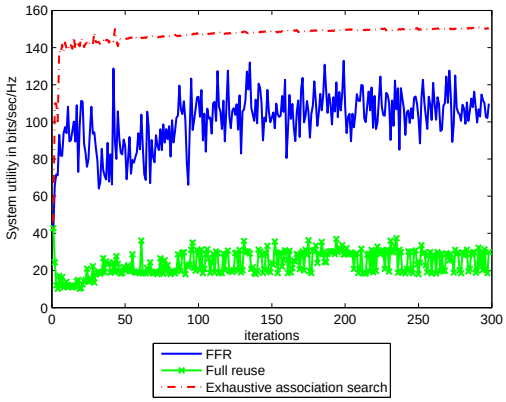


Figure 2.7: *Network utility for* $\alpha = 0$.

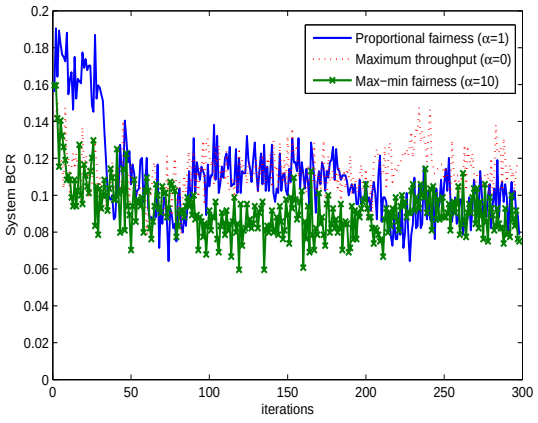


Figure 2.8: *Block call rate for different values of* α.

2.5.4 Robustness and scalability

Next, we evaluate in this section a seamless adaptation of our algorithms to a dynamic environment. We simulate a discrete time system over several iterations, and generate a burst of users arrival at a specific time instant during the simulation time. Indeed, while new arrivals generally occur every minute in the cellular systems, our association

algorithm converges at the order of a few *ms*. This speed of reactivity and adaptation show improved performances of our hierarchical algorithm and is traduced in figures 2.9, 2.10 and 2.11. We consider two rings of a small cells network where each base station is featured with 4 PRBs and contain each 3 users randomly positioned inside the cell at the beginning of the simulation. In this new setting, we assume that the algorithms iteration scale is 1 for 100. For each of the simulated schemes, we assume that each BSs can exchange interferences information only with a subset of all the interfering neighbors. This consideration helps to understand how the proposed scheme reacts when the amount of information exchanged between BSs is limited.

For the first (figure 2.9), second (figure 2.10) and third (figure 2.11) scenario, we consider that each BS exchange data respectively with the first ring neighbors, only with the 2 first closest neighbors and finally with all the interfering neighbors from the two considered rings. By comparison of the different scenarios, it is observed that more exchanged information lead, as one can intuitively expect, to an increased outcome in utility. However, although this can be imputed to randomness, when comparing figures 2.9 and 2.11 we see that the system stability is not necessarily insured by an increase of exchanged information rate. On another hand, even with very few amount of exchanged information our algorithm preserves a convergence to 97% of the exhaustive association search utility. From the same figures, we address the scalability of our algorithms, with the introduction of a burst of new arrivals in the system at iteration 200. Although this event in not clearly captured in figure 2.10, case when less information is exchanged, we can observe from figures 2.9 and 2.11 that our mechanism adapts very fast to the system evolution in order to reach the new point of convergence.

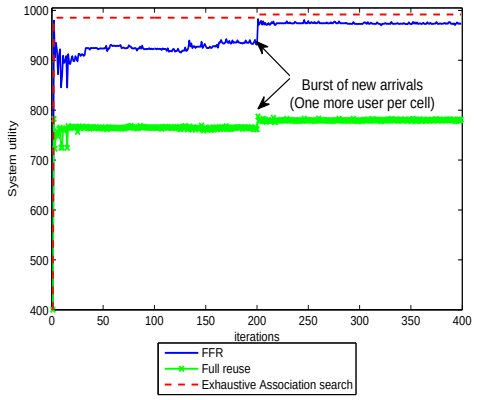


Figure 2.9: *Network utility for $\alpha = 0.98$ with first ring interferences informations.*

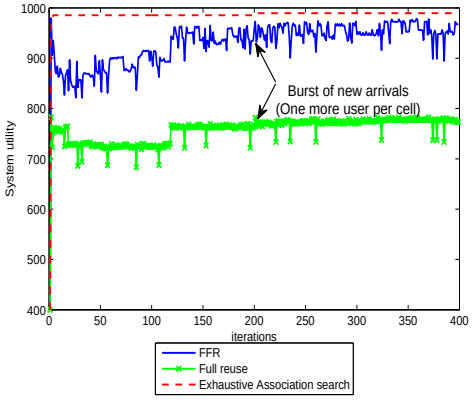


Figure 2.10: *Network utility for $\alpha = 0.98$ with two neighbors interferences informations.*

2.6 Conclusions

In this chapter, we have investigated the idea of a hierarchical learning game for fractional frequency reuse in an OFDMA network. In this framework, both the network

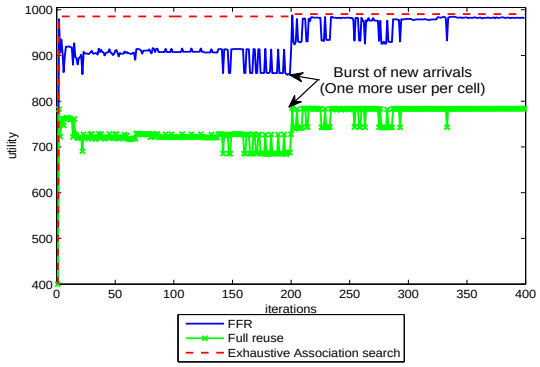


Figure 2.11: *Network utility for* $\alpha = 0.98$ *with all neighbors interferences informations.*

manager and mobile users learn to reach an equilibrium that optimizes the global network utility while ensuring individual utility optimization for mobile users. We have first proposed formally a game model to define how the network manager and mobile users can obtain their respective equilibria by means of a Stackelberg formulation. Then, we have presented a two-stage learning algorithm for finding a Stackelberg equilibrium and the corresponding mobiles' association strategies. Practical directions for implementability of our solution are also presented. We have showed using several numerical examples the efficiency of the obtained equilibrium compared to the exhaustive (optimal) solution and a fixed full frequency reuse pattern. In particular, in the case of proportional-fair policy, the proposed FFR approach offers approximately 90% of the optimal association policy and 40% of gain with respect to the fixed full reuse. Indeed, for implementation purposes and in order to adapt to the dynamic of the mobile environment, the number of iterations before convergence should remain in the order of a few tens. Eventually, we have addressed interesting issues such as fairness, robustness and scalability and offered insights into how to design such scenario in a wireless network environment.

Part III

Energy efficient games in decentralized networks in ad-hoc mode

Chapter 3

Delay Tolerant Networks in partially overlapped networks: A non-cooperative game approach

3.1 Introduction

Delay tolerant mobile ad-hoc networks have gained attention in recent research. Instantaneous connectivity is not needed any more and messages can arrive at their destination thanks to the mobility of some subset of nodes that carry copies of the message. A naive approach in forwarding a message to the destination consists in the use of an epidemic routing strategy, in which any mobile that has the message keeps on relaying it to any other mobile that arrives within its transmission range and which does not still have the message. This would minimize the delivery probability at a cost of inefficient use of network resources in terms of energy used for transmission. The need for a more efficient use of network resources has motivated the use of more economic packet forwarding strategies such as the two-hop routing protocols, in which the source transmits copies of its message to all mobiles it encounters, but these relay the message only if they come in contact with the destination. The performance of the two-hop forwarding protocol along with the effect of the timers have been evaluated in [4]. In this study we consider an alternative approach that offers a way of analyzing the successful delivery probability and energy consumption. We aim to provide a scheme which maximizes the expected delivery rate while satisfying a certain constant on the number of forwardings per message. To do this, we assume that each mobile may decide which routing protocol it wants to use for delivering packets. We restrict the case that only two routing protocols are available to mobiles: epidemic routing and two-hops. This scheme allows us to exploit the trade-off between delivery delay and resource consumption. The higher number of users use epidemic (resp. two hops) routing , the higher (resp. lower) probability of success and the higher (resp. lower) consumption of resource.

In our study we assume that each mobile like to find the routing protocol that maxi-

mizes his utility function. But, as this utility depends on the action of the other mobiles, the system can be described as a non-cooperative game. We show that this game has at least one Nash equilibrium, and we designed a distributed algorithm to reach it. This algorithm is implemented at each node, allowing the system to reach the Nash equilibrium in a completely distributed way. Since the estimation of some parameters of the system, is very difficult in DTN, due to the lack of persistent connectivity, the proposed algorithm also allows the nodes to converge to the Nash equilibrium without any information.

Delay Tolerant Networks (DTNs) have recently attracted attention of the research community. Delay Tolerant Networks (DTNs) are sparse and/or highly mobile wireless ad hoc networks where no continuous connectivity guarantee can be assumed [122, 100]. There are several results of real experiments on DTNs [56, 31, 48]. In [127], the authors studied the optimal static and dynamic control problems using a fluid model that represents the mean field limit as the number of mobiles becomes very large. In [35], the optimal dynamic control problem was solved in a discrete time setting. The optimality of a threshold type policy, already established in [36] for the fluid limit framework, was shown to hold in [35] for the actual discrete control problem. A game problem between two groups of DTN networks was further studied in [35].

3.1.1 Main contributions

The major contributions in this chapter are the following:

- We address a routing configuration game in DTN by allowing users to change dynamically their routing policy according to network reward and other users configurations.
- Our game is energy efficient and leads to an optimal trade-off between the successful delivery probability and the number of infected relays at the equilibrium.
- We give some analytical insight to model the propagation of the message from the source to the destination as a fluid model using mean field approximation.
- We define a reward mechanism to give the incentive to users of using one strategy or the other by introducing a reward on participation in successful message delivery.
- We eventually rely on a learning algorithm to allow users to determine the best strategy to adopt and converge to a NE.

The rest of the chapter is organized as follow. In section 3.2 we decline the network model and DTN framework. In section 3.3 we establish the existence of a Nash equilibrium for our game. Section 3.4 presents the stochastic approximation for our leaning process and we define our reinforcement learning algorithm. We analyze in section 3.5 the network efficiency and compare to global optimum using the price of anarchy. We eventually conclude the chapter in section 3.6.

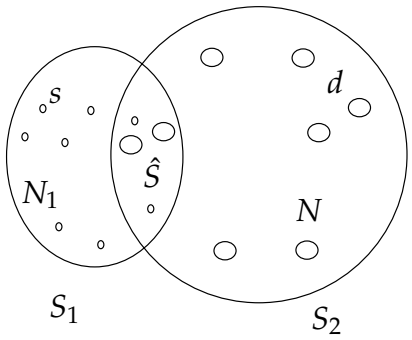


Figure 3.1: *Overlapped Network Region $\hat{S}$*

3.2 The Model

Let us consider two overlapping network regions, where source and destination nodes are each in distinct regions. By network region we mean a region with moving nodes that can establish a connection between them. We assume that nodes have random way-point mobility (see [49]) which is confined to the region it is associated . In context of DTN the transportation of data relies mainly on mobility, so the overlapping region plays an important role. Overlapping regions are the only place where nodes can exchange data from one region to another. Consider that network region S_1 contains a source S, and N_1 mobile nodes, and that network region S_2 contains the destination node d and N_2 mobile nodes. Since source and destinations are in different regions, data can be transported from source to destination by mobile nodes only through the overlapping region $\hat{S}$. Let us parameterize the overlapped(normalized) region, denoted by $\tilde{S} = \hat{S}/\max\{S_1, S_2\}$. Notice that the overlapping region $\tilde{S}$, when parameterized reduces to (assume $S_1 = S_2$ for simplicity) the following special cases : "Unified network", i.e., when $\tilde{S} = S_1 = S_2$, and "Overlapped network" when $0 < \tilde{S} < 1$.

We assume that each mobile node is equipped with some form of proximity wireless communications device. The network regions are assumed to be sparse, so that, at any time instant, nodes are isolated with high probability. Communication opportunities arise whenever, due to mobility patterns, two nodes get within mutual communication range. We refer to such events as "contacts". The time between subsequent contacts of any pair of nodes is assumed to follow an exponential distribution. The validity of this model for synthetic mobility models (including, e.g., Random Walk, Random Direction, Random Waypoint) has been discussed in [4]. In [49], the authors derived the following estimation of the pairwise meeting rate λ :

$$\lambda = \frac{2wRE[V^*]}{S}, \tag{3.1}$$

where w is a constant specific to the mobility model, $E[V^*]$ is the average relative speed between two nodes and R is the range. Let λ_1 (resp. λ_2) be the rate of meeting of any pair of nodes in region S_1 (resp. S_2). Let λ_S denote the rate of meeting between the

source and a node in region S_2. From (3.1), we have

$$\lambda_1 = \frac{2wRE[V_1^*]}{S_1},\ \lambda_2 = \frac{2wRE[V_2^*]}{S_2} \text{ and } \lambda_s = \frac{2wRE[V_s^*]}{S_1}.$$

Similarly, the rate of meeting between a node (resp. source) in S_1 and a node in S_2 is given by $\lambda_{12} = \frac{2wRE[V_{12}^*]}{\hat{S}}$, $\lambda_{s_2} = \frac{2wRE[V_{s_2}^*]}{\hat{S}}$,

where V_{s_2} is the average relative speed between source and a node in region S_2. There can be multiple source-destination pairs, but we assume that at a given time there is a single message, eventually with many copies, spreading in the network. For simplicity we consider the message originated at time t = 0. We also assume that the message that is transmitted is relevant only during some time τ. The message contains a time stamp reporting its generation time, so that it can be deleted at all nodes when it becomes irrelevant.

A mobile terminal is assumed to have a message to send to a destination node. We consider in our approach two types of routing in DTN networks: epidemic routing and two-hop routing.

- **Epidemic routing:** At each encounter between a mobile that has the message and another one that does not, the message is relayed to the one that does not have it.
- **Two-hop routing:** At each encounter between the source and a mobile that does not have the message, the message is relayed to that mobile. If a mobile that is not the source has the message and it is in contact with another mobile then it transfers the message if and only if the other mobile is the destination node.

In this chapter we study the competition between individual mobiles in a game theoretical setting. Each mobile can decide whether to use epidemic or two-hop routing, depending on which strategy maximizes his utility function. We assume that the source node S stays in region S_1 while the destination node d stays in region S_2. Naturally, the nodes in S_1 needs to forward the packet to the nodes in S_2. Hence, the nodes in S_1 are of "Epidemic" type only, while nodes in S_2 may be of either type.

Consider that there are N_1 mobiles among the total N_{tot_1} in region S_1 which participate in forwarding the packet using epidemic routing. We assume that N mobiles among N_{tot} in region S_2 can choose between epidemic and two-hop routing. Let N_e^0 (resp. N_t^0) be the number of mobiles that always use epidemic (resp. two-hop) routing. Then, we have:

$$N_{tot} = N + N_e^0 + N_t^0$$

The source in region S_1 has a packet generated at time 0 that wishes to send to the destination d in region S_2. In region S_2, let N_e (resp. N_t) be the number of users that use epidemic routing (resp. two-hop routing). Let $X_e(t)$ (resp. $X_t(t)$) be the number of mobile nodes (excluding the destination and source) that use epidemic routing (resp. two-hop) and have at time t a copy of the packet. Denote by $D_i(\tau)$ the probability of a successful delivery of the packet by time τ. Then, given the process X_i (for which

a fluid approximation will be used), we have the probability of successful delivery of packet as:

$$P_{succ}(\tau) = 1 - e^{\left(-\lambda_d \int_0^\tau (X_e(t)+X_t(t))dt\right)} \tag{3.2}$$

where λ_d denotes the inter-meeting rate between the destination and a node in S_2. Consider that on successful delivery of the packet is rewarded with $\bar{\alpha}$ which is shared among all the participating nodes. Let the reward is shared among the two region as α_{S_1} for region S_1 and α for S_2, where $\bar{\alpha} = \alpha_{S_1} + \alpha$. In region S_1 there are only epidemic type user, the reward is shared equally among $X_1(\tau)$ users. While in region S_2, the reward α is further shared as α_e (resp. $\alpha_t = \alpha - \alpha_e$) among the mobiles that have at time τ a copy of the message and use epidemic (resp. two-hop) routing. Hence, the utility U_e (resp. U_t) for a player using epidemic (resp. two-hop) routing is given by

$$U_e(N_e) = \left(\frac{\alpha_e P_{succ}(\tau)}{X_e(\tau)} - \beta\tau\right)\P_1 \tag{3.3}$$

Similarly, the utility for a player use two-hop routing is given by

$$U_t(N_e) = \left(\frac{\alpha_t P_{succ}(\tau)}{X_t(\tau)} - \gamma\tau\right)\P_1 \tag{3.4}$$

where β and γ are the energy cost ,and $\P_1(t) = 1 - e^{-\int_0^t (\lambda_{s_2} + \lambda_{12} X_1(s) + \lambda_2 X_e(s) ds)}$ which denotes that the probability of receiving a packet by time t.

3.2.1 Fluid Approximation

We consider the following standard fluid approximation (based on mean field analysis)

$$\frac{dX_1(t)}{dt} = (\lambda_s + \lambda_1 X_1(t) + X_e(t)\lambda_{21})(N_1 - X_1(t)), \tag{3.5}$$

$$\frac{dX_e(t)}{dt} = (\lambda_{s_2} + \lambda_{12} X_1(t) + X_e(t)\lambda_2)(N_e - X_e(t)), \tag{3.6}$$

$$\frac{dX_t(t)}{dt} = (\lambda_{s_2} + \lambda_{12} X_1(t) + X_e(t)\lambda_2)(N_t - X_t(t)). \tag{3.7}$$

. The message is spread directionally, which means that nodes from region S_1 can forward the packet to nodes in S_2, while the reverse is not allowed,so $\lambda_{21} = 0$. On solving the ODE's given in eq. (3.5)-(3.7) using the suitable initial conditions, we obtain

$$X_1(t) = \frac{\lambda_s N_1 \left(1 - \exp\left(-t\left(\lambda_s + \lambda_1 N_1\right)\right)\right)}{\lambda_s + \lambda_1 N_1 \exp\left(-t\left(\lambda_s + \lambda_1 N_1\right)\right)}, \tag{3.8}$$

$$X_e(t) = \frac{N_e\left[\psi(t)\left(1 - N_e \int_0^t \frac{\lambda_2}{\psi(u)} du\right) - 1\right]}{\psi(t)\left(1 - N_e \int_0^t \frac{\lambda_2}{\psi(u)} du\right)}, \tag{3.9}$$

$$X_t(t) = N_t\Big(1 - \exp\Big[-\lambda_{12}\int_0^t X_1(u)du + \lambda_2 \int_0^t X_e(u)du + t\lambda_{s_2}\Big]\Big). \tag{3.10}$$

where $\psi(t) = \exp\left(\int_0^t (\lambda_{s_2} + \lambda_{12} X_1(u) + \lambda_2 N_e)\, du\right)$.

3.3 The DTN game

As explained before, there is but a single choice for the nodes in region S_1, i.e., to participate or not in epidemic forwarding. However in region S_2, a node can choose between participating or not, and, if so, it can choose between epidemic forwarding or two hop forwarding to deliver the packet to destination. This raised the game situation among the players to choose a strategy. A strategy of a mobile is to choose between epidemic and two-hop routing. Every mobile would like to find the strategy that maximizes his individual utility. But, as his utility depends on the actions of the other mobiles, the system can be described as a non-cooperative game. As the game is symmetric, a Nash equilibrium (NE) N_e^* is given by the two conditions:

$$U_e(N_e^*) \geq U_s(N_e^* - 1) \text{ and } U_t(N_e^*) \geq U_e(N_e^* + 1)$$

The previous definition means that no user using epidemic routing (resp. two-hop routing), has an incentive to use two-hop routing (resp. epidemic routing). It can be shown that the equilibrium is given by the equivalent condition

$$\begin{aligned} U_e(N_e^*) \geq \max\{ & \frac{\alpha_e(N_e^* + N_0^e)}{\alpha_t(N - N_e^* + N_0^t)}(U_e(N_e^* - 1) + \beta) - \gamma, \\ & \frac{\alpha_t(N - N_e^* + N_0^t)}{\alpha_e(N_e^* + N_0^e)}(U_e(N_e^* + 1) + \gamma) - \beta\} \end{aligned} \tag{3.11}$$

.

Proposition 3.3.1. *For each N, total number of players, there is at least one Nash equilibrium which is characterized by inequality (3.11).*

The proof of the proposition follows from [98].

3.4 Stochastic approximation for Nash equilibrium

In this section we introduce a distributed method to achieve the Nash equilibrium in the case where some parameters (i.e., N, λ and λ_s) are unknown. We show that simple iterative algorithms may be implemented at each node, allowing them to discover the Nash equilibrium in spite of the lack of information on such parameters. Note that the estimation of N, λ and λ_s, is very difficult in DTN because of the lack of persistent connectivity. This distributed algorithm proposed in [91] was proved, for a fixed number of players, that if it converges, it will always do to a Nash equilibrium. In order to increase the speed of convergence, each user decides to stop his update mechanism after reaching a given threshold [89]. It is not a global convergence criteria, as we can find in centralized algorithms, but an individual convergence criteria that let each user stop

calculations. The algorithm is based on a reinforcement of mixed strategies and players are synchronized in such a way that the decision of all players (playing pure strategy) induce the utility perceived for each one.

The algorithm works in rounds. Each round corresponds to the delivery of a message by the source. Let $N_e(t)$ be the number of players that use epidemic routing at round t. At each round t, each user i chooses epidemic routing over the set $C = \{e, t\}$ of strategies, with probability p_t (and chooses the two-hop routing with probability $1 - p_t$). The utility perceived by user i at round t depends on his action and on the actions of the other mobiles. This utility u_t^i is expressed as follows:

$$u_t^i = \mathbb{1}_{\{c_t=e\}} \cdot U_e(N_e(t)) + \mathbb{1}_{\{c_t=t\}} \cdot U_t(N_e(t)) \tag{3.12}$$

Then, each player updates his probability according to the following rule (see Algorithm 2):

$$p_t^i = p_{t-1}^i + b \cdot \left(\mathbb{1}_{\{c_t=e\}} - p_{t-1}^i\right) \cdot u_t^i, \tag{3.13}$$

Algorithm 2 Dynamic Distributed Algorithm.

1. Initialize p_0^i as starting probability for the player i.
2. For each player i:
   (a) If player i has converged move to the next player.
   (b) Player i performs a choice over C, according to $p_{t-1}^{(i)}$.
   (c) Player i updates his probability $p_t^{(i)}$ according to his choice using (3.13).
   (d) If $\left|p_t^{(i)} - p_{t-1}^{(i)}\right| < \epsilon$ then player i has converged.
3. Make $t = t + 1$ and go to step 2.

Figure 3.2 shows the evolution of the probabilities and the convergence to Nash Equilibrium for a set of 10 players, using a treshold of convergence at $\epsilon = 10^{-6}$.

3.5 Global optimum repartition and Nash Equilibrium

In this section, we are interested in the network efficiency as the maximization of the global optimum of the system. We want to optimize the overall network energy-efficiency with respect to the aforementioned degrees of freedom. For this purpose, we consider the optimal social welfare, which is well known in game theoretic studies, and compare it with the performance achieved at Nash Equilibrium.

The following simulations allow us to see the range of values for different parameters which minimizes the gap in total utility between the Nash equilibrium and the global optimum. For different rates of λ_s and different values of the reward on epidemic routing α_e, we compute the price of anarchy, using the total utility at the global optimum repartition and at Nash Equilibrium.

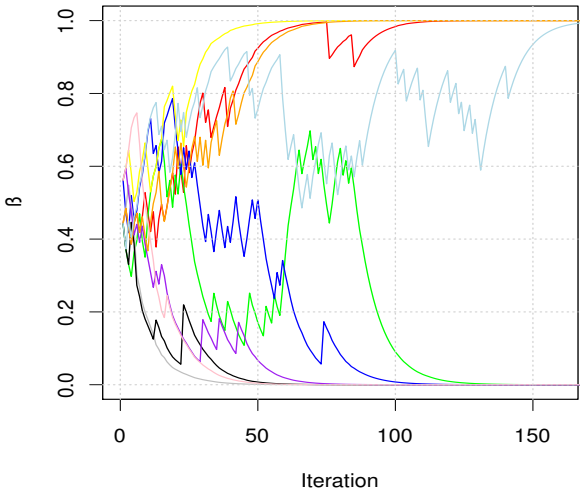


Figure 3.2: *Convergence to Nash Equilibrium*

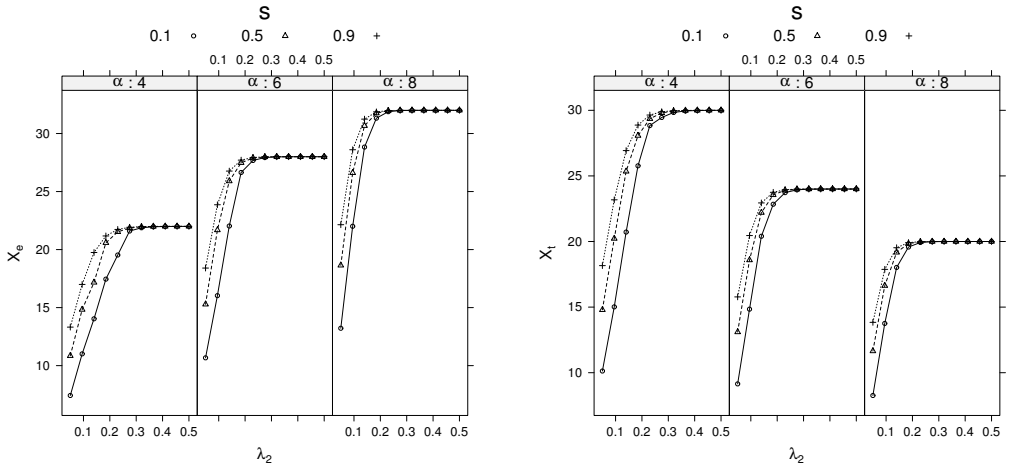
The social welfare of the network is measured by the total utility of the system expressed by

$$W_s = X_e(\tau)U_e(N_e) + X_t(\tau)U_t(N_e) \tag{3.14}$$

and the price of anarchy is measured as follows:

$$PoA = (W_s^{Opt} - W_s^{NE})/W_s^{Opt} \tag{3.15}$$

where W_s^{Opt} (resp. W_s^{NE}) is the social welfare at the global optimum (resp. at the Nash Equilibrium.)



a) Epidemic routing b) Two-hop routing

Figure 3.3: *Infected users using epidemic or two-hop routing*

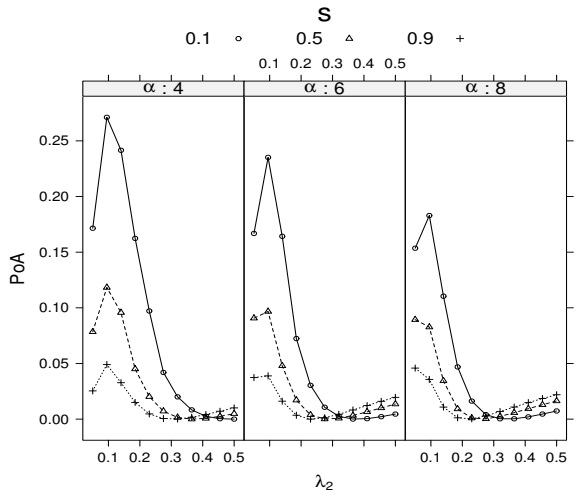


Figure 3.4: *Price of anarchy depending on* λ_2

Through the different simulations for several set of values for the main parameters of our DTN network, we observe the network stability and efficiency. In figure 3.3 we plot the evolution of the number of users infected using either two hops or epidemic routing. As we can notice, the rate of infection of users using epidemic routing increases with the inter-meeting rate in the second region before reaching a stability point, that is mainly influenced by the relevant time of packet delivery which increases the probability of success and makes the infection rate independent on λ_2. This rate is always bigger with the surface of overlapping and the reward on using epidemic routing. We observe the same behavior for the infection rate of users using two-hop routing, except that the infection rate become smaller with the reward on using epidemic routing. Figure 3.4 present on the other hand the price of anarchy (PoA) at Nash Equilibrium. For small values of the inter-meeting rate λ_2 in the second region, the PoA takes it highest values and is almost independent on $\tilde{S}$. The optimality of the Nash equilibrium (obtained when the PoA is near or equal to zero) is achived for small values of λ_2 by increasing α_e or $\tilde{S}$.

3.6 Conclusion

This chapter presents a framework to analyze the trade-off between the successful data delivery probability and energy costs. We formulate the problem as a non-cooperative game in which each mobile has to decide which routing protocol it wants to use for packet delivering: Epidemic routing or Two-hop routing. We explore the scenario where the source and the destination mobiles are enclosed in two different regions, which are partially overlapped. We showed the impact of overlapping area on price of anarchy and Nash equilibrium. To complete this contribution, we plan to analyze the system when there are new arrivals to the area of interaction and mobiles within this area will be active for a limited period of time. This configuration makes the system dynamic in the number of mobiles, a more realistic approach to a DTN case.

Chapter 4

Evolutionary forwarding games in Delay Tolerant Networks: equilibria, mechanism design and stochastic approximation

4.1 Introduction

Cellular telecommunication networks have enabled voice and data communications for mobile users, achieving global connectivity. This crucial advance has revolutionized everyday habits in our society: nevertheless, mobile applications are posing new technical challenges. Telecommunication networks in fact are nowadays struggling to support the fast adoption of data-centric applications running on newly available platforms such as M2M modules, smart-phones or tablets. Thus, operators of telecommunication networks are seriously considering off-loading techniques such as encouraging the use of Wifi at home. Furthermore, Internet protocols have been designed for network access with very slow user's mobility and cannot efficiently support mobile applications. This is mainly due to the intermittent connectivity with the infrastructure as experienced, e.g., in the metro or in very dense areas.

But, mobile applications call also for more flexible ways to exchange data, e.g, in ad hoc mode with other co-localized users. Thus, next generation wireless systems are expected to enable versatile and flexible communications between mobile and autonomous users even in case where no infrastructure is available. Such flexibility comes at a price though. In these regimes, in fact, due to nodes' mobility, network topology may change rapidly and in a non-deterministic way in time. All customary network functionalities such as topology discovery, routing and messaging have therefore to be handled by the mobile nodes, thereby creating the need for efficient decentralized algorithms. The design of such algorithms under continuous topology changes and using only local information available at mobiles requires a specific design effort; this is also

the main motivation for this work.

On one hand, high mobility and frequent network partitioning rule out Internet routing protocols which operate poorly under uncertain networking conditions, high mobility and frequent network partitioning. But, on the other hand, many users carry advanced computing devices such as smart-phones, netbooks, etc. Such devices are equipped with wireless interfaces so that it is possible to sustain communication by leveraging intermediate nodes acting as relays, the so called *carry-store-and-forward* strategy. Messages can arrive at their destination thanks to the mobility of some subset of nodes that carry copies of the message stored in their local memory. Networks with such characteristics are named in literature Delay (or Disruption) Tolerant Networks (DTNs). Sometimes they are also known as opportunistic networks because communication opportunities appear at random, e.g., when a novel device enters radio range of a mobile node, and data exchange is possible while in radio range. However, as a result of mobility, a full path between source and destination may break too frequently or may never exist, thus preventing the adoption of end-to-end communications.

The idea of designing DTNs to sustain communications in spite of lack of persistent end-to-end connectivity is well documented in literature [5, 99, 70] and several real experiments over DTNs were performed in the past [56, 31] [1]

Yet, the fundamental issue for DTNs is how to trade-off between the number of released message copies and the delay for a message to reach the destination. In fact, the more the copies relays have, the higher the probability to deliver the message within a given time-line, but, the more the energy the network consumes.

As a consequence, several mechanisms for message forwarding in DTNs have been proposed. For instance, if mobiles having the message keep on relaying to any other mobile that enters its transmission range and which does not have the message yet [127], one would maximize the delivery probability. This is the well known epidemic routing, which is very expensive in terms of network resources. A smarter forwarding strategy such as the two hop routing protocol acts in a more efficient way. The source transmits copies of its message to all mobiles it encounters: relays in turn are forced to relay the message only to the destination.

In what follows, we confine our analysis to the two hop routing protocol because of two major advantages: first, it performs natively a good trade-off between the number of released copies and the delivery probability [36]. Second, forwarding control can be implemented on board of the source node: under two hop routing, the source can control the number of message copies released to mobiles it encounters. As it will be clear in the rest of this work, this is a convenient feature to connect the game-theoretical mechanism design and the stochastic approximation techniques in order to attain decentralized blind online optimal control.

In literature, several previous works elaborated mechanisms for efficient message

[1] We observe that DTNs in principle do not restrict to mobile networks: the DTN concept was actually conceived for satellite and space communications where long periods of out-of-reach conditions were inherently part of the communication requirements.

forwarding in DTNs using tools from both control theory and game theory (see Sec. 4.2 for related works). In this chapter, we make use of *evolutionary games*, which are novel in this context: within such framework, we are able to model the competition of relays in a DTN and reduce it to a distributed control problem. Compared with existing literature the forwarding policy is determined by strategies played by the relay nodes. Even more important, forwarding dynamics is determined by the fraction of the population of relays that comply to each given strategy. There are two advantages of this novel approach: first, we can provide a strong notion of equilibrium for the system which permits to identify robustness of stable system configurations, namely Evolutionary Stable Strategies (ESSs) – ESSs are defined against deviations of a certain fraction of the population of mobiles. Second, we can apply the general convergence theory of replicator dynamics, and several stability results that we introduce in future sections.

We consider that relays can adopt two types of behaviors: they can either undergo full activation or decide for partial activation for energy saving reasons. It is clear that devices owners would indeed turn off any relaying functionality for the sake of battery lifetime. However, a rewarding mechanism can be designed to incentive relays to participate to the forwarding process. This is the main contribution of this chapter: we model the competition of relays as a distributed control problem where the forwarding policy is determined by the strategies played by the relays themselves in order to increase their utility.

An additional degree of freedom is that the strategies played by relays evolve with time, e.g., due to some periodic strategy revision policy. Hence, we characterize the equilibria of the DTN forwarding dynamics. It is possible to identify cases in which at ESS, only one population prevails (namely, an ESS in pure strategies) and others, in which an equilibrium between several population types is obtained (namely, ESS in mixed strategies). Once determined the possible ESS equilibrium, we could also determine feasibility conditions for optimal forwarding control at the source node through a controlled rewarding policy. Observe the novelty compared to similar control problems seen in DTNs literature: here the source node cannot hope to just increase the success delivery probability by unilaterally increasing the number of released message copies. The point is that in some cases this unilateral action results in decreasing the utility of relays and reducing the number of those which take part to message forwarding with full activation. Finally, this in turn can produce exactly the opposite effect, i.e., the success probability decreases.

4.1.1 Main contributions

In this chapter, we aim at devising a so called *mechanism design* for controlling the evolutionary dynamics through the choice of appropriate forwarding control at the source. In this way it is possible to govern the replicator dynamics: such an approach provides very interesting insight into the feasibility of optimal mechanism design and, the technique appears per se novel compared to known results in literature.

Finally, we propose a hierarchical algorithm that allows the source to achieve the

optimal forwarding control, with the objective of maximizing the probability of success. The proposed algorithm is based on stochastic approximation theory. As clarified in the related section, the power of such techniques lies in the capability to infer online the optimal control in a blind fashion.

The chapter is organized as follows. In Sec.4.2 we report on previous works in the field of DTNs and evolutionary games. In Sec. 4.3 we introduced the network model and notions of evolutionary game theory used in the chapter. The equilibria of the system are also characterized in Sec. 4.3.5. A specific reward mechanism that will be used in the rest of the chapter is presented in Sec. 4.4, whereas Sec. 4.5 is devoted to mechanism design techniques for both static and dynamic control of forwarding at the source node. A stochastic approximation algorithm is proposed in Sec 4.6 and numerical outcomes are collected in Sec. 4.7. We eventually conclude our developments in Sec. 4.8.

4.2 Related Works

Several previous works address the control of forwarding schemes in DTNs [122, 83, 4, 36]. The work [4] proposed to control two hop forwarding and optimized the system performance by choosing the average duration of timers for message discarding. In [57], the authors describe an epidemic forwarding protocol and show that it is possible to increase the message delivery probability by tuning the parameters of the underlying SIR model. In [127], the authors studied optimal static and dynamic control problems using a fluid model that represents the mean field limit as the number of mobiles becomes very large. In [122] and its follow-up [114], the authors optimize network performance by designing the mobility of message relays. Some other works that explicitly address the control of forwarding and related to our work are [6, 83]. The optimality of a threshold type policy, already established in [36] for the fluid limit framework, was shown to hold in [35] for the actual discrete control problem. Another work [33] proposes the optimization of two hop forwarding based on the theory of linear-quadratic regulators.

Energy consumption is a major issue that should be taken into account in order to increase the lifetime of the network. Some devices may have very limited energy sources, like a small battery. The main issue is how nodes with finite energy budget can optimally decide that if and when to activate in order to be able to take part to the forwarding protocol. Several solutions have been proposed to overcome this problem in a homogeneous network [34], [52], [85].

Evolutionary games

The theory of evolutionary games has been developed by biologists to predict population dynamics in the context of *interactions between populations* [106]. Although ESS has been defined in the context of biological systems, it is highly relevant to engineering as

well [34, 52]. Global dynamics is determined through the description of local interactions, i.e., interactions that characterize the competition within a certain portion of the population (in our case, the local interaction will be represented by the delivery of a message from a certain source to a certain destination). This formalism studies Evolutionary Stability, and *Evolutionary Game Dynamics*. The Evolutionarily Stable Strategy (ESS), first defined in [106], is characterized by robustness against invaders (mutations): (i) if an ESS is reached, then the proportions of each population do not change in time, (ii) at ESS, the populations are immune from being invaded by other small populations. Observe that this notion is stronger than Nash equilibrium: for a Nash equilibrium, in fact, it is only requested that an *individual* would not benefit by a change (mutation) of its behavior. Standard references for evolutionary games are [124] and [121]. In the biological context, the replicator dynamics is a model to explain observed variations in populations' size. In engineering, we can go beyond characterizing and modeling existing evolution. The evolution of protocols in DTNs can be engineered by providing guidelines or regulations for the way to upgrade existing ones and to determine critical parameters which can impact both the performance of the network and services deployed on top of the network through a properly designed incentive mechanism.

4.3 Network Model

We consider a Delay Tolerant Network with several sources s_i, destinations d_i and a large number of mobiles acting as relay nodes in the system. We assume mobiles are randomly placed over a plan following a distribution process. Each mobile is equipped with a wireless interface allowing communication with other mobiles in their proximity. Messages are generated at the source nodes and need to be delivered to the destination nodes; however, each such message is relevant for a time interval of length τ: this is also the horizon by which we intend to optimize network performance. The network is assumed to be sparse: at any time instant, nodes are isolated with high probability.[2] Nevertheless, due to mobility patterns, communication opportunities arise whenever two nodes get within mutual communication range, i.e., a "contact" occurs. The time between subsequent contacts between any two nodes is assumed to follow an exponential distribution with parameter λ. The assumption for synthetic mobility models has been discussed in [4] and has been widely adopted to make analytical model tractable [127, 96]. Consider a message generated at $t = 0$: each source node attempts to deliver the message to its destination; it does so eventually with several copies spread between the relays nodes. Each such message contains a time stamp reporting its age and can be deleted when it becomes irrelevant, e.g., after time τ. Due to lack of permanent connectivity, we exclude the use of feedback that allows the sources or other mobiles to know whether the message has been successfully delivered to its destination or not. For the same reason, the design of our activation mechanism should not require centralized coordination and any such scheme should indeed run fully distributed on board of the

[2] This is also the case when disruption caused by mobility occurs at a fast pace compared to the typical operation time of protocols, e.g., the TPC/IP protocol suite.

relay nodes. This is a standard description of DTN frameworks that we will be using in the following chapters as well.

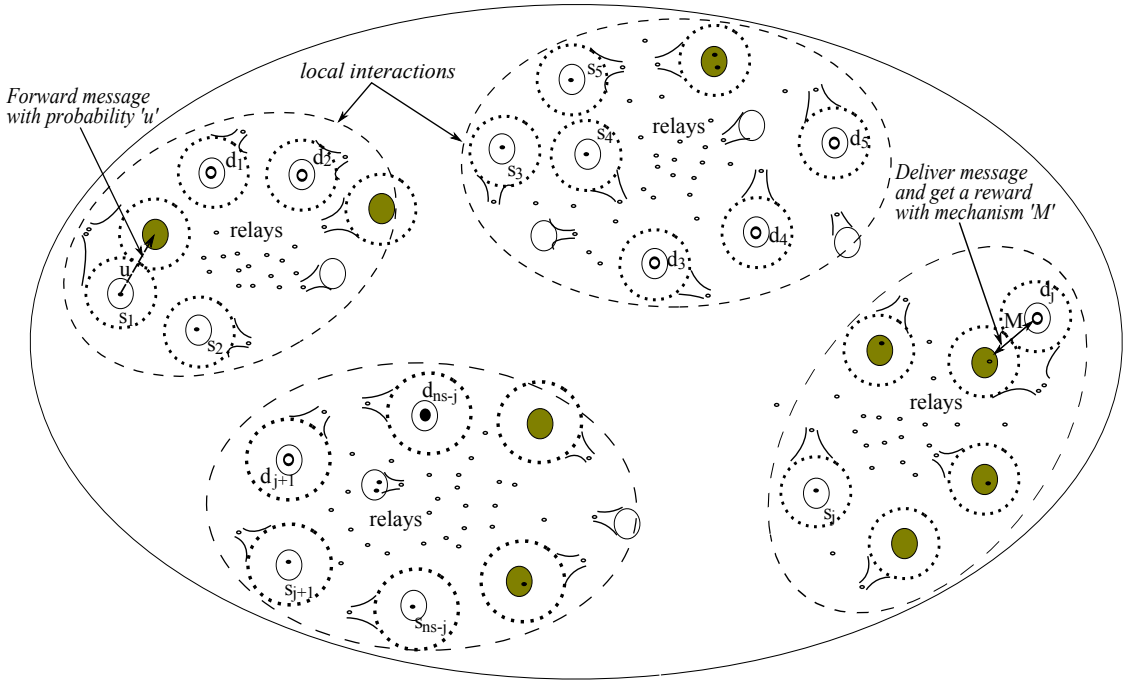


Figure 4.1: *Snapshot of the network: the figure is portraying the typical working conditions for the DTN.*

4.3.1 Network Game

In a conventional game, the objective of a player is to choose a strategy that maximize its utility. In contrast, evolutionary games are played repeatedly by players drawn from a population. In this section we apply evolutionary games to non-cooperative 'live time' selection: the aim is to study a competition framework among individual mobile relays. Basically, we formulate a DTN routing game, where the relays represent the players of the game. There are again many local interactions among players belonging to large populations of mobile relays. Those players need to take some actions with respect to messages that are generated by severals source-destination pairs in a local interaction (see figure 4.1).

Now, we detail the utility structure of the proposed mechanism. When a message is generated by source nodes in a local interaction, the competition takes place during the message lifetime, i.e., with duration τ. The message live time in turn is dictated by the strategy adopted by relays: a relay node can decide the duration of the interval of time during which it participates to the forwarding process. For simplicity, let us assume that each relay node chooses two different live times for the message: τ, i.e., full activation, and τ', i.e., partial activation. Without loss of generality let $\tau' < \tau$. Notice that switching off the radio interface at $\tau' < \tau$ acts as a power saving technique since

the relay reduces battery depletion due to idle listening and beaconing. Moreover, relay interacts severals times with other relays, the actions of the relay along with those with which it interacts determine the immediate reward of the relay. An mobile relay can be seen as trying to maximize the expected sum of its immediate reward during the game. Without loss of generality, we assume that at a local interaction, there is n_s source-destination pairs in which the source has packet generated at some time $t_i = i\tau$, which is then relevant up to time $t_i + \tau = t_{i+1}$, when another message is released and local interaction $(i+1)$-th begins. In that local interaction, a mobile relay may interact with the actions of some N (possibly random number of) other relays. Hence, the immediate reward is designed in such a way that, upon successful delivery of message to destinations, a relay can receive a reward based on the reward mechanism used by the source-destination pairs in the system. In what follows, there are also a few assumptions that are not standard in evolutionary games; they follow from the fact that, in the type of games we deal with, a message is possibly delivered leveraging several relays, i.e., players, at once. To this respect, in our framework, a fundamental assumption is that the incentive to relay a message is a certain reward that each relay may receive upon delivering the message. However, competing relays perceive each other as *interfering* to their final goal to attain the reward. Let $(y, 1-y)$ be the distribution of strategies among the population of relays, where y represents the proportion of mobile relays choosing strategy τ (full active). A distribution y is sometimes called the "state" or the "profile" of the population.

- The number of mobile relays interfering with a given randomly selected user is a random variable $K \in \mathbb{N}$.
- A mobile relay does not know how many players would interfere with it.
- Let K be the number of relays in a local interaction and $P^s_{succ}(a, K, y)$ be the probability that a relay receives a reward from the source s when its action is a and profile of the population is y. By assuming that all relays have enough buffer capacity to store all messages forwarded by sources in a local interaction, the expected immediate reward can be written as

$$R(a, K, y, S) = \sum_{(s,d)\in S} r^{s,d} P_{succ}(a, K, y, (s, d)) - ga$$

  where S is the set of source-destination pairs in a local interaction and g is the energy spent by a relay by unit of time when it remains active. We observe that during the interval $[0, \tau']$, all relays are active and the probability that a relay receives a reward before τ' doesn't depend on the profile of population, i.e., $P_{succ}(\tau', K, y, (s, d)) = P_{succ}(\tau', K, (s, d))$ for all $(s, d), y$.

  For ease of presentation, in the following we consider the homogeneous case:

  - $P_{succ}(a, K, y, (s, d)) = P_{succ}(a, K, y)$, in which the relay and sources have similar physical characteristic, e.g. transmission range, mobility patterns, energy capacities, and
  - The reward mechanism is fixed for all source-destination pairs, i.e., $r^{s,d} = r$ for all (s, d).

– The number of source-destination pairs in a local interaction is a random variable follows a general (arbitrary) distribution with mean n_s.

The expected reward becomes

$$R(a,K,y) = n_s r P_{succ}(a,K,y) - ga$$

The expected payoff of a relay playing strategy a when the state of the population is y, is given by

$$U_{av}(a,y) = \sum_{k\geq 0} \mathbb{P}(K=k) R(a,K,y),$$

Hence the average payoff of a population in state y is given by

$$F(y,y) = \sum_{i=1}^{I} y_j U_{av}(j,y).$$

- The game is played many times; we call each round of the game a *local interaction* and there are many local interactions at the same time;

Remark 4.3.2. *We observe that in our system, relays do not need to be synchronized to the source clock. In fact, it is sufficient that they decide their strategy at the time when they meet the source. At that time, they can be made aware of the deadline τ using a time-to-live counter that is decreased over time at the source node.*

Remark 4.3.3. *Our model may also cover the scenario in which a mobile node can be a source and a relay at once. By assuming that the arrival of messages at a mobile node follows a Poisson process with rate μ and the mobile does not generate new message as long as a previous message is not yet transmitted, the distribution of the number of source-destination pairs in a local interaction with K relays can be computed a follows:*

$$\mathbb{P}(N_s = n_s) = C_{n_s}^{K} (q_a)^{n_s} (1-q_a)^{K-n_s}$$

where q_a is the probability that a source generates a new message during a slot. In sight of the Poisson arrival assumption, we have $q_a = 1 - e^{-\mu\tau}$. Hence, the expected reward becomes

$$\begin{aligned} R(a,K,y) &= \sum_{n_s=1}^{K} \mathbb{P}(N_s = n_s)\; r P_{succ}(a,K,y) - ga \\ &= K q_a r P_{succ}(a,K,y) - ga \end{aligned}$$

4.3.4 Evolutionary Stable Strategy

The evolutionarily stable strategy (ESS), first defined in [106], is characterized by robustness against invaders (mutations). This notion is stronger than Nash equilibrium in which it is only requested that a single user would not benefit by a change (mutation) of its behavior. Although ESS has been defined in the context of biological systems, it is highly relevant to engineering as well [37, 116]. There are two advantages in doing so within the framework of evolutionary games:

- it provides the stronger concept of equilibria, the ESS, which allows us to identify robustness against deviations of more than one mobile, and
- it allows us to apply the generic convergence theory of replicator dynamics, and stability results that we introduce in future sections.

Now, we detail the ESS concept for our network game. Suppose that, initially, the population profile of relays is $(y, 1-y)$. Now suppose that a small group of mutants relays enters the population playing according to a different profile $(mut, 1-mut)$. If we call $\epsilon \in (0,1)$ the size of the subpopulation of mutants relays after normalization, then the population profile after mutation will be $\epsilon \cdot mut + (1-\epsilon)y$. After mutation, the average payoff of non-mutants will be given by

$$F(y, \epsilon \cdot mut + (1-\epsilon)y).$$

Note that U_{av} need not to be linear in the second variable. Analogously, the average payoff of a mutant is

$$F(mut, \epsilon \cdot mut + (1-\epsilon)x).$$

Notice that the second argument of the average payoff is expressing the "average" interferer profile that is faced by a relay picked at random.

Definition 4.3.4.1. *A strategy $y^* \in M$ is an ESS if for any $mut \neq y^*$, there exists some $\epsilon_{mut} \in (0,1)$, which may depend on mut, such that for all $\epsilon \in (0, \epsilon_{mut})$ one has*

$$F(y^*, \epsilon\, mut + (1-\epsilon)y^*) > F(mut, \epsilon \cdot mut + (1-\epsilon)y^*)$$

which can be rewritten as

$$\sum_{j=1}^{N} (y_j^* - mut_j) U_{av}(j, \epsilon \cdot mut + (1-\epsilon)y^*) > 0.$$

That is, y^* is an ESS if, after mutation, non-mutants are more successful than mutants. Borrowing the expression from population dynamics [102], under an ESS profile, mutants cannot invade the population and will eventually get extinct.

When the ESS is such that the whole population plays a certain pure strategy, we say that the ESS is in pure strategy. If not, we say the ESS is in mixed strategies. We now state the assumption required for the function $P_{succ}(\tau, k, s)$

Assumption A

The function $P^s_{succ}(\tau, k, Y)$ is decreasing in y, i.e., number of active relays.

The above assumption reflects that the bigger the number of relays participating to the message delivery with full activation, the higher the delivery probability for the message, but indeed the less the chance for the tagged relay to receive a reward from the system. The global activation target settles the number of opponents of a randomly tagged relay, i.e., the active fraction of the population.

4.3.5 Existence and uniqueness of ESS

In the rest of the chapter, we will employ an auxiliary function $\tilde{H}$ defined as

$$\begin{aligned} H : y \in (0,1) \quad &\rightarrow \quad U_{av}(\tau, y) - U_{av}(\tau', y) \\ &= \sum_{N=1}^{\infty} P(K = N)\Big(P_{succ}(\tau, K, y) - P_{succ}(\tau', K)\Big) - \frac{g(\tau - \tau')}{n_s r} \end{aligned} \quad (4.1)$$

In this section, we characterize the existence and uniqueness of the ESS. A very compact result exists that ties together the main parameters of the system:
Theorem 4.3.5.1. *The ESS exists and it is unique, furthermore*

(1) The strategy τ dominates the strategy τ' if and only if

$$\sum_{N=1}^{\infty} P(K = N)\Big(P_{succ}(\tau, K, 1) - P_{succ}(\tau', K)\Big) - \frac{g(\tau - \tau')}{n_s r} \geq 0.$$

so that $y = 1$ is the ESS in pure strategy

(2) The strategy τ' dominates the strategy τ if and only if

$$\sum_{N=1}^{\infty} P(K = N)\Big(P_{succ}(\tau, K, 0) - P_{succ}(\tau', K)\Big) - \frac{g(\tau - \tau')}{n_s r} \leq 0$$

so that $y = 0$ the ESS in pure strategy

(3) If $\sum_{N=1}^{\infty} P(K = N)\Big(P_{succ}(\tau, K, 0) - P_{succ}(\tau', K)\Big) - \frac{g(\tau - \tau')}{n_s r} > 0$ and $\sum_{N=1}^{\infty} P(K = N)\Big(P_{succ}(\tau, K, y) - P_{succ}(\tau', K)\Big) - \frac{g(\tau - \tau')}{n_s r} < 0$, then there exists an unique ESS y^* which is given by

$$y^* = H^{-1}\Big(0\Big)$$

Proof. (1) The strategy τ dominates the strategy τ' if and only if $U_{av}(\tau, y) \geq U_{av}(\tau', y)$ for all $y \in [0,1]$. Since H is a decreasing function and $H(1) = \sum_{N=1}^{\infty} P(K = N)\Big(P_{succ}(\tau, K, 1) - P_{succ}(\tau', K)\Big) - \frac{g(\tau - \tau')}{n_s r} \geq 0$, thus $H(y) = U_{av}(\tau, y) - U_{av}(\tau', y) > 0$ for all $y \in [0,1]$. This completes the proof for (1)

(2) The strategy τ' dominates the strategy τ if and only if $U_{av}(\tau, y) \leq U_{av}(\tau', y)$ for all $y \in [0,1]$. Since the function H is a decreasing function and $H(0) = \sum_{N=1}^{\infty} P(K = N)\Big(P_{succ}(\tau, K, 0) - P_{succ}\tau', K)\Big) - \frac{g(\tau - \tau')}{n_s r} \leq 0$, thus $H(y) = U_{av}(\tau, y) - U_{av}(\tau', y) \leq 0$ for all $\beta \in [0, (1 - Q_\tau)]$. This completes the proof for (2).

(3) A strictly mixed equilibrium y^* is characterized $U_{av}(\tau, y^*) = U_{av}(\tau', y^*)$ i.e $H(y^*) = 0$. The function H is continuous and strictly decreasing monotone on $(0,1)$ with $H(0) > 0$ and $H(1) < 0$. Then the equation $H(y^*) = 0$ has a unique solution in the interval $(0,1)$. This completes the proof.

□

4.4 Reward mechanism : First place winner

In this section, we consider the following mechanisms used by the system for receiving a reward by relays. In particular, for each message generated by a source, a mobile may receive a unit of reward if it is the first to deliver a copy of the packet to the destination. A utility function is introduced as the difference between the expected reward and the energy cost, i.e., the cost spent by the relay to sustain forwarding operations. Hence, during a certain local interaction, the probability for a tagged mobile to relay the copy of the packet to the destination within live time τ is then given by $1 - Q_\tau$ where Q_τ is given by

$$Q_\tau = (1 + \lambda\tau)e^{-\lambda\tau}$$

and the probability that it relays the copy of the message if it chooses live time τ' is given by $1 - Q_{\tau'}$ where $Q_{\tau'} = (1 + \lambda\tau')e^{-\lambda\tau'}$. Hence the probability that a relay receives a reward from a source when its action is τ', becomes

$$P_{succ}(\tau', N) = (1 - Q_{\tau'}) \sum_{k=1}^{N-1} C_{k-1}^{N-1} \frac{(1 - Q_{\tau'})^{k-1}(1 - (1 - Q_{\tau'}))^{N-k}}{k} = \frac{1 - Q_{\tau'}^N}{N} \quad (4.2)$$

The utility for a mobile using live time τ' is

$$U_{av}(\tau', y) = \sum_{N=1}^{\infty} P(K = N) \frac{r \cdot n_s(1 - Q_{\tau'}^N)}{N} - g\tau'$$

where g is energy cost spent for being active during a unit of time. The linear energy expenditure, in particular, may express the cost of activating the RF interface and the cost of periodic beaconing. Now the probability that a mobile receives the unit award, if it chooses live time τ, is given by

$$\begin{aligned} P_{succ}(\tau, N, y) &= P_{succ}(\tau', N, y) + (Q_{\tau'})^N \beta \sum_{k=1}^{N-1} C_{k-1}^{N-1} \frac{\beta^{k-1} y^{k-1}(1 - y\beta)^{N-k}}{k} \\ &= P_{succ}(\tau', N) + (Q_{\tau'})^N \frac{1 - (1 - \beta y)^N}{Ny} \end{aligned} \quad (4.3)$$

where $\beta = 1 - \dfrac{Q_\tau}{Q_{\tau'}} = 1 - \dfrac{1 + \lambda\tau}{1 + \lambda\tau'} e^{-\lambda(\tau - \tau')}$.

From theorem 4.3.5.1, we have directly the following results :

Corollary 4.4.1. *The ESS exists and it is unique, furthermore*

(1) The strategy τ dominates the strategy τ' if and only if

$$\sum_{N=1}^{\infty} P(K=N)\frac{Q_{\tau'}^N(1-(1-\beta)^N)}{N} \geq \frac{g(\tau-\tau')}{r\cdot n_s}.$$

so that $y=1$ is the ESS in pure strategy

(2) The strategy τ' dominates the strategy τ if and only if

$$\sum_{N=1}^{\infty} P(K=N)Q_{\tau'}^N\beta - g(\tau-\tau') \leq \frac{g(\tau-\tau')}{r\cdot n_s}$$

so that $y=0$ the ESS in pure strategy

(3) If $\sum_{N=1}^{\infty} P(K=N)Q_{\tau'}^N\beta > \frac{g(\tau-\tau')}{r\cdot n_s}$ and $\sum_{N=1}^{\infty} P(K=N)\frac{Q_{\tau'}^N(1-(1-\beta)^N)}{N} < \frac{g(\tau-\tau')}{r\cdot n_s}$, then there exists an unique ESS y^* which is given by

$$y^* = H^{-1}\Big(0\Big)$$

The above result states that we should always expect a unique ESS; also we should either expect a ESS in pure strategy, namely in cases (1) and (2), or a unique ESS in mixed strategies, namely in case (3). In the following we consider specific examples for the above result in the case when the statistics of the number of nodes K meeting in local interactions is known.

4.4.2 Poisson distribution

Let nodes be distributed over a plane following a Poisson distribution with density γ. The probability that there exist N nodes during a local interaction is given by the following distribution: $\mathbb{P}(K=k) = \frac{\gamma^{k-1}}{(k-1)!}e^{-\gamma}$, $k \geq 1$. The expression in Thm. 4.3.5.1(3) can be derived in closed form, since the unique ESS y^* is the unique solution of the following equation: $\frac{e^{-\gamma(1-Q_{\tau'})}(1-e^{-Q_{\tau'}\beta y^*\gamma})}{\gamma y^*} = \frac{g(\tau-\tau')}{r\cdot n_s}$. Thus, the equilibrium is given by

$$y^* = \frac{1}{\gamma Q_{\tau'}\beta}\Big(LambertW(-Ze^{-Z}) + Z\Big)$$

with $Z = \frac{r\cdot n_s\, Q_{\tau'}\beta e^{-\gamma(1-Q_{\tau'})}}{g(\tau-\tau')}$ and $LambertW$ is the inverse function of $f(u) = ue^u$.

4.4.3 Dirac distribution

This is the case when there exists a fixed number of nodes in a local interaction. We suppose that the population of nodes is composed by many local interactions between

N nodes where $N > 2$. The unique ESS y^* of this game is the unique solution of the following equation:

$$(Q_{\tau'})^N - (Q_{\tau'} - Q_\tau y)^N = y\frac{g(\tau - \tau')}{r \cdot n_s}$$

Since this polynomial has degree N, we can have an explicit expression only for $N \leq 5$. We find here the closed form expression of the ESS for some values of N. For example:

$$N = 2 \implies y(2Q_{\tau'}Q_\tau - Q_\tau^2 y - \frac{g(\tau - \tau')}{r \cdot n_s}) = 0$$

which gives, $y^* = 0$ or $y^* = \frac{1}{Q_\tau^2}(2Q_\tau Q_{\tau'} - \frac{g(\tau - \tau')}{r \cdot n_s})$. We also observe that $y^* > 0 \iff 2Q_\tau Q_{\tau'} > \frac{g(\tau - \tau')}{r \cdot n_s}$.

For $N = 3$, y^* is the solution of the equation

$$yQ_\tau(y^2 + 3Q_\tau Q_{\tau'} y - 3Q_{\tau'}^2 + \frac{g(\tau - \tau')}{r \cdot n_s Q_\tau}) = 0$$

The feasible solutions are 0, and the pair $(\frac{3Q_{\tau'} + \sqrt{k}}{2Q_\tau}, \frac{3Q_{\tau'} - \sqrt{k}}{2Q_\tau})$, given that $\frac{g(\tau - \tau')}{r \cdot n_s} > \frac{3}{4}Q_{\tau'}^2 Q_\tau$ for the positivity of the discriminant.

4.5 Mechanism design

It sight of the characterization of the ESS for the system described before, we are interested in controlling the system in order to optimize for energy consumption and delivery probability. Let us assume that the system controls the forwarding of message copies: during a local interaction a copy of the message is relayed with probability $u = u(t)$ upon meeting a node without a message copy, i.e., using a static or dynamic forwarding policy [36]. In our analysis, the main quantity of interest is denoted P_s and it is the success probability of a message at a local interaction; under the same assumptions of linearity in [36], the average energy expenditure in a local interaction is $\mathcal{E} = \varepsilon\Psi$, where $\varepsilon > 0$ is the source energy expenditure per relayed message copy and Ψ is the corresponding expected number of copies released at that local interaction.

4.5.1 Static forwarding policy

Let us consider first the static forwarding control, i.e., u is a constant. We assume that the sources wish to control the system in order to optimize for the energy consumption, i.e., the number of message copies, and the delivery probability by static forwarding

policies. In this case, a tagged mobile relay playing τ, (resp. τ') may deliver a copy of the message to the destination within time τ (resp. τ') with probability $1-Q^u_\tau$ (resp. $1-Q^u_{\tau'}$) where Q^u_τ is given by $Q^u_\tau = \dfrac{e^{-\lambda u\tau} - ue^{-\lambda\tau}}{1-u}$ and $Q^u_{\tau'} = \dfrac{e^{-\lambda u\tau'} - ue^{-\lambda\tau'}}{1-u}$ so that the success probability in a local interaction with N mobiles at the equilibrium is

$$\begin{aligned} P_s(u,N) &= 1 - \Big[\Big(Q^u_{\tau'}\Big)^{N(1-y(u))} \cdot \Big(Q^u_\tau\Big)^{Ny(u)}\Big] \\ \Rightarrow P_s(u) &= 1 - \sum_{k=0}^{\infty} P(N=k)\Big[\Big(Q^u_{\tau'}\Big)^{k(1-y(u))} \cdot \Big(Q^u_\tau\Big)^{ky(u)}\Big] \end{aligned} \tag{4.4}$$

where $y(u)$ is the fraction of mobiles playing action τ when k nodes are present in the local interaction. We can determine the the above the expressions of Q^u_τ and $Q^u_{\tau'}$ by letting A, B two random variables determining the time spent by a given relay to respectively, receive a copy of the message from the source and deliver the message to the destination once it was received at the source node.

$$\begin{aligned} Q^u_{\tau'} &= 1 - P(A+B \le \tau') = 1 - \int_0^{\tau'} P(B \le \tau' - t \mid A = t)\, P(A=t)dt \\ &= 1 - \int_0^{\tau'} (1 - e^{-\lambda(\tau'-t)})\, \lambda u e^{-\lambda t} dt \\ &= 1 - (-e^{-\lambda u\tau'} + 1) + u\int_0^{\tau'} (\lambda e^{-\lambda(\tau'-t+ut)})\, dt \\ &= e^{-\lambda u\tau'} + \frac{u}{1-u}(e^{-\lambda\tau'} + e^{-\lambda\tau' u}) \\ &= \frac{e^{-\lambda u\tau'} - ue^{-\lambda\tau'}}{1-u}. \end{aligned}$$

We replace τ' by τ to find the expression of Q^u_τ. It is then possible to define all relevant quantities as a function of u, i.e., $U^u(\tau, y)$, $U^u(\tau', y)$ and also $H^u(\cdot, k)$. Recall that the system wants to maximize the delivery probability of the message to the destination and meet a given constraint on the energy expenditure, i.e., message copies. The expected number of copies generated by a sources given k nodes in a local interaction is given by the expected number relays the source meets in $[0, \tau']$, namely $k \cdot (1 - e^{-\lambda u\tau'})$, plus the expected number of relays under full activation which the source meets in $[\tau', \tau]$, namely $k \cdot y(u) \cdot (e^{-\lambda u\tau'} - e^{-\lambda u\tau})$. Since there is (in average) n_s source-destination pairs, the constraints can be expressed as

$$n_s\Big(\mathbf{E}\{N\}(1 - e^{-\lambda u\tau'}) + \mathbf{E}\{Ny(u)\}(e^{-\lambda u\tau'} - e^{-\lambda u\tau})\Big) = \Psi \tag{4.5}$$

Using the previous relation, it is possible to design an optimal control mechanism at each source in the system.

Proposition 4.5.2. *Let us assume $0 < \tau' \le \tau$, then the following holds:*

1. *Let* $u^* = min\left\{\frac{-\ln(1-\frac{\Psi}{n_s \cdot \boldsymbol{E}\{N\}}))}{\lambda\tau}, 1\right\}$*: if* $\sum_{N=1}^{\infty} P(K=N)(Q_{\tau'})^N\beta \leq \frac{g(\tau-\tau')}{r\cdot n_s}$*, then* u^* *is the optimal control. If not go to step 2.*
2. *Let* $u^* = min\left\{\frac{-\ln(1-\frac{\Psi}{n_s \cdot \boldsymbol{E}\{N\}}))}{\lambda\tau'}, 1\right\}$*: if* $\sum_{N=1}^{\infty} P(K=N)\frac{Q_{\tau'}^N(1-(1-\beta)^N)}{N} \geq \frac{g(\tau-\tau')}{r\cdot n_s}$*, then* u^* *is the optimal control. If not go to step 3.*
3. *Let* u^* *solve for* $\frac{\frac{\Psi}{n_s} - \boldsymbol{E}\{N\}(1-e^{-\lambda u\tau'})}{e^{-\lambda u\tau'} - e^{-\lambda u\tau}} = \boldsymbol{E}\left\{NH_u^{-1}(0,N))\right\}$*, then* u^* *is the optimal control.*

Notice that in this case an optimal static forwarding policy always exists: this is a consequence of the fact that all relays are active. The proof of proposition 4.5.2 derives directly from relation (eq:constraints). Also, should we consider the case $0 = \tau' \leq \tau$, then the following result on the monotonicity of the ESS holds.

Corollary 4.5.3. *If we consider the case of activation control (i.e* $\tau' = 0$*), then the ESS* $y^*(u)$ *is an increasing function of the source control* u*.*

Proof. The idea of the proof is to use in a first place the fact that the fitness function H, which depends on the utility attained for being active or not, is decreasing with the population profile y. Then we show that the function H is increasing in u to conclude eventually that the ESS y^* is also an increasing function of the source control u. Recall that $y^* = H^{-1}(0)$, where y^* is projected onto interval $[0,1]$.

(1) Let's show that H is a decreasing function of y. We borrow the expression of $H(y,u)$ from the live time control:

$$H(y,u) = r\cdot n_s \sum_{N=1}^{\infty} P(K=N)(Q_{\tau'}^u)^N \frac{1-(1-\beta y)^N}{Ny} - g(\tau-\tau')$$

Let $f(y) = \frac{1-(1-\rho^u y)^N}{Ny}$. It is easy to check If f is decreasing with y then H is also decreasing with y. We have

$$\begin{aligned}
\frac{df(y)}{dy} &= \frac{N\rho^u(1-\rho^u y)^{N-1}Ny - N\left(1-(1-\rho^u y)^N\right)}{(Ny)^2} \\
&= \frac{N\rho^u y(1-\rho^u y)^{N-1} - 1 + (1-\rho^u y)^N}{Ny^2} \\
&= \frac{(1-\rho^u y)^{N-1}(N\rho^u y + 1 - \rho^u y) - 1}{Ny^2} \\
&= \frac{(1-\rho^u y)^{N-1}((N-1)\rho^u y + 1) - 1}{Ny^2}
\end{aligned}$$

Let's show that $(1-\rho^u y)^{N-1}((N-1)\rho^u y+1)-1$ is negative. In fact we have

$$(1-\rho^u y)^{N-1}((N-1)\rho^u y+1)-1 \leq 0 \Leftrightarrow (N-1)\rho^u y+1 \leq \frac{1}{(1-\rho^u y)^{N-1}}$$

which is true since $\frac{d((N-1)\rho^u y+1)}{dy} \mid_{(y=0)} = \frac{d\left(\frac{1}{(1-\rho^u y)^{N-1}}\right)}{dy} \mid_{(y=0)} = \rho^u(N-1)$ and $\frac{d^2\left(\frac{1}{(1-\rho^u y)^{N-1}}\right)}{d^2 y} = \frac{(\rho^u)^2(N-1)(N+1)}{(1-\rho^u y)^{N+2}} > 0$. Thus H is a non-increasing function of y.

(2) Now we show that H is increasing with u. As done for the previous point, we use the function f. For fixed y, we have,

$$\frac{df(y,u)}{du} = \frac{d\rho^u}{du}(1-\rho^u y)^{N-1}$$

Since ρ^u is increasing then $\frac{df(y,u)}{du} > 0$. It follows that H is increasing with u.

From (2) we know that H is increasing with u for any given y. This implies that $if \exists\ u\ s.t\ H(u)=0$ for a given y then, u is unique in $[0,1]$.
From (1) we have, $\forall(y_1, y_2)\ s.t.\ y_1 \leq y_2, \Rightarrow H(y_1,u) \geq H(y_2,u)$. It follows that $if \exists\ u_1, u_2\ s.t.\ H(y_1,u_1) = H(y_2,u_2) = 0$ then $u_1 \leq u_2$ and $y^*(u)$ is an increasing function of u as stated in the corollary and this completes the proof. □

4.5.4 Dynamic forwarding policy

We now consider the case of dynamic forwarding policies, i.e., when control $u = u(t) \in [u_{min}, 1]$. Again, aim is to optimize for the energy consumption by trading off the number of message copies for the delivery probability.

In [35] authors show that optimal forwarding control at the source node has threshold form: the source should forward the message with probability u_{max} up to a certain time t_{th} from the message generation time, and with probability u_{min} after time t_{th}. Thus the control has form:

$$u(t) = \begin{cases} u_{max} & \text{if } t \leq t_{th} \\ u_{min} & \text{if } t > t_{th} \end{cases}$$

Without loss of generality we assume that $u_{max} = 1$. Taking into account the threshold policy, the expression for Q_τ and $Q_{\tau'}$ are slightly modified: in particular, a tagged mobile relay playing τ' (resp. τ) may deliver a copy of the message to the destination

within time τ' (resp. τ) with probability $1 - Q_{\tau'}(t_{th})$ (resp. $1 - Q_{\tau}(t_{th})$) where expression for $Q_{\tau'}(t_{th})$ (resp. $Q_{\tau}(t_{th})$) is given by

$$Q_{\tau'}(t_{th}) = \begin{cases} e^{-\lambda t_{th}}(1 + \lambda t_{th} e^{-\lambda(\tau' - t_{th})}) - \dfrac{u_{min} e^{-\lambda u_{min} t_{th}}}{1 - u_{min}}(1 - e^{-\lambda(\tau' - t_{th})}) & \text{if } t_{th} \leq \tau', \\ (1 + \lambda\tau')e^{-\lambda\tau'} & \text{if } t_{th} > \tau'. \end{cases} \tag{4.6}$$

and

$$Q_{\tau}(t_{th}) = e^{-\lambda t_{th}}(1 + \lambda t_{th} e^{-\lambda(\tau - t_{th})}) - \frac{u_{min} e^{-\lambda u_{min} t_{th}}}{1 - u_{min}}(1 - e^{-\lambda(\tau - t_{th})}) \tag{4.7}$$

so that the success probability in a local interaction with k mobiles at the equilibrium is

$$\begin{aligned} P_s(t_{th}, k) &= 1 - \left[\left(Q_{\tau'}(t_{th}) \right)^{k(1 - y(t_{th}))} \cdot \left(Q_{\tau}(t_{th}) \right)^{k y(t_{th})} \right] \\ \Rightarrow P_s(t_{th}) &= 1 - \sum_{k=0}^{\infty} P(N = k) \left[\left(Q_{\tau'}(t_{th}) \right)^{k(1 - y(t_{th}))} \cdot \left(Q_{\tau}(t_{th}) \right)^{k y(t_{th})} \right] \end{aligned} \tag{4.8}$$

where $y(t_{th})$ is the fraction of mobiles playing τ when k nodes are present in the local interaction. As done before for the static control, we can define all relevant quantities as a function of t_{th}, i.e., $U^{t_{th}}(\tau, y)$, $U^{t_{th}}(\tau', y)$ and also $H^{t_{th}}(\cdot, k)$. Considering a Poisson distribution of nodes, the ESS y^* is the unique solution of the following equation:

$$\frac{e^{-\gamma(1 - Q_{\tau'}(t_{th}))}(1 - e^{-Q_{\tau'}(t_{th})\beta(t_{th})y^*\gamma})}{\gamma y^*} = \frac{g(\tau - \tau')}{n_s r}$$

Hence, at the equilibrium we have:

$$y^*(t_{th}) = \frac{1}{\gamma Q_{\tau'}(t_{th})\beta(t_{th})} \left(LambertW(-Z(t_{th}) e^{-Z(t_{th})}) + Z(t_{th}) \right)$$

with $Z(t_{th}) = \dfrac{n_s r Q_{\tau'}(t_{th})\beta(t_{th}) e^{-\gamma(1 - Q_{\tau'}(t_{th}))}}{g(\tau - \tau')}$ and $\beta(t_{th}) = 1 - \dfrac{Q_{\tau}(t_{th})}{Q_{\tau'}(t_{th})}$.

The probability of success is:

$$P_s(t_{th}) = 1 - e^{-\gamma} \left[e^{\gamma (Q_{\tau'}(t_{th}))^{(1 - y(t_{th}))} \cdot (Q_{\tau}(t_{th}))^{y(t_{th})}} - 1 \right].$$

In the following section we will present a learning algorithm to find the threshold t_{th} as a function of the source control. Here we provide the derivation that leads to (4.6) and (4.7). Again, let A, B two random variables determining the time spent by a given relay to respectively, receive a copy of the message from the source and deliver the message

to the destination once it was received by the source node. If $t_{th} \geq \tau'$ then $u = 1$,

$$\begin{aligned} Q_{\tau'}(t_{th}) &= 1 - P(A + B \leq \tau') \\ &= 1 - \int_0^{\tau'} (P(B \leq \tau' - t | A = t) \cdot P(A = t)) dt \\ &= 1 - \int_0^{\tau'} (1 - e^{-\lambda(\tau' - t)}) \cdot \lambda e^{-\lambda t} dt \\ &= 1 + \tau' \lambda e^{-\lambda \tau'} + e^{-\lambda \tau'} - 1 \\ &= (1 + \tau' \lambda) e^{-\lambda \tau'}. \end{aligned}$$

Similarly, if $t_{th} \leq \tau'$, then

$$\begin{aligned} Q_{\tau'}(t_{th}) = & \quad 1 - \int_0^{t_{th}} (1 - e^{-\lambda(\tau' - t)}) \cdot \lambda e^{-\lambda t} dt - \int_{t_{th}}^{\tau'} (P(B \leq \tau' - t | A = t) \cdot P(A = t)) dt \\ = & \quad (1 + (\tau' - t_{th}) \lambda e^{-\lambda(\tau' - t_{th})}) e^{-\lambda t_{th}} - \int_{t_{th}}^{\tau'} (P(B \leq \tau' - t / A = t) \cdot P(A = t)) dt \\ = & \quad (1 + (\tau' - t_{th}) \lambda e^{-\lambda(\tau' - t_{th})}) e^{-\lambda t_{th}} - \int_{t_{th}}^{\tau'} (1 - e^{-\lambda(\tau' - t)}) \cdot u_{min} \lambda e^{-u_{min} \lambda t} dt \\ = & \quad e^{-\lambda t_{th}} + \lambda t_{th} e^{-\lambda \tau'} - \left(\frac{-e^{-\lambda \tau'}}{1 - u_{min}} + e^{-u_{min} \lambda t_{th}} + \frac{u_{min}}{1 - u_{min}} e^{-\lambda(\tau' - t_{th} + u_{min} t_{th})} \right) \\ = & \quad e^{-\lambda t_{th}} + \lambda t_{th} e^{-\lambda \tau'} + \frac{u_{min}}{1 - u_{min}} \left(e^{-\lambda t_{th} u_{min}} - e^{-\lambda(\tau' - t_{th} + u_{min} t_{th})} \right) \\ = & \quad e^{-\lambda t_{th}} (1 + \lambda t_{th} e^{-\lambda(\tau' - t_{th})}) - \frac{u_{min} e^{-\lambda u_{min} t_{th}}}{1 - u_{min}} (1 - e^{-\lambda(\tau' - t_{th})}). \end{aligned}$$

The expression for $Q_{\tau}(t_{th})$ can be derived using exactly the same reasoning and it is omitted for the sake of space.

Recall that the source wants to maximize the delivery probability of the message to the destination and meet a given constraint on the energy expenditure, i.e., number of message copies released. In this case, the number of copies generated given k nodes in local interaction is $k\, n_s \Big(y(t_{th}) \, (e^{-\lambda(\tau' - t_{th}) u_{min}} - e^{-\lambda(\tau - t_{th}) u_{min}}) + (1 - e^{-\lambda(\tau' - t_{th}) u_{min}}) + (1 - e^{-\lambda t_{th}}) \Big)$ if $t_{th} \leq \tau'$ and $k\, n_s \Big(y(t_{th}) \, (e^{-\lambda \tau'} - e^{-\lambda t_{th}} + (1 - e^{-\lambda(\tau - t_{th}) u_{min}})) + (1 - e^{-\lambda \tau'}) \Big)$ if $t_{th} > \tau'$. From those expressions, the constraint writes

$$\Psi = X(t_{th})$$

where $X(t_{th})$ is the number of copies generated by n_s sources in a local interaction which is given by

$$X(t_{th}) = \begin{cases} n_s \mathbf{E}\{N y(t_{th})\} \Big(e^{-\lambda(\tau' - t_{th}) u_{min}} - e^{-\lambda(\tau - t_{th}) u_{min}} \Big) + n_s \mathbf{E}\{N\} \Big(1 - e^{-\lambda(\tau' - t_{th}) u_{min}} \Big) \\ \qquad + n_s \mathbf{E}\{N\} \Big(1 - e^{-\lambda t_{th}} \Big) & \text{if } t_{th} \leq \tau' \\ n_s \, \mathbf{E}\{N y(t_{th})\} \Big[e^{-\lambda \tau'} - e^{-\lambda t_{th}} + (1 - e^{-\lambda(\tau - t_{th}) u_{min}}) \Big] + n_s \, \mathbf{E}\{N\} \Big(1 - e^{-\lambda \tau'} \Big) & \text{if } t_{th} > \tau'. \end{cases} \quad (4.9)$$

Using the previous relation, it is possible to design an optimal threshold mechanism at the source node to find the switching time t_{th}.

Proposition 4.5.5. *Assume $0 < \tau' \leq \tau$ and let $0 \leq t_{th} \leq \tau$.*
If $u_{min} \geq \dfrac{\ln(1-\Psi/(n_s, \boldsymbol{E}\{N\}))}{\lambda\tau'}$ then there is no possible threshold policy. Otherwise there exist an optimal threshold policy as given in the following.

1. *Let $\tilde{t}_{th}$ solve for $e^{-\lambda(\tau'-t_{th})u_{min}} + e^{-\lambda t_{th}} = 2 - \Psi/(n_s \boldsymbol{E}\{N\})$. Let $t^*_{th} = min\{\tilde{t}_{th}, \tau\}$: given that $t_{th} \leq \tau'$ and $\sum_{N=1}^{\infty} P(K=N)(Q_{\tau'})^N \beta \leq \dfrac{g(\tau-\tau')}{n_s r}$, then t^*_{th} is the optimal threshold. If not go to step 2.*
2. *Let $\tilde{t}_{th}$ solve for $e^{-\lambda(\tau-t_{th})u_{min}} + e^{-\lambda t_{th}} = 2 - \Psi/n_s \cdot \boldsymbol{E}\{N\}$. Let $t^*_{th} = min\{\tilde{t}_{th}, \tau\}$: if $\sum_{N=1}^{\infty} P(K=N)\dfrac{Q^N_{\tau'}(1-(1-\beta)^N)}{N} \geq \dfrac{g(\tau-\tau')}{n_s r}$, then t^*_{th} is the optimal control. If not go to step 3.*
3. *Let t^*_{th} solve for $\Psi - X(t_{th}) = 0$, if exists, then t^*_{th} is the optimal threshold. Otherwise use $t_{th} = \tau$.*

Notice that in case 1., if $t_{th} > \tau'$ then u is constant and $u = 1$ until time τ'. Since action τ' dominates action τ, no user will use action τ thus, the game ends at τ' and there will be no switching.

Proof. A threshold policy that respects the constraint is such that the number of copies generated by n_s sources during the interval $[0, \tau]$, is not larger than Ψ. We hence determine the conditions such that, $X(t_{th}) \geq \Psi \ \forall t_{th} \Rightarrow \min_{t_{th}}(X(t_{th})) \geq \Psi$.
We have, $\min_{t_{th}}(X(t_{th})) = \min_{t_{th}}(n_s \mathbf{E}\{N\}(1-e^{-\lambda(\tau'-t_{th})u_{min}}) + n_s \mathbf{E}\{N\}(1-e^{-\lambda t_{th}}))$. Let

$$f(t_{th}) = n_s \mathbf{E}\{N\}(1-e^{-\lambda(\tau'-t_{th})u_{min}}) + n_s r \mathbf{E}\{N\}(1-e^{-\lambda t_{th}})$$

$$f'(t_{th}) = \lambda n_s \mathbf{E}\{N\}(1 - u_{min}e^{-\lambda\tau'}e^{\lambda t_{th}(1+u_{min})})$$
$$f'(t_{th}) \leq 0 \Rightarrow 1 - u_{min}e^{-\lambda\tau'}e^{\lambda t_{th}(1+u_{min})} \leq 0$$

It gives that,

$$t_{th} \geq \frac{-1}{\lambda(1+u_{min})}(\ln(u_{min}) - \lambda\tau' u_{min}) \tag{4.10}$$

From (4.10) we first deduce that X has at most one absolute maximum. Secondly, for $t_{th} \ll 1$, (4.10) is not verified and X is increasing. Otherwise X is decreasing and as a consequence has only one maximum at $t_{th} = \dfrac{-1}{\lambda(1+u_{min})}(\ln(u_{min}) - \lambda\tau' u_{min})$. It follows that $\min_{t_{th}}(X) = \min\{X(0), X(\tau')\} = \min\{n_s \mathbf{E}\{N\}(1-e^{-\lambda\tau' u_{min}}), n_s \mathbf{E}\{N\}(1-e^{-\lambda\tau'})\} = n_s r \mathbf{E}\{N\}(1-e^{-\lambda\tau' u_{min}})$. Therefore,

$$\min_{t_{th}}(X(t_{th})) \geq \Psi \Rightarrow n_s \mathbf{E}\{N\}(1-e^{-\lambda\tau' u_{min}}) \geq \Psi \Rightarrow u_{min} \geq \frac{\ln(1-\Psi/(n_s \mathbf{E}\{N\}))}{\lambda\tau'}.$$

1. and 2. follow respectively when the constraint Ψ is saturated in case that the strategy of choosing life-time τ' dominates the strategy of choosing τ and vice-versa. 3. follows when there exists an ESS in the interval $(0,1)$: See Theorem 4.3.5.1 for the conditions of existence of the ESS. In the case the constraint is not saturated, it means that number of copies generated by n_s sources during the interval $[0,\tau]$ is always less than Ψ. Thus the source will rather use $u = 1$ all the time and then switch with policy $t_{th} = \tau$. This completes the proof.

□

As obtained in the static policy case, if we consider the case of activation control with $0 = \tau' \leq \tau$, the same result on the monotonicity of the ESS holds.

Corollary 4.5.6. *If we consider the case of activation control (i.e $\tau' = 0$), then the ESS y^* is an increasing function of the threshold policy t_{th}.*

The proof of this corollary comes from the fact that $Q_\tau(t_{th})$ is decreasing in t_{th} and we can use the same reasoning as for corollary 4.5.3.

4.6 Learning of Optimal threshold strategy

One issue that makes the control of DTNs an interesting technical challenge is that collecting network parameters, e.g.,, number of nodes and intermeeting intensities, is per se a difficult task due to lack of persistent connectivity. We prove in the next Section that it is possible to design distributed online protocols that attain the optimal control described in Proposition 4.5.2 and in Proposition 4.5.5 at the source node in a blind fashion. In practice this means that the source node can perform the optimal control of forwarding – as devised before through mechanism design – by tracking the number of infected relays only. Observe that for the two hop routing this is indeed a quantity available at the source node. For the sake of clearness, we restrict our analysis to the case of $\tau' = 0$.

4.6.1 Evolutionary Game Dynamics

In order to study the behavior of a population of relays, it is convenient to represent this game as a dynamical system. The evolutionary system is described by differential equations that govern the rate of change of subpopulations playing a particular strategy. As mentioned above, it is assumed here that individuals play only pure strategies, mixed strategies result when fractions of the population play different pure strategies. Evolutionary game dynamics give a tool for observing the evolution of strategies in the population in time. The most famous one is the replicator dynamics, based on replication by imitation, which we consider in this subsection for observing the evolution of the activation rate of the population of relays.

Theorem 4.6.1.1. *The ESS given in Thm. 4.3.5.1 is an asymptotically stable equilibrium of the replicator dynamics for every non trivial state y_0*

Proof. The replicator dynamic when the source uses policy u, is given by

$$\begin{aligned} \frac{dy(t,u)}{dt} &= y(t,u)(U^u_{av}(\tau,y,u) - F^u(y,y,u)) && (4.11)\\ &= y(t,u)(U^u_{av}(\tau,y(t)) - y(t)U^u_{av}(\tau,y(t,u))) && (4.12)\\ &= y(t,u)(1-y(t,u))H^u(y(t,u)) && (4.13) \end{aligned}$$

We can consider the Lyapunov function $V(y) = (y-y_0)^2$, and non trivial ESS $0 < y_0 < 1$, then

$$\dot{V}(y) = \frac{d}{dy}V(y)\,\dot{y} = 2(y-y_0)\Big(y(1-y)H^u(y)\Big) \quad (4.14)$$

By definition, $y_0 = H^{-1}(0)$, so that $H(y_0) = 0$. Also, recall that function $H^u(y)$ is decreasing on $(0,1)$ (see Corollary 4.5.3). For $y < y_0$, $H^u(y) > 0$ so that $\dot{V}(y) < 0$, and for $y > y_0$, $H^u(y) < 0$ so that $\dot{V}(y) < 0$. Indeed, $\dot{V} \leq 0$ and $\dot{V}(y) = 0$ iff $y = y_0$, so that the ESS is asymptotically stable according to the Lyapunov stability method. □

This result makes it possible to formulate some qualitative statements about the ESS by analyzing the corresponding replicator dynamics. Hence the replicator dynamic can be used by sources in order to observe the evolution of strategies in the population. In the next subsection, we propose a new online algorithm that allows the sources to converge to the optimal solution.

4.6.2 Stochastic algorithm for adaptive optimization

In this subsection we propose an on-line algorithm to attain optimal control of forwarding for both the static case and the dynamic case. It is designed based on stochastic approximation theory. This algorithm is *blind*: it does not require a priori knowledge of network parameters (inter-meeting intensities, density of relays). Let us denote $\alpha = \int_0^\tau u(t)\,dt$ determined from policy u. For the static policy, $u(t) = u = \alpha/\tau$, while for the threshold policy it holds $t_{th} = \alpha$. The algorithm works in rounds and each round of the algorithm corresponds to the delivery of a set of messages. During a round, the source adopts a certain control policy.

Let Ψ/n_s be the maximum number of relay mobiles that the source can infect in a local interaction. For the sake of notation, we let $\hat{X}^s(\alpha(n), y(n))$ the number of copies that are potentially delivered by the source s using policy $\alpha^s(n)$ and the profile of relay population is $y(n)$ (the fraction of relay mobiles using full activation action τ) in the current round. This quantity can be estimated by averaging over several consecutive slots (a slot corresponds to the delivery of a message to destination). Although it is obvious that the source node knows the exact number of relay nodes met at every round, by simply keeping track of the successive contacts, in our simulation it is necessary to approximate the randomness of the contact and infection events. We maintain at each

round the vector $\tilde{X}_n$ that records the average number of nodes that are potentially infected at any time instant up to time τ. Using interpolation, the source node is able to obtain an estimate the average number of copies $\hat{X}(\alpha(n), y(n))$ that could potentially be infected by the source using policy $\alpha^s(n)$.

Using $\bar{X}_n^s = \hat{X}^s(\alpha^s(n), y(n))$, we update $\alpha^s(n)$ according to the following relation:

$$\alpha^s(n+1) = \Pi_{[0,\tau]}\Big(\alpha^s(n) + a_n(\frac{\Psi}{n_s} - \bar{X}_n^s)\Big).$$

where $\Pi_{[0,\tau]}$ is the projection of the values of $\alpha(n)$ over the set $[0,\tau]$ at each iteration. The algorithm is described in Algorithm 4.6.2.

In the following theorem we resume the convergence properties of our algorithm.

Theorem 4.6.2.1. *If the sequence $\{a_n\}$ is chosen such that (a_n) verifies: $a_n > 0, \forall n \sum_{n=0}^{+\infty} a_n = +\infty$ and $\sum_{n=0}^{+\infty} a_n^2 < +\infty$, then the sequence $(\alpha(n))$ converges to the optimal threshold $t_s^*(y^*)$ where y^* is the ESS.*

Proof. First we show that the sequence $\alpha(n)$ converges to some limit set of the following Ordinary Differential Equation (ODE)

$$\dot{\alpha}^s(t) = G(\alpha^s(t)) + z(t) = \Psi - E[X^s|\ \alpha^s(t), y^*(\alpha^s(t))] + z(t), \quad z \in -C(\alpha^s) \tag{4.15}$$

where $y^*(\alpha)$ is the solution of the replicator dynamic when the source uses policy α, i.e.,

$$\dot{y}(t, \alpha^s) = y(t)(1 - y(t))\tilde{H}^{\alpha}(y(t, \alpha^s)) \tag{4.16}$$

and z is the minimum force needed to keep the solution α^s in $I = [0,\tau]$. If α^s is in I on some time interval, then $z(.)$ is zero on that interval ($C(\alpha^s)$ contains only the zero element). If α^s is on the interior of a boundary of I (i.e., α^s equals either 0 or τ) and $G(\alpha^s)$ points out of I, then $z(.)$ points backward inside I, i.e. $C(\alpha)$ is the infinite convex cone generated by the outer normals at α^s of the faces on which α^s lies. For example, let $\alpha^s = \tau$, with $G(\alpha^s) > 0$, then, $z(t) = -G(\alpha^s)$. Note that the solution $y^*(\alpha^s)$ is global asymptotically stable equilibrium of the ODE (4.16). As discussed in [66], the convergence of such stochastic algorithm is guaranteed when the sequence (a_n) verify, $a_n > 0, \forall n$, $\sum_{n=0}^{+\infty} a_n = +\infty$ and $\sum_{n=0}^{+\infty} a_n^2 < +\infty$.

We now need to show that $(\alpha^*, y^*(\alpha^*))$ is globally asymptotically stable solution of the system (4.15)-(4.16). From theorem 4.6.1.1, proposition 4.5.2 for the static policy and proposition 4.5.5 for dynamic policy, it is easy to see that $(\alpha^*, y^*(\alpha^*))$ solves (4.15)-(4.16). In order to show the stability of the ODE (4.15), we need to show that the function $E[X^s|\ \alpha^s, y^*(\alpha^s)]$ is increasing function in α on $[0,\tau]$. From equations (4.9) and (4.5), we have

$$E[X^s|\ \alpha^s, y^*(\alpha^s)] = \begin{cases} \mathbf{E}\{Ny(\alpha^s)\}(1-e^{-\lambda\alpha^s}) & \text{for the static policy} \\ \mathbf{E}\{Ny(\alpha^s)\}\left[2-(e^{-\lambda\alpha^s}+e^{-\lambda(\tau-\alpha^s)u_{min}})\right] & \text{for the dynamic policy} \end{cases} \tag{4.17}$$

From Corollary 4.5.3 and Corollary 4.5.6, for either static or dynamic case, the function $y(\alpha^s)$ is increasing function. It follows from (4.17), that $E[X^s|\ \alpha^s, y^*(\alpha^s)]$ is increasing function. Then the derivative of the function $\frac{\Psi}{n_s} - E[X|\ \alpha^s, y^*(\alpha^s)]$ at the optimal solution α^* is negative. Hence the optimal solution is asymptotically stable. □

Algorithm: ADAPTIVE OPTIMIZATION

Initialize the values of $t_{th}(0)$ and X_0.
At each round n

1. Update t_{th} with $t_{th}(n+1) = \Pi_{[0,\tau]}\Big(t_{th}(n) + a_n(\Psi - \hat{X}(t_{th}(n)))\Big)$
2. Find the corresponding value of $\hat{X}(t_{th}(n+1))$:
   - Generate randomly the number k of users in a local interaction
   - Use the replicator dynamic[a] to find the profile $y(t_{th}(n+1))$ of the population.
   - Estimate X_{n+1} and maintain $\tilde{X}_{n+1} = \dfrac{n\cdot\tilde{X}_n + X_{n+1}}{n+1}$
3. Estimate $\hat{X}(t_{th}(n+1)) = \text{interp}(\tilde{X}_n, t_{th}(n))$
4. If $|t_{th}(n+1) - t_{th}(n)| < \epsilon$ stop, else go to 1.

End Algorithm

[a] source nodes in local interactions uses the replicator dynamic to determine the population profile. Note that the replicator dynamic converges exponentially, so that the discovery of the population profile can be achieved over a limited number of slots.

4.7 Numerical analysis

We extend our analysis with numerical experiments. In particular, the population of competing nodes is assumed to be Poisson distributed, with mean γ. This means that the number of mobiles interfering in each local interaction is a Poisson random variable, as discussed in previous sections. The inter-meeting rate λ between a pair of relay nodes follows a Random Way-point mobility process. In lemma 4. of [96] authors derived the following estimation of the pair-wise meeting rate :

$$\lambda = \frac{2\omega r \mathbf{E}\{V^*\}}{L^2}$$

where, $\omega = 1.3683$ is a constant specific to the mobility model, here the random way point mobility, r is the communication range radii and L is the playground size. In our simulations, we have considered a playground of size, $L = 1000m$, $r = 50m$

and $\mathbf{E}\{V^*\} = 2.5m/s$ when, $[v_{min}, v_{max}] = [4, 10]Km/h$. Other parameters are set in Tab. 4.1.

-	λ	τ	τ'	u_{min}	γ
Settings ($g = 0.000015$)	0.0004	1000	0	0.01	100

Table 4.1: *Parameters values*

We depict in Fig. 5.3 the different values of the number of nodes that would be infected depending on the threshold policy t_{th}. We simulate the forwarding of message copies and track the average number of them released within live-time τ, according to expression (4.9). We observe that the set of parameters chosen for the numerical experiments indeed allows for the existence of an optimal threshold policy within the interval $(0, \tau)$, i.e., the forwarding control can saturate the energy constraint by infecting the maximum allowed number of infected nodes Ψ (see Prop. 4.5.5). In the same figure we can also observe the target optimal threshold policy, as computed using a centralized approach, i.e., under the assumption that every parameter of the network is known.
However, in the rest of our simulations we focus on the performance of our learning algorithm, which is based on a decentralized approach.

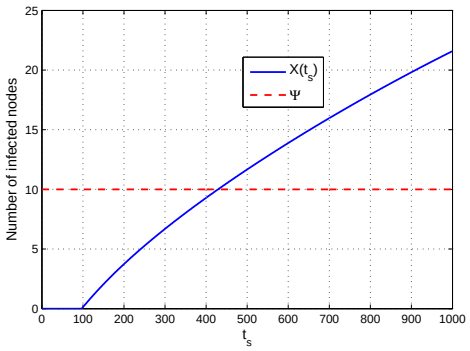


Figure 4.2: *Offline expression of the number of infected nodes in function of the threshold control t_{th}*

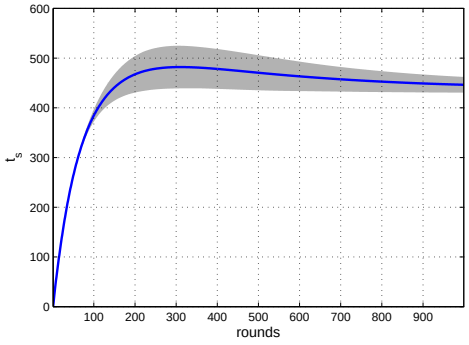


Figure 4.3: *Learning algorithm convergence to the optimal threshold policy for several rounds*

In Fig. 5.4, we plot the trajectory of sequence $(t_{th}(n))$ and a confidence interval of 90% with 1000 runs each consisting of several rounds of our stochastic adaptive algorithm, meant to measure convergence to the optimal threshold policy $t_{th}*$. Fig. 4.4 shows several runs of our learning algorithm for the sequence $(t_{th}(n))$ without averaging. In Fig. 4.5, we plot the evolution of the number of infected nodes for several rounds of our learning algorithm. This can be seen for each run, as the cost spent during learning, in terms of constrain overflow.

In the following, we evaluate the behavior of the learning algorithm once the learning process converged at the equilibrium (ESS). Our objective here is to observe the

influence on the ESS and on the probability of success of parameters such as partial live-time activation τ' and infection constraint Ψ. In figure 4.6 we plot the ESS y^* and probability of success P_s as a function of τ': by increasing τ' the difference $(\tau - \tau')$ decreases, untill it no longer satisfies existence condition of a threshold policy (see Prop. 4.5.5). It can be noticed that the probability of success is maximized for a value of τ' that approaches τ but this value is just above the value of the probability of success when τ' is null: in that regime, a large proportion of the population is fully active. As τ' approaches τ there are less and less relays nodes that select to be fully active, so that the population profile of fully active nodes decrease with τ'. A joint analysis of both plots from figure 4.6 shows that at the value of τ' *s.t.* $y^*(\tau') = 0$, the plot of the probability of success changes suddenly the sens of its monotonicity. This is due to the fact that the probability of success decreases when y^* decreases and increases in case y^* is constant.

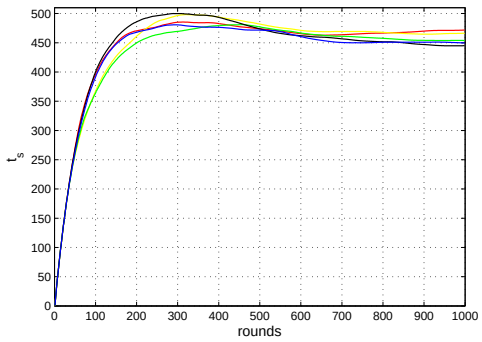


Figure 4.4: *Several runs* (5) *of the learning process to find the optimal threshold*

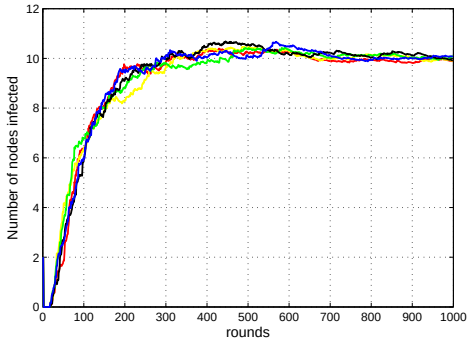


Figure 4.5: *Evolution of the number of infected nodes during several* (5) *runs of the learning process*

Fig. 4.7 shows the behavior of the probability of success P_s and ESS y^* depending on Ψ. As expected, the less the system is constrained, i.e., the larger Ψ, the larger the number of nodes that take part to live-time τ strategy to relay the message, and larger the success probability. However, saturation appears in Ψ: this confirms the intuition that the probability of success is not constantly increasing with the constrain Ψ, but attains a maximum that depends on the set of physical parameters, such as τ, λ and γ.

For small values of Ψ, on the other hand, the condition of existence of a threshold policy is not attained. In this case, the source not will either not transmit or will be transmitting with probability u_{min} all the time as the constraint is too tight and cannot be fulfilled. Using the settings adopted before, this happens for $\Psi < 1$. Finally we depict in figure 4.8 the evolution of the probability of success as a function of the inter-meeting rate λ. We observe that as λ increases the probability of success approaches unity. Indeed the higher the inter-meeting rate the larger the chance for a message to reach its destination. The small variations observed around $\lambda = 0.01$ are due to simulations artifact when λ takes too big values.

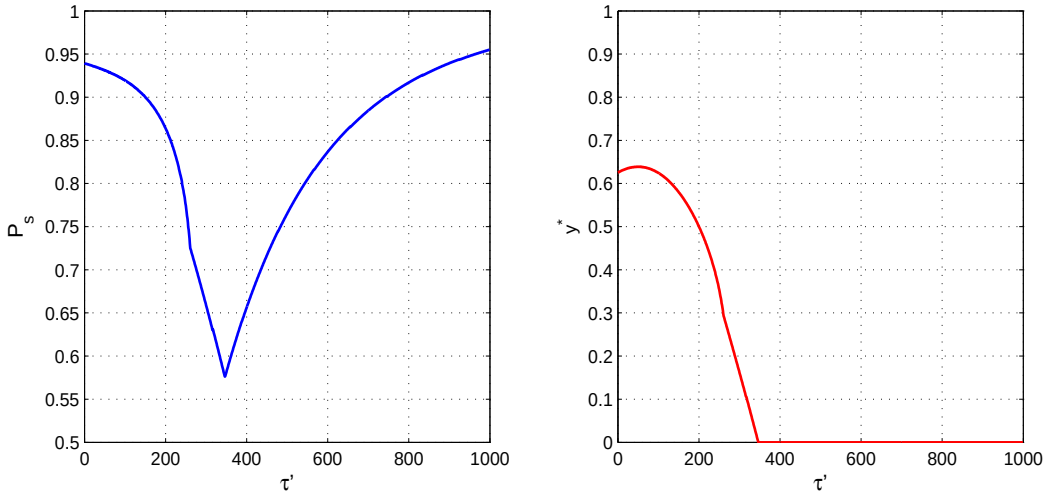


Figure 4.6: *Equilibrium analysis : Evolution of the ESS y^* and the probability of success P_s with τ'*

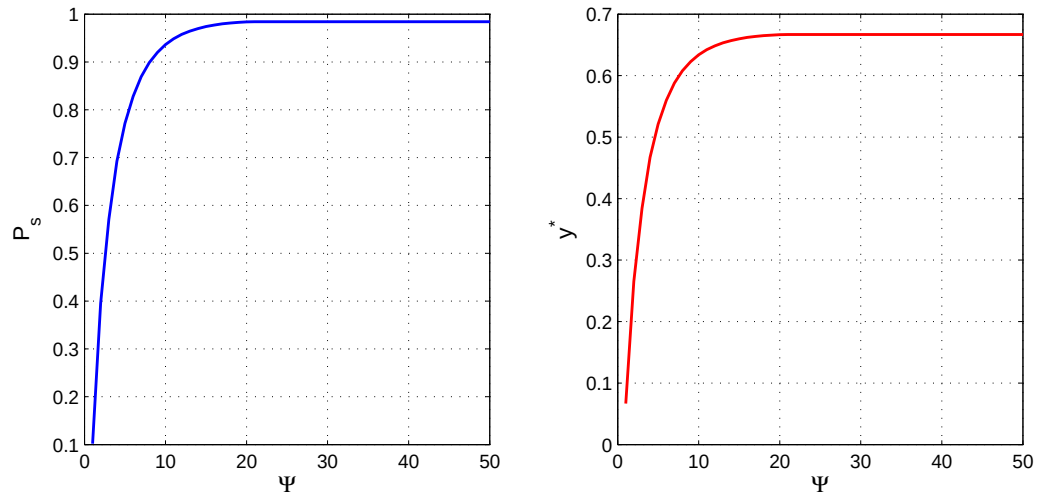


Figure 4.7: *Equilibrium analysis : Evolution of the ESS y^* and the probability of success P_s with Ψ*

4.8 Conclusions

This work introduces a novel general framework for competitive forwarding in DTNs ruled by two hop routing. Within the context of routing games, part of the forwarding control is demanded to relays. Relays, in fact, may accept to spend some energy to participate to the forwarding process. However, since forwarding has some cost, relays can trade energy expenditure for some reward upon delivery to the destination. This type of incentive mechanism may prove crucial in order to incentive owners of mobiles

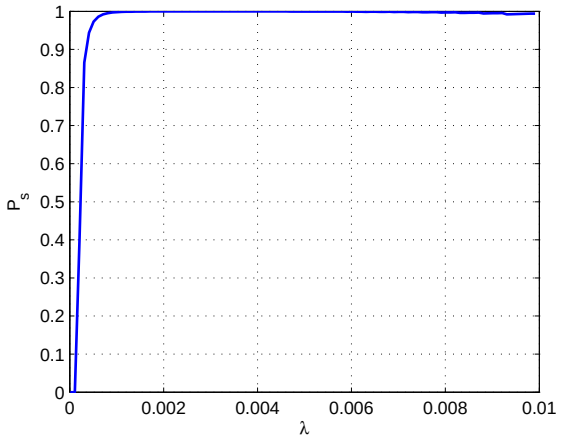


Figure 4.8: *Evolution of the probability of success in function of* λ

to sacrifice part of the battery charge of their devices in order to take part to the relay process. Eventually, if several relays do compete for reward, the delivery probability is increased. In turn, the challenging technical problem is how to design routing in this context, i.e., exploit efficiently the incentive mechanism by means of the forwarding control operated by the source. In this chapter, we have proposed a complete model for such incentive mechanisms accounting for several aspects of DTN forwarding, including energy expenditure at relays and number of released message copies. We have been using competitive games: we assumed that the activation of a relay during a local interaction depends solely on the expected reward in delivering with success the message to the destination and on the energy expenditure to perform forwarding operations. Evolutionary game theory was employed to elaborate necessary and sufficient conditions for the existence of evolutionary stable strategies, depending on energy and delivery probability only. Compared to existing works in literature, this is a novel context for routing control in DTNs because the forwarding control at the source depends on the whole evolutionary dynamics and ultimately on the unique ESS that is reachable given the forwarding control chosen at the source. This has been developed precisely as a mechanism design problem, both in the case of optimal static control policies and dynamic ones. Finally, we demonstrated that such a mechanism can be implemented in a decentralized fashion, with no need to estimate system parameters at runtime, e.g., the number of nodes or inter-meeting intensity. In fact, a stochastic approximation algorithm can attain the optimal forwarding policy. However, with respect to the last result, we had to restrict to activation control case ($\tau' = 0$). With this choice, it was possible to ensure the asymptotic stability of the equilibrium of our learning process. In future work, however, we will investigate the general case of of live-time control ($\tau' \geq 0$), and the conditions for stability of the learning process.

Chapter 5

Markov Decision Evolutionary Game for Energy Management in Delay Tolerant Networks

5.1 Introduction

Delay tolerant networks (DTNs) emerged recently as a novel communication paradigm. Throughout this chapter, we focus on a specific class of DTNs where persistent connectivity cannot be guaranteed due to limited coverage and high mobility [122]. For such networks, forwarding strategies have been designed purposely to solve the problem of intermittent connectivity: a message is delivered to the intended destination, leveraging the motion of a subset of nodes, i.e., relays, which carry copies of the message stored in their local memory. This is the so called *carry-store-and-forward* routing. The DTN paradigm has been validated by several experimental deployments [56, 31].

In order to reach the destination, a straightforward strategy is to disseminate multiple copies of the message in the network. This approach is known as epidemic forwarding [120], in analogy to spread of infectious diseases. The aim in doing so is to let some of such message copies reach the destination with high probability within some target deadline [108, 36]. We confine our analysis to the two hop routing protocol. In fact, it has two major advantages: first, compared to epidemic routing, it performs natively a better trade-off between the number of released copies and the delivery probability [36]. Second, forwarding control can be implemented on board of the source node. Under two hop routing, the source transmits a message copy to mobiles devices it encounters. Relays, conversely, forward the message to the destination only.

In this context, the higher the number of relays joining the forwarding process, the higher the success probability. However, battery lifetime of mobile devices may deplete due to continuous beaconing operations [10], which may be a critical factor discouraging the usage of mobile devices as relays for DTN-based applications. A solution is to design reward-based forwarding mechanisms where the probability of forwarding

becomes function of the competition within a population of mobiles: a relay receives a unit of reward if it is the first to deliver the message to the destination. For each message generated by the source, a relay may choose two different actions that affect message relaying: full activation or partial activation, i.e., being active for a shorter time period and then go back to low power mode, thus saving batteries.

5.1.1 Main contributions

In this chapter we extend a similar framework studied by El-Azouzi et al. [37]. The novelty here is that the strategy of a mobile relay determines not only the immediate reward but also the transition probability to its next battery energy state. The problem is formulated as a Markov Decision Evolutionary Game (MDEG), where each relay wishes to maximize the expected utility. We characterize the Evolutionary Stable Strategies (ESS) for these games and show a method to compute them. Once determined the possible equilibria for the game, the optimal forwarding control at the source node that maximizes the forwarding probability has been derived. We show that the success probability is not always increasing with the number of message copies, and may well *decrease* under some conditions, which is adding an intriguing novel facet to the control of forwarding in DTNs.

5.2 Basic notions on evolutionary games

We consider the standard setting of evolutionary games :

- There is one population of users. The number of users in the population is large.
- We assume that there are finitely many pure strategies or actions. Each member of the population chooses from the same set of strategies $\mathcal{A} = \{1, 2, \ldots, I\}$.
- Let $M := \{(y_1, \ldots, y_I) \mid y_j \geq 0, \sum_{j=1}^{I} y_j = 1\}$ be the set of probability distributions over the I pure actions. I can be interpreted as the set of mixed strategies. It is also interpreted as the set of distributions of strategies among the population, where y_j represents the proportion of users choosing the strategy j.
- The number of users interfering with a given randomly selected user is a random variable K in the set $\{0, 1, \ldots\}$.
- A player does not know how many players would interact with it.
- The payoff function of all players depends of the player's own behavior and the behavior of the other players. The expected payoff of a user playing strategy j when the state of the population is y, is given by $U_{av}(j, y) = \sum_{k \geq 0} \mathbb{P}(K = k) U(j, k, y)$, where $U(j, k, y)$ is the payoff of a user playing strategy j when the state of the population is y and given that the number of users interfering with a

given randomly selected user is k. Hence the average payoff of a population in state y is given by $F(y,y) = \sum_{i=1}^{I} y_j U_{av}(j,y)$.

- The game is played many times and there are many local interactions at the same time

This framework defines the necessary tools for application of the concept of ESS defined in 4.3.4 of chapter 4.

5.3 Model

Consider a standard description of a DTN framework with several sources and destinations under two-hops routing(see Chapter 4: section 4.3). We use evolutionary games to study the competition between individual mobiles in a routing game in DTN based on non-cooperative live-time selection.

Specifically, we assume that each relay node can choose two different live times for the message: τ, i.e. full activation and τ' i.e., partial activation, where $\tau' < \tau$. At each slot, a mobile has to take a decision to be fully active or partially active, based on his battery energy sate. To simplify, we assume that the state can take three values : $\{F, A, E\}$ for Full, Almost empty or Empty. At state F only action τ is available, and at E participation on forwarding message is not possible any more. In state A, each mobile chooses between the two actions τ (full activation) or τ' (partial activation). The life time of mobile is defined as the number of slots during which its battery is nonempty.
Local Interaction : Without loss of generality, we assume that the network is composed of several local interactions. In each local interaction, there is a source-destination pair in which the source has packet generated at each time $t_{i+1} - t_i = n\tau$ where $i = 1, 2..$ and $t_0 = 0$. Let N be the number of mobiles (possibly random) in an area which is assumed fixed during time slot. We denote by y (resp. $1-y$) the fraction of mobiles sharing the strategy S_τ (resp. $S_{\tau'}$), playing action τ(resp. τ'), in state A.

Some notation

Consider an active mobile in a local interaction with source s_i, destination d_i and N opponents. We introduce the following notations:

- $P_i(a)$ is the probability of remaining at energy level i when using action a. Since at state F only action τ is available, we write P_F instead of $P_F(\tau)$.
- Let $M := \{(y, 1-y)\}$ be the set of probability distributions over the 2 pure actions τ and τ' in state A. M can be interpreted as the set of mixed strategies.

We denote by α the proportion of mobiles that uses the action τ in the population whatever their state. This proportion is exactly equal to the fraction of time that a

mobile uses action τ during his life. Then we have the following relation between α and y:

$$\alpha(y) = \frac{T(F)}{T_\tau(A) + T(F)} + y\frac{T_\tau(A)}{T_\tau(A) + T(F)}, \tag{5.1}$$

where $T_\tau(A) = \dfrac{1}{1 - P_A(\tau)}$ (resp. $T(F) = \dfrac{1}{1 - P_F}$) is the expected time a mobile spends at state A (resp. F).

The probability that the tagged mobile relays the copy of the packet to the destination within live time τ is given by $1 - Q_\tau$ where Q_τ is given by $Q_\tau = (1 + \lambda\tau)e^{-\lambda\tau}$ and the probability that it relays the copy of the message if it chooses live time τ' is given by $1 - Q_{\tau'}$ where $Q_{\tau'} = (1 + \lambda\tau')e^{-\lambda\tau'}$. Let $P_{succ}(\tau, N, \alpha(y))$ (resp. $P_{succ}(\tau', N, \alpha(y))$) be the probability that the tagged mobile receives the unit reward, if it chooses live time τ (resp. τ'). Now this success probability is expressed by:

$$P_{succ}(\tau', N, \alpha(y)) = (1 - Q_{\tau'})\sum_{k=1}^{N} C_{k-1}^{N-1} \frac{(1 - Q_{\tau'})^{k-1}(1 - (1 - Q_{\tau'}))^{N-k}}{k} = \frac{1 - (Q_{\tau'})^N}{N}.$$

The gain obtained by a mobile using live time τ' is given by

$$U(\tau', y) = \sum_{N=1}^{\infty} P(K = N) P_{succ}(\tau', N, \alpha(y))$$

Now the probability that a mobile receives the unit reward, if it chooses live time τ, is given by

$$\begin{aligned} P_{succ}(\tau, N, \alpha(y)) &= P_{succ}(\tau', N, \alpha(y)) + (Q_{\tau'})^N \beta \sum_{k=1}^{N} C_{k-1}^{N-1} \frac{\beta^{k-1}\alpha(y)^{k-1}(1 - y\beta)^{N-k}}{k} \\ &= P_{succ}(\tau', N, \alpha(y)) + (Q_{\tau'})^N \frac{1 - (1 - \beta\alpha(y))^N}{N\alpha(y)} \end{aligned}.$$

where $\beta = 1 - \dfrac{Q_\tau}{Q_{\tau'}}$. The utility for a mobile using live time τ is thus,

$$U(\tau, \alpha(y)) = \sum_{N=1}^{\infty} P(K = N) P_{succ}(\tau, N, \alpha(y)).$$

A general policy u is a sequence $u = (u_1, u_2, ...)$ where u_i is the strategy used at time t_i if the state is A. We shall use a pure stationary policy in which there exist two pure stationary policies; the one that always choose τ and the one that always choose τ'.

Fitness

Let assume that an active mobile during $[t_i, t_{i+1}]$, will receive a unit of reward r if it is the first to deliver a copy of the packet to the destination. Assume that y is fixed

and does not change in time (Note that assuming that y is fixed in time does not mean that the actions of each player are fixed in time. It only reflects a situation in which the system attains a stationary regime due to the averaging over a very large population, and the fact that all mobiles choose an action in a given individual state using the same probability low). Then the expected optimal fitness of an individual starting at a given initial state can be computed using the standard theory of total-cost dynamic programming, that states in particular that there exists an optimal stationary policy (i.e. a policy for which at any time that the individual is at state A, the action a_i of choosing $T \in \{\tau, \tau'\}$ is the same). We shall therefore *restrict to stationary policies* unless stated otherwise.

Let $V_\tau(i, \alpha(y))$ (respectively $V_{\tau'}(i, \alpha(y))$) be the total expected fitness of a user given that it is in state i, that it uses action τ(respectively τ') and given the population profile $\alpha(y)$.

We proceed by computing the individual's expected total utility and remaining lifetime that correspond to a given initial state and a stationary policy. We have $V_\tau(A, \alpha(y)) = U(\tau, \alpha(y)) + P_A(\tau)V_\tau(A, \alpha(y))$ which gives that

$$V_\tau(A, \alpha(y)) = \frac{U(\tau, \alpha(y))}{1 - P_A(\tau)}.$$

The total expected utility for a mobile starting from state F and using strategy S_τ, is given by

$$\begin{aligned} V_\tau(F, \alpha(y)) &= U(\tau, \alpha(y)) + P_F V_\tau(F, \alpha(y)) + (1 - P_F)V_\tau(A, \alpha(y)), \\ &= U(\tau, \alpha(y))(\frac{1}{1 - P_F} + \frac{1}{1 - P_A(\tau)}). \end{aligned}$$

Similarly, the total expected utility for a mobile starting from state F and using strategy $S_{\tau'}$, is given by

$$V_{\tau'}(F, \alpha(y)) = \frac{U(\tau, \alpha(y))}{1 - P_F} + \frac{U(\tau', \alpha(y))}{1 - P_A(\tau')}.$$

Our game has a more complex structure than a standard evolutionary game. In particular, the fitness that is maximized is not the outcome of a single interaction but of the sum of fitness obtained during all the opportunities in the mobile's lifetime. Let H be the function defined as

$$\begin{aligned} H : y \in (0,1) \rightarrow & \quad V_\tau(F, \alpha(y)) - V_{\tau'}(F, \alpha(y)) \\ = & \quad \frac{U(\tau, \alpha(y))}{1 - P_A(\tau)} - \frac{U(\tau', \alpha(y))}{1 - P_A(\tau')}) \\ = & \quad \frac{U(\tau, \alpha(y))(1 - P_A(\tau')) - U(\tau', \alpha(y))(1 - P_A(\tau))}{(1 - P_A(\tau'))(1 - P_A(\tau))} \\ = & \quad \frac{\tilde{H}(y)}{(1 - P_A(\tau'))(1 - P_A(\tau))}, \end{aligned}$$

with $\tilde{H}(y) := U(\tau, \alpha(y))(1 - P_A(\tau')) - U(\tau', \alpha(y))(1 - P_A(\tau))$. Thus this function can be written as:

$$H(y) = \frac{\sum_{N=1}^{\infty} P(K = N)\Big[(Q_{\tau'})^N \frac{1-(1-\beta\alpha(y))^N}{N\alpha(y)}(1 - P_A(\tau')) - \frac{1-(Q_{\tau'})^N}{N}(P_A(\tau') - P_A(\tau))\Big]}{(1 - P_A(\tau'))(1 - P_A(\tau))}.$$

Existence and Uniqueness of ESS

In this section, we are now looking at the existence and uniqueness of the ESS
Proposition 5.3.1.

(1) The strategy S_τ dominates the strategy $S_{\tau'}$ if and only if

$$P_A(\tau') - P_A(\tau) \leq \sum_{N=1}^{\infty} P(K = N)\Big[Q_\tau'^N\Big((1-\beta)^N(1 - P_A(\tau'))\Big) + 1 - P_A(\tau)\Big] \triangleq A_1$$

(2) The strategy $S_{\tau'}$ dominates the strategy S_τ if and only if

$$P_A(\tau') - P_A(\tau) \geq \sum_{N=1}^{\infty} P(K = N)Q_\tau^N\Big(N\beta(1 - P_A(\tau')) + P_A(\tau') - P_A(\tau')\Big) \triangleq A_0$$

(3) If $P_A(\tau') - P_A(\tau) > A_1$ and $P_A(\tau') - P_A(\tau) < A_0$, then there exists an unique ESS y^* which is given by $y^* = \tilde{H}^{-1}(0)$

Proof. (1) The strategy S_τ dominates the strategy $S_{\tau'}$ if and only if $V_\tau(F, \alpha(y)) \geq V_{\tau'}(F, \alpha(y))$ for all $y \in [0, 1]$. Since $\tilde{H}$ is a decreasing function and $\tilde{H}(1) = A_1 - (P_A(\tau') - P_A(\tau)) \geq 0$, thus $\tilde{H}(y) \geq 0$ for all $y \in (0, 1)$. Then $V_\tau(F, \alpha(y)) - V_{\tau'}(F, \alpha(y)) = H(y) = \frac{\tilde{H}(y)}{(1 - P_A(\tau'))(1 - P_A(\tau))} \geq 0$ for all $y_T \in [0, 1]$. This completes the proof for (1)

(2) The strategy $S_{\tau'}$ dominates the strategy S_τ if and only if $V_{\tau'}(F, y) \geq V_\tau(F, y)$ for all $y \in [0, 1]$. Since the function $\tilde{H}$ a is decreasing function and $\tilde{H}(0) = A_0 - (P_A(\tau') - P_A(\tau)) \leq 0$, thus $\tilde{H}(y) \leq 0$ for all $y \in (0, 1)$. Then $V_{\tau'}(F, y) - V_\tau(F, y) = \tilde{H}(y) = \frac{\tilde{H}(y)}{(1 - P_A(\tau'))(1 - P_A(\tau))} \leq 0$ for all $\beta \in [0, (1 - Q_\tau)]$. This completes the proof for (2)

(3) A strictly mixed equilibrium y^* is characterized $V_\tau(F, \alpha(y^*)) = V_{\tau'}(F, \alpha(y^*))$. The function $\tilde{H}$ is continuous and strictly decreasing monotone on $(0, 1)$ with $\tilde{H}(0) > 0$ and $\tilde{H}(1) < 0$. Then the equation $\tilde{H}(y) = 0$ has a unique solution in the interval $(0, 1)$. This completes the proof.

□

5.3.2 Poisson distribution

We consider that nodes are distributed over a plan following a Poisson distribution with density γ. The probability that there is N nodes in local interaction is given by the

following distribution : $\mathbb{P}(K=k) = \frac{\gamma^{k-1}}{(k-1)!}e^{-\gamma}$, $k \geq 1$. Using a Poisson distribution of the nodes and from previous theorems, the unique ESS y^* is unique solution of the following equation:

$$\frac{e^{\gamma Q_{\tau'}} - e^{Q_{\tau'}(1-\beta\alpha(y^*))\gamma}}{\alpha(y^*)} = (e^{\gamma} - e^{Q_{\tau'}\gamma})\frac{P_A(\tau') - P_A(\tau)}{1 - P_A(\tau)}$$

Thus, the equilibrium is given by

$$\alpha(y^*) = \frac{LambertW(-\frac{\rho\beta e^{-\frac{\rho(\beta e^{\rho}-c)}{c}}}{c})c + \rho\beta e^{\rho}}{c\rho\beta}, \tag{5.2}$$

where $\rho = Q_{\tau'}\gamma$ and $c = (e^{\gamma} - e^{Q_{\tau'}\gamma})\frac{P_A(\tau') - P_A(\tau)}{1 - P_A(\tau)}$

Dirac distribution

We consider that at a given time there is a fixed number of nodes in a local interaction. In this part, we suppose that the population of nodes is composed with many local interaction between N nodes where $N > 2$. Using the Dirac distribution, the unique ESS y^* of this game is the unique solution of the following equation:

$$\frac{1 - (1 - \beta\alpha(y^*))^N}{\alpha(y^*)} = \frac{1 - (Q_{\tau'})^N}{(Q_{\tau'})^N}. \tag{5.3}$$

Since (5.3) corresponds to a polynome of order N, we can only have an explicit expression for $N \leq 5$. Therefore, we restrict to show some properties of the stable equilibrium by numerical computations in the next section.

For example:

$$N = 2 \implies \alpha(y^*) = \frac{(1 - (Q_{\tau'})^2)G - 2Q_{\tau'}(Q_{\tau'} - Q_{\tau})}{(Q_{\tau'} - Q_{\tau})^2}$$

with $G = \frac{P_A(\tau') - P_A(\tau)}{1 - P_A(\tau)}$. One can easily show that $\alpha(y^*) > 1$ thus $y^* = 1$.

For $N = 3$, $\alpha(y^*)$ is the solution of the equation

$$y^2 + \frac{3}{\beta}y - k = 0$$

with $k = \frac{3}{\beta^2} - \frac{1 - (Q_{\tau'})^3}{\beta^3(Q_{\tau'})^3}G$. $\delta = \frac{9}{\beta^2} + 4k$ with

$$k = \frac{3(Q_{\tau'})^2[Q_{\tau'} - Q_{\tau}]^2 + (Q_{\tau'})^4 - [Q_{\tau'} - Q_{\tau}]}{\beta^2(Q_{\tau'})^2[Q_{\tau'} - Q_{\tau}]^2} > 0$$

The realizable solution is $\alpha(y^*) = \dfrac{-\frac{3}{\beta} + \sqrt{\delta}}{2}$. In this case the solution $\alpha(y^*)$ is positive when the conditions from proposition 5.3.1 are satisfied.

5.4 Mechanism design

In sight of the characterization of the ESS, we are interested in controlling the system in order to optimize for the energy consumption and the delivery probability. Let us assume that the source node controls the forwarding of message copies: a copy of the message is relayed with constant probability u upon meeting a node with no message during a local interaction, i.e., using a static forwarding policy. The main quantity of interest is denoted P_s and it is the success probability of a message at a local interaction. Under the same assumptions of linearity in [36], the average energy expenditure at the source node is $\mathcal{E} = \varepsilon\Psi$, where $\varepsilon > 0$ is the source energy expenditure per relayed message copy and Ψ is the corresponding expected number of copies released.

Live-time control

Now consider the live time control. The probability that the tagged mobile relays the copy of the packet to the destination within live time τ is given by $1 - Q_\tau^u$ where Q_τ^u is given by $Q_\tau^u = \dfrac{e^{-\lambda u\tau} - ue^{-\lambda\tau}}{1-u}$ and the probability that it relays the copy of the message if it chooses live time τ' is given by $1 - Q_{\tau'}^u$ where $Q_{\tau'}^u$ is given by $Q_{\tau'}^u = \dfrac{e^{-\lambda u\tau'} - ue^{-\lambda\tau'}}{1-u}$. Then the success probability in a local interaction with N mobiles is:

$$P_s(u|N=k) = 1 - \left[\left(Q_{\tau'}^u\right)^{k(1-\alpha(y(u)))} . \left(Q_\tau^u\right)^{k\alpha(y(u))}\right]$$

$$\Rightarrow P_s(u) = 1 - \sum_{k=1}^{\infty} P(N=k)\left[\left(Q_{\tau'}^u\right)^{k(1-\alpha(y(u)))} . \left(Q_\tau^u\right)^{k\alpha(y(u))}\right].$$

Using the same notations for the ESS and the Poisson distribution, at the equilibrium we have :

$$\alpha(y^*(u)) = \frac{LambertW(-\frac{\rho\beta e^{-\frac{\rho(\beta e^\rho - c)}{c}}}{c})c + \rho\beta e^\rho}{c\rho\beta}$$

where $\rho = Q_{\tau'}^u\gamma$, $c = (e^\gamma - e^{Q_{\tau'}^u\gamma})\dfrac{P_A(\tau') - P_A(\tau)}{1 - P_A(\tau)}$ and $\beta = 1 - \dfrac{Q_\tau^u}{Q_{\tau'}^u}$.

The probability of success is:

$$P_s(u) = 1 - e^{-\gamma}\left[e^{\gamma\left(Q_{\tau'}^u\right)^{(1-\alpha(y(u)))} . (Q_\tau^u)^{\alpha(y(u))}} - 1\right].$$

In the following theorem we give some results about the probability of success according to the behavior of the ESS when the controls change at the source.

Theorem 5.4.0.1. *The maximum value of the probability of success is attained for* $y* = argmax\{P_s(1), P_s(u_0)\}$, *where* u_0 *satisfies* $y^*(u_0) = 1$ *and* $y^*(u_0 + \delta) \leq 1,\ \delta > 0$.

We avoid the proof of the theorem here for clarity of the developments. The reader can refer to appendix 7.5 for a detailed proof of this theorem. Analysis from simulations validate our result that increasing the control at the source does not alway insure a higher probability of success given that the ESS $y*$ changes accordingly. A similar observation can also be deduced for the Dirac distribution.

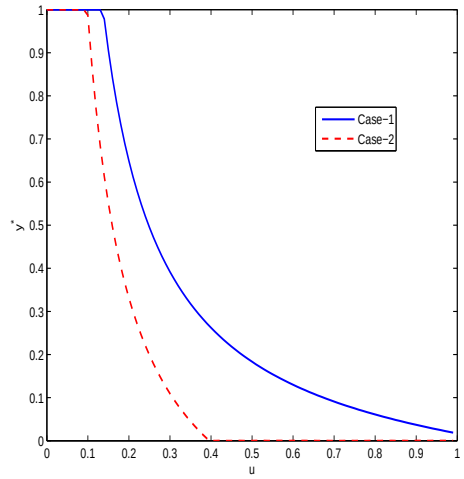


Figure 5.1: *Evolution of the ESS using a poisson distribution*

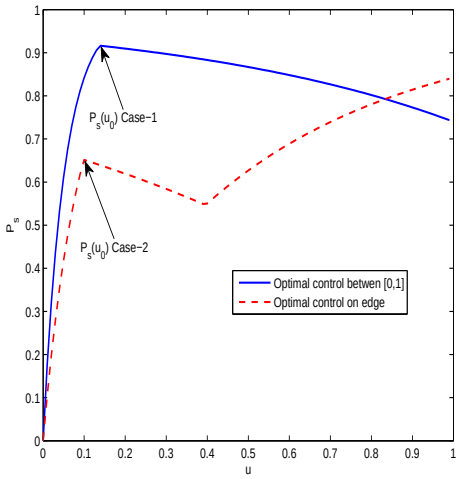


Figure 5.2: *Evolution of the probability of success using a poisson distribution*

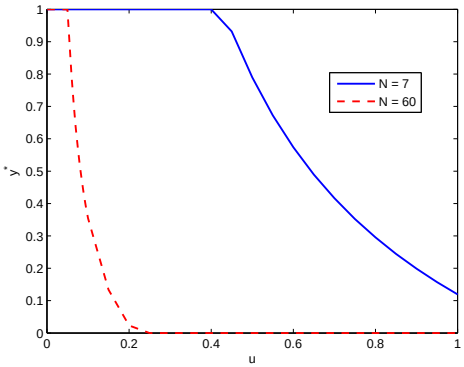


Figure 5.3: *Evolution of the ESS using a Dirac distribution*

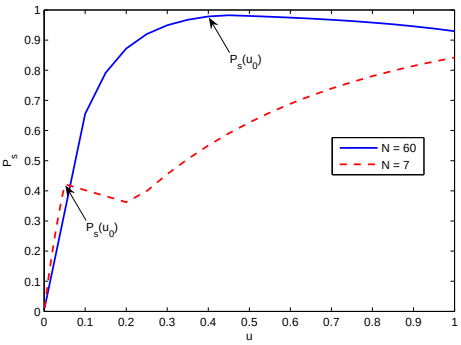


Figure 5.4: *Evolution of the probability of success using a Dirac distribution*

5.5 Replicator Dynamics

Another important feature of the Evolutionary Game framework is the Replicator Dynamics[124]. Through differential equations, also known as the replicator equations, they describe how the fraction of the population that uses different actions evolves over time. These equations are obtained by making assumptions on the way that the probability of using an action increases as a function of the fitness that an individual that has chosen the action is receiving. Those equations are based on the fact that the average growth rate per individual that uses a given strategy is proportional to the excess of fitness of that action with respect to the average fitness.

Recall that α (resp. α') represents the proportion of the population that uses strategy τ (resp. τ') and y is the proportion of individuals who play strategy τ in state A. Then, the replicator dynamic equation which describes the evolution of the proportion y is given by:

$$\begin{aligned} \dot{y} &= y\left[V_\tau(F,\alpha(y)) - (yV_\tau(F,\alpha(y)) + (1-y)V_{\tau'}(F,\alpha(y)))\right], \\ &= y(1-y)(V_\tau(F,\alpha(y)) - V_{\tau'}(F,\alpha(y))), \\ &= y(1-y)H(y). \end{aligned} \tag{5.4}$$

Using the argument proposed in [101], we conclude our replicator dynamic has the property of positive correlation which ensures that any equilibria of our game are the stationary points of the replicator dynamics. Finally, given this important property, for all interior point $\alpha_0 \in]0,1[$, the replicator dynamic defined in equation (5.4) converges to the ESS y^*.

5.6 Conclusion

Throughout this chapter Markov decision evolutionary games are used to model competition between individual mobiles acting as relay nodes in a DTN routing game. The objective of the source node is to maximize the probability of success of delivering a message to destination. However, mobiles decide to join message relaying based on their current energy state, which in turn is influenced by the forwarding control used by the source, in trade for reward.

Under this framework, we studied a source-controlled evolutionary game aimed at optimizing the energy consumed by relays. We observed a clear trade-off, where the optimal solution in general does not correspond to forwarding at full rate at the source node, and we showed cases where such a greedy strategy is well sub-optimal in maximizing the probability of success at the equilibrium.

Chapter 6

Energy efficient Minority Game for Delay Tolerant Networks

6.1 Introduction

Delay Tolerant Networks (DTNs) are designed to cope with scarce coverage. Thus, the standard problem is how to maximize the delivery probability of a message under constraints on the resources spent to forward it to destination. To this respect, efficient routing was studied first. Aim is to avoid greedy solutions such as epidemic routing where the success probability is maximized together with the number of message copies [127, 92]. In an effort to optimize the network performance under various resources constraints, several other studies have further included the use of activation and/or forwarding control at relays [83, 6, 38]. However, due to limited energy or memory capacity, not always relays can be active and participate to message routing. For instance, owners of relay devices such as smartphones or tables may not be willing to have battery depleted to sustain DTNs communications. From the forwarding standpoint, in turn, massive de-activation of relays becomes a core threat. Under two hop routing, for instance, a linear decrease in the number of relays determines the exponential decay in the delivery probability. In our framework we assume that the decision to participate to relaying or not, is taken autonomously by relay nodes according to an incentive scheme. Incentives engender a competition among relays that play strategies on their activation. The objective here is to attain an operating point for the DTN which is the solution of a joint optimization problem involving the number of active relays *and* the energy cost. The relay activation control in turn is fully decentralized and does not require additional control messages. In order to do so, we use a novel and specific utility structure. Such utility is rooted on the following trade off: the success of a tagged relay depends explicitly on the number of opponents met, namely, nodes adopting the same strategy. In fact, the bigger the number of relays participating to the message delivery, the higher the delivery probability for the message, but indeed the less the chance for the tagged relay to receive a reward from the system. The global activation target settles the number of opponents of a randomly tagged relay, i.e., the active frac-

tion of the population. Overall, the chapter is pivoted around this new approach: by modeling competition of relay nodes as a coordination game we show that it is possible to enforce a behavior of cooperation within a population of relays through competition. In fact, we will rely on the theory of the Minority Game (MG) [78] which is rooted on dynamical competition. MG does not require explicit coordination among the relay nodes: this makes it attractive because control messages among DTN nodes may experience unpredictable delays due to lack of persistent connectivity. In the last part of this chapter, we investigate how to account for the presence of heterogeneous agents. The MG rules performance of competing relays and welfare of the DTN (number of message copies and delivery message) and thus configures as an appropriate tool to drive the network to a desired operating point. We thoroughly investigate the properties of our coordination game in which relays compete to be in the population minority. Finally, since the MG scheme rules the number of active relays, the message source can achieve a target performance figure, e.g., the probability of successful message delivery, by setting the rewarding mechanism appropriately. Conversely, the source can reduce the quality of service in order to reduce the relays energy consumption. Thus, our incentive mechanism can match quality of service metrics such as delivery probability to the available resources.

Compared to existing literature, the novelty of this approach stands in the way the activation and forwarding process is jointly controlled by the operator of the network acting on a distributed mechanism which takes place among competing relays based on the MG. We will specialize the new mechanism in two frameworks: the first one is the single-class model (namely, the homogeneous DTN case), the second takes into consideration the existence of several classes of nodes (namely, the heterogeneous DTN case). Finally, we provide an algorithmic formulation of the game and demonstrate that the solution of the MG can be attained under adaptation of each ones expectation about the future.

6.1.1 Background and contribution

The minority game studies how individuals of a population of heterogeneous agents may reach a form of coordination when sharing resources for which the utility decreases in the number of competitors. Upon introducing adaptation of strategies based on each one's expectation about the future, the game can describe a dynamical system with many interacting degrees of freedom where cooperation is implicitly induced among agents. The MG was first introduced in literature as a simplification of the El Farol Bar's attendance problem [78, 59]. In the El Farol bar problem [46] N users decide independently whether to go to the unique bar in Santa Fe that offers entertainment. However, the bar is small, and they enjoy only if at most Ψ of the possible N attendees are present, in which case they obtain a reward r at a cost $0 \leq c \leq r$ for going to the bar. Otherwise, they can stay home and watch stars with utility 0. Players have two actions: go if they expect the attendance to be less than Ψ people or stay at home and watch stars if they expect the bar will be overcrowded. The original formulation as a single stage game, El Farol bar game has $\binom{N}{\Psi}$ pure Nash equilibria and a single symmetric

mixed Nash equilibrium at zero utility where each player uses the value of p such that $\sum_{h=0}^{\Psi-1} \binom{N-1}{h} p^h = c$.

The extension of the game introduces a learning component based on the belief of future attendance that every player has: the only information available is the number of people who came to El Farol in past weeks. In particular, [104] and its follow up [103] apply the concept of evolutionary MG to complex networks including random and scale-free networks. Authors of [74] apply MGs to cognitive radio networks for the design of MAC layers. All those works consider an odd number of interacting agents and do not suggest the exact analysis of equilibrium points as we suggest in this work; a further key added value of our work is the application of a standard economic estimator, namely, the logit belief model, which provides a suitable convergence framework for our mechanism design. Finally, from the application standpoint, and to the best of our knowledge, it is the first time the concept of MG is applied to DTNs with the aim to derive a mechanism to induce coordination in a non-cooperative fashion.

The remainder of the chapter is organized as follows. In Sec. 6.2 we introduce the system model and the notation used throughout the chapter. Results for the equilibria of the MG are derived in in Sec. 6.3. The extension to the multiclass DTN case is provided Sec. 6.4. A distributed reinforcement learning algorithm able to drive the system to the desired operating point is derived in Sec. 6.5. In Sec. 6.6 we study a particular case of application. Numerical results for validating the outcomes of the theoretical analysis are reported and discussed in Sec. 6.7. Final remarks are reported in Sec. 6.8

6.2 Network model

Consider a standard description of a DTN framework with several sources and destinations (see chapter 4: section 4.3). In a particular scenario in section 6.6, we will consider the two hop routing scheme, in which any mobile that receives a copy of the packet from a given source can only forward it to its destination.

Remark 6.2.1. *Since in DTNs the sources cannot predict neither the forwarding path nor the minority community nodes, some rewarding models assume for example that, the reward is distributed by the current intermediate nodes without the involvement of the source. This can be realized using, for example, the* layered coin *method proposed in [130].*

6.2.2 Network Game

In this section we detail the payoff structure of the proposed mechanism. When a message is generated by a source node, the competition takes place during the message lifetime, i.e., with duration τ. Each mobile has two strategies: either to participate to forwarding, i.e., pure strategy *transmit* (T), or not to participate, i.e., pure strategy *silent* (S). Mixed strategies, i.e., probability distributions over the two possible actions, are also possible and will be described later on.

Each strategy corresponds to a certain utility for the relay. Let's now detail how the minority game develops. First, let $\Psi > 0$ be the threshold fixed by some operator (e.g., the source nodes): it defines the majority/minority of nodes using the two policies. Hence, the utility of player is designed in such a way that, upon successful delivery of message to the destination, an active mobiles may receive a positive expected reward conditional to the fact the actives mobiles represent the minority and the mechanism selected by network operator. Other nodes receive in this case the opposite as a non-positive expected reward. The customary way to interpret this non-positive reward is that of a regret for abstention. Formally, let N be the total number of nodes involved in the competition. The probability that an active mobile relays the copy of the packet to the destination within time τ is denoted by $1 - Q_\tau$ where Q_τ is the probability for the tagged relay for not succeeding in message relaying to destination. At time $t = 0$, each relay plays T or plays S: players who take the minority action win, whereas the majority loses. Now, let $N = N_T + N_S$, where N_T (resp. N_S) is the number of agents selecting strategy T (resp. S). A tagged relay playing strategy T is member of the minority if $N_T \leq \Psi$, otherwise it loses; silent agents win as $N_S \leq N - \Psi$. The probability of receiving a reward R, for an active relay is a function of inter-meeting rate, live time, reward mechanism used by the operator and number of active relays. The total reward $R = \sum_s r^s P^s_{succ}(T, k, s)$ with $P_{succ}(T, k, s)$, the probability of an active node to receive a reward r^s from source s when k nodes are active. We denote by g the energy spent by a relay node when it remains active during $[0, \tau]$. From the sources point of view, performance should be guaranteed above some target level: $D^s_{succ} \geq D^{th}_{succ}$, where D^s_{succ} is the probability of successful delivery of a message:

$$D^s_{succ}(N_T) \qquad = 1 - \prod_{k=1}^{N_T} Q^k_\tau = 1 - Q^{N_T}_\tau \tag{6.1}$$

and D^{th}_{succ} is the performance threshold imposed by the source. Recall the fundamental trade-off: larger successful delivery comes at the price of larger value of N_T and then larger energy cost for active nodes. The connection between the network performance and the game depends on the total reward R set by the network operator for successful delivery where each r^s is decided by the source s: larger rewards causes more nodes to be active which yields a higher delivery probability at the expense of battery depletion, and network's lifetime. How to define the reward in order to attain a given performance level: we let threshold Ψ obey to the relation

$$\sum_s r^s \cdot P^s_{succ}(T, \Psi, s) = g\tau$$

where $g \geq 0$ is a constant cost of activation per second for each relay. Note that Ψ is chosen such as to equalize the total energy cost spent by nodes for being active in $[0, \tau]$ and the expected reward obtained for a successful delivery (see Fig. 6.1). In the homogeneous case ($P^s_{succ} = P_{succ}\ \forall s$), in which the relay and sources have similar physical characteristic, e.g. transmission range, mobility patterns, energy capacities etc, the last relation becomes

$$n_s r \cdot P_{succ}(T, \Psi) = g\tau \tag{6.2}$$

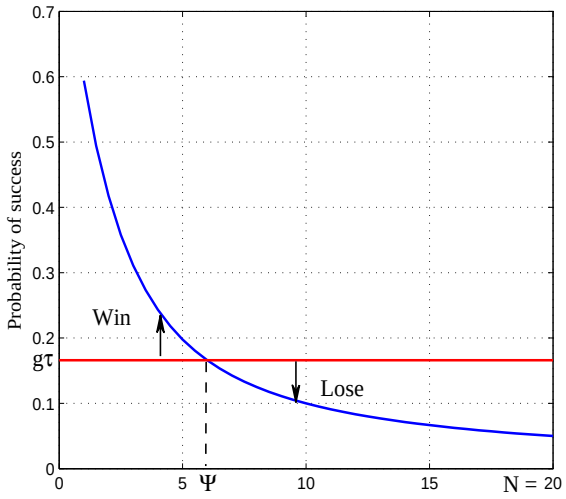


Figure 6.1: *Outcome picture of the game as observed by an active node: the intersection corresponds to the threshold value for the minority being attained by active nodes, i.e.,* $N_T = \Psi$.

where n_s is the number of sources in the network. We now state the assumption required for the function $P_{succ}(T, k, s)$

Assumption A

The function $P^s_{succ}(T, k, s)$ is decreasing in k, i.e., number of active relays. Now we can introduce two utility functions for our game, under the assumption that the population of sources is homogeneous: $P^s_{succ}(T, k, s) = P_{succ}(T, k)\ \forall s$:

Scenario 1 Zero-sum utility

$$U(T, N_T) = \sum_s r^s \cdot P^s_{succ}(T, N_T, s) - g\tau, \quad U(S, N_S) = -U(T, N_T)$$

Scenario 2 Fixed regret utility

$$U(T, N_T) = \sum_s r^s \cdot P^s_{succ}(T, N_T, s) - g\tau, \quad U(S, N_S) = -\alpha, \forall\ N_S$$

where in the second case the utility of non-active nodes expresses the regret or satisfaction for not participating to message relaying. In particular, we assume $\alpha \geq 0$, and we define N^α_T such that $U(T, N^\alpha_T) = -\alpha$.

The formulation of **Scenario 1**, requires nodes to estimate P_{succ}. This can be calculated over time by interrogating neighboring nodes and averaging their success rate: this amounts to run a pairwise averaging protocol as in [50]. In case we want to avoid the use of gossip mechanisms, we can model regret of non-active nodes as a constant negative perceived utility, which corresponds to Scenario 2.

Remark 6.2.3. *In minority games with odd number of opponents, different types of equilibria have been characterized numerically, e.g., see Challet and Zhang [25], Moro [78]. The minority*

rule sets the comfort level at $(N_T, N_S) = (\Psi, N - \Psi)$ *and computer simulations show that the participation rate fluctuates around* Ψ *in a* $(\Psi, N - \Psi)$ *configuration of people that participate or not. For fixed regret, it can be anticipated that the comfort level will set around* $(N_T^\alpha, N - N_T^\alpha)$.

6.3 Characterization of equilibria

In this section we provide the exact characterization of the equilibria induced by the game: we distinguish pure Nash equilibria and mixed Nash equilibria.

6.3.1 Pure Nash Equilibrium

The definition of a Nash Equilibrium in pure strategy for our game requires the following two conditions to be satisfied:

$$U(S, N_T) \geq U(T, N_T + 1) \tag{6.3}$$
$$U(S, N_T - 1) \leq U(T, N_T) \tag{6.4}$$

Thus, no player can improve its utility by unilaterally deviating from the equilibrium.
Proposition 6.3.2. *Under assumption A, there exists a pure Nash Equilibrium for our game. Moreover*
(i) for **scenario 1**, *there exists a unique NE obtained when exactly* Ψ *among the total population of* N *nodes play* T.
(ii) **scenario 2**, *there exists two Nash equilibria which are obtained when the total number of active relays is such that:* $N_T \in \{N_T^\alpha, N_T^\alpha - 1\}$

Proof. Scenario 1: First, we show that $N_T = \Psi$ is a pure Nash equilibrium:

$$U(S, \Psi) = U(T, \Psi) = 0 \geq U(T, \Psi + 1).$$

which is first condition (6.3). In the same way

$$U(S, \Psi - 1) = -U(T, \Psi - 1) \leq 0 = U(T, \Psi)$$

and we have second condition (6.4).

Second, we show that at the NE: $(N_T, N_S) = (\Psi, N - \Psi)$. By contradiction: let $N_T > \Psi \Rightarrow U(S, N_T) \geq U(T, N_T + 1)$, i.e., (6.3) holds. However,

$$U(S, N_T - 1) = -U(T, N_T - 1) \geq 0 > U(T, N_T)$$

and (6.4) fails. Conversely, let $N_T < \Psi \Rightarrow U(S, N_T - 1) \leq U(T, N_T)$ so that (6.4) holds. But,

$$U(S, N_T) = -U(T, N_T) < 0 \leq U(T, N_T + 1)$$

and (6.3) fails. Hence, $N_T = \Psi$ is the only possible pure Nash equilibrium. Scenario 2: Let $N_T \in \{N_T^\alpha, N_T^\alpha - 1\}$ we have,

$$\begin{cases} U(S, N_T) = -\alpha = U(T, N_T^\alpha) & \geq & U(T, N_T + 1), \\ U(S, N_T - 1) = -\alpha = U(T, N_T^\alpha) & \leq & U(T, N_T), \end{cases}$$

where equality holds in the first relation for $N_T = N_T^\alpha - 1$ and in the second for $N_T = N_T^\alpha$. We show that if $N_T \notin \{N_T^\alpha, N_T^\alpha - 1\}$ then $(N_T, N - N_T)$ then (6.3) or (6.4) fails. In fact, if $N_T > N_T^\alpha$ we have,

$$\begin{aligned} U(S, N_T) = -\alpha = U(T, N_T^\alpha) \geq U(T, N_T + 1), \text{but:} \\ U(S, N_T - 1) = -\alpha = U(T, N_T^\alpha) > U(T, N_T) \end{aligned}$$

Second, if $N_T < N_T^\alpha - 1$ we have,

$$\begin{aligned} U(S, N_T - 1) = -\alpha = U(T, N_T^\alpha) < U(T, N_T), \text{but:} \\ U(S, N_T) = -\alpha = U(T, N_T^\alpha) < U(T, N_T + 1) \end{aligned}$$

Which concludes the proof for the second scenario. □

Remark 6.3.3. *A crucial design issue is how to relate the parameters of the game to the performance of the DTN at the equilibrium. From (6.1), the number of active nodes required to attain D_{succ}^{th} is $N_T^{th} = \dfrac{\log(1 - D_{succ}^{th})}{\log(Q_\tau)}$. Besides, from Proposition 6.3.2 it must be $\Psi = N_T^{th}$. Replacing in (6.2) we obtain:*

$$r^* = g\tau \frac{1}{n_s P_{succ}(T, N_T^{th})}$$

Message reward r at the equilibrium is proportional to energy cost g through a positive constant.

6.3.4 Mixed Nash Equilibrium

Let's consider now that relay nodes maintain a probability distribution over the two actions. Compared to pure strategy game, in the mixed strategy game every node can define the strategy by which it will be active only for a fraction of the time and stay silent the rest of the time. This kind of equilibrium is desirable for an homogeneous population of nodes with similar energy constraints.

In the mixed strategy game, node i can choose to play action T with probability p_i and play S with probability $(1 - p_i)$. We let, $\mathbf{p} = (p_1, p_2, ..., p_N)$, $p_i \geq 0$, $\forall i$ the mixed strategy profile of our game. If $0 < p_i < 1$, $\forall i$ then $\mathbf{p}$ is a fully mixed strategy profile of the game. A standard companion notation that we use for $\mathbf{p}$ is $(p_i, \mathbf{p_{-i}})$: it denotes the strategy profile of the game when relay i uses strategy p_i and others use $\mathbf{p}_{-i} = (p_1, .., p_{i-1}, p_{i+1}, .., p_N)$. Let's denote by $V^i(\tilde{p}, \mathbf{p}_{-i})$ the utility of node i playing action T with probability $\tilde{p}$. We have the following definition of the mixed strategy Nash Equilibrium:

Definition 6.3.4.1. *(i) A **mixed strategy Nash Equilibrium** specifies a mixed strategy $p_i^* \in [0,1]$ for each player i (where $i = 1 \dots N$) such that :*

$$V^i(p_1^*, .., p_{i-1}^*, p_i^*, p_{i+1}^*, .., p_N^*) \geq V^i(p_1^*, .., p_{i-1}^*, p_i, p_{i+1}^*, .., p_N^*) \tag{6.5}$$

for every mixed strategy $p_i \in [0,1]$.

*(ii) We call a **Fully mixed Nash Equilibrium** a mixed strategy Nash equilibrium $\mathbf{p} = (p_1, .., p_i, .., p_N)$ with $p_i \notin \{0,1\}, \forall i$.*

From now on we will denote by the term **'mixer'** a relay who uses a mixed strategy $0 < p_i < 1$. The following proposition states that any mixed equilibrium $\mathbf{p}$ with $p_i \notin \{0,1\} \forall i$, is symmetric, i.e. $p_i = p \ \forall i$. This result comes from the fact that given any pair of mixers, a player is better off if the other chooses differently. Moreover, at the equilibrium each player must be indifferent on whether it is active or silent.

Proposition 6.3.5. *Assume assumption A holds. Let $\mathbf{p}$ be the mixed strategy profile of our game s.t $p_i \notin \{0,1\}$, then at the equilibrium, all mixers must use the same probability p, i.e., $p_i = p_j \ \forall$ mixer i, j.*

Proof. Assume that the set of mixers is not empty and let suppose that there are l relays that select pure strategy T and r pure strategy S. Without loss of generality let the strategy profile at the equilibrium :

$$\mathbf{p} = (p_1, \dots, p_{N-l-r}, 1, \dots, 1, 0, \dots, 0)$$

Scenario 1: The utility for a mixer relay i writes

$$V^i(\tilde{p}, \mathbf{p}_{-i}) = (2\tilde{p}_i - 1) F(p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_N)$$

with

$$\begin{aligned} F(p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_N) \quad = \quad & \prod_{j \neq i}^{N-l-r} (1 - p_j) U(T, l+1) + \sum_{j \neq i}^{N-l-r} p_j \prod_{j' \notin \{i,j\}}^{N-l-r} (1 - p_{j'}) U(T, l+2) + \\ & \sum_{j,j' \neq i}^{N-l-r} p_j p_{j'} \prod_{j'' \notin \{i,j,j'\}}^{N-l-r} (1 - p_{j''}) U(T, l+3) + \dots + \prod_{j \neq i}^{N-l-r} p_j U(T, N-r). \end{aligned}$$

Note about this function that:

- F is strictly decreasing by any unilateral increase of p_j by node j. This comes from the fact that the utility function of an active node is decreasing with the number of active nodes (assumption A).
- For any two mixers $j \neq j'$, p_j and $p_{j'}$ are indifferently interchangeable variables in F.

At mixed equilibrium $\mathbf{p}$, $\dfrac{\partial V^i(\mathbf{p})}{\partial \tilde{p}_i} = 0 \ \forall \ i \in \{1, \dots, N-l-r\}$. This implies that:

$$F(p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_N) = 0, \forall \text{ mixer } i$$

. Now suppose that there exists two mixers i and j, s.t. $p_i^* \neq p_j^*$. Without lost of generality assume that $p_i^* < p_j^*$, then

$$\begin{aligned} 0 = F(p,..,p_{i-1},p_{i+1},..,p_j,..,p_N) &> F(p_1,..,p_{i-1},p_{i+1},..,p_i,..,p_N) \\ &= F(p_1,..,p_{j-1},p_{j+1},..,p_N) > 0 \end{aligned}$$

which is absurd. Thus $p_i = p_j$, $\forall$ mixers i, j.

Scenario 2: As in **Scenario 1**, the utility perceived by a given mixer i when the strategy profile is $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ is given by:

$$\begin{aligned} V^i(\tilde{p}, \mathbf{p}_{-i}) &= \tilde{p}_i F(p_{-i}) - \alpha(1-\tilde{p}_i)\Big[\prod_{j\neq i}^{N-l-r}(1-p_j) + \sum_{j\neq i}^{N-l-r} p_j \prod_{j'\notin\{i,j\}}^{N-l-r}(1-p_{j'}) \\ &\quad + \sum_{j,j'\neq i}^{N-l-r} p_j p_{j'} \prod_{j''\notin\{i,j,j'\}}^{N-l-r}(1-p_{j''}) + \ldots + \prod_{j\neq i}^{N-l-r} p_j\Big] \end{aligned}$$

At the equilibrium we have, $\forall$ mixer i, $\dfrac{\partial V^i(\mathbf{P})}{\partial p_i} = F'(p_{-i}) = 0$, where F' has exactly the same shape as F with $U(T,k)$ replaced by $U(T,k)+\alpha$, $k \in \{l+1,\ldots,N-r\}$. We then use the same reasoning as done with function F and conclude that, $p_i^* = p_j^*$, $\forall$ mixers i, j.

□

In the following corollary, we restrain the result of proposition 6.3.5 to the special case when every nodes act as mixers.

Corollary 6.3.6. *Under assumption A, any fully mixed equilibrium* $\mathbf{p}$ *with* $p_i \notin \{0,1\}, \forall i$, *is symmetric, i.e.* $p_i = p$ $\forall i$.

The following proposition characterize the existence and uniqueness of a fully mixed Nash Equilibrium.

Proposition 6.3.7. *Under assumption A, there exists a unique fully mixed Nash Equilibrium* $\mathbf{p}^*$. *Moreover,* $\mathbf{p}^*$ *is solution to:*

- **Scenario 1** *:*

$$A(N,p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1}(p^*)^{k-1}(1-p^*)^{N-k}U(T,k) = 0. \tag{6.6}$$

- **Scenario 2** *:*

$$A'(N,p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1}(p^*)^{k-1}(1-p^*)^{N-k}[U(T,k)+\alpha] = 0.$$

Proof. Let p the symmetric mixed strategy adopted by every node in the game, $p_i = p$, $\forall i$.

Scenario 1: The utility of one relay i when the strategy profile (p_i, p_{-i}) is played is given by:

$$\begin{aligned} V^i(\tilde{p}_i, p_{-i}) &= \tilde{p}_i \sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1} (1-p_{-i})^{N-k} U(T,k) + (1-\tilde{p}_i) \sum_{k=0}^{N-1} C_k^{N-1} p_{-i}^{k} (1-p_{-i})^{N-k-1} U(S,k+1) \\ &= \tilde{p}_i \sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1} (1-p_{-i})^{N-k} U(T,k) + (1-\tilde{p}_i) \sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1} (1-p_{-i})^{N-k} U(S,k) \\ &= (2\tilde{p}_i - 1) \sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1} (1-p_{-i})^{N-k} U(T,k) \end{aligned}$$

Let $A(N, p_{-i}) = \sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1} (1-p_{-i})^{N-k} U(T,k)$

if $A(N, p_{-i}) < 0$, $p_i = 0$ is the best response for player i and conversely, $p = 1$ is a best response when $A(N, p_{-i}) > 0$. A mixed strategy is obtained when $A(N, p_{-i}) = 0$. Also, we have

$$A(N,0) = U(T,1) > 0 > A(N,1) = U(T,N)$$

thus there exists a mixed symmetric Nash Equilibrium which is unique since $A(N, p_{-i})$ is strictly decreasing with p. The mixed equilibrium is thus characterized by equation (6.6).

$$A(N, p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1} (p^*)^{k-1} (1-p^*)^{N-k} U(T,k) = 0.$$

Scenario 2: The utility of one relay i when the strategy profile $(\tilde{p}_i, p_{-i})$ is played is given by:

$$V^i(\tilde{p}_i, p_{-i}) = \tilde{p}_i \sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1} (1-p_{-i})^{N-k} U(T,k) - \alpha(1-\tilde{p}_i)$$

At the Nash equilibrium we have, $\forall$ player i, $\dfrac{\partial V^i(p^*)}{\partial p^*} = A'(N, p^*) = 0$ with

$$A'(N, p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1} (p^*)^{k-1} (1-p^*)^{N-k} [U(T,k) + \alpha]$$

Since α is a fixed positive constant, $A'(N, p^*)$ has the same properties as $A(N, p^*)$ from the proof of scenario 1. Then we easily conclude that, p^* is unique and characterized by :

$$A'(N, p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1} (p^*)^{k-1} (1-p^*)^{N-k} [U(T,k) + \alpha] = 0.$$

□

6.3.8 Equilibrium with mixers and non-mixers

We study here the existence of equilibrium when the population of agents is composed of pure strategy players: active or non-active, as well as mixers. In this case, a non-pure Nash equilibrium can be represented by the triplet (l, r, p^*), where $l, r \in \{0, 1, \ldots, N\}$ denote respectively the number of agents choosing pure strategy T or S, and $p^* \in (0, 1)$ the probability with which the remaining $N - l - r$ mixers choose strategy T. Moreover, we denote by $v_T(l, r, p)$(resp. $v_S(l, r, p)$) the expected payoff to a player choosing T(resp. S). The expressions of $v_T(l, r, p)$ and $v_S(l, r, p)$ write as follow:

$$v_T(l, r, p) = \sum_{k=0}^{N-l-r} C_k^{N-l-r} p^k (1-p)^{N-l-r-k} U(T, l+k) \tag{6.7}$$

and

$$v_S(l, r, p) = -\sum_{k=0}^{N-l-r} C_k^{N-l-r} p^k (1-p)^{N-l-r-k} U(T, l+k) \tag{6.8}$$

Proposition 6.3.9. *Using the previous notations, a strategy profile of type (l, r, p^*) is a Nash equilibrium with at least one mixer if and only if:*

$$v_T(l+1, r, p^*) = v_S(l, r+1, p^*) \tag{6.9}$$

We prove that this result holds for Zero-sum utility and fixed regret utility for non-active nodes(resp. **Scenario 1** *and* **Scenario 1***)*

Proof. The condition (6.9) describes that a mixer is indifferent whether it chooses a pure strategy T or S. This is a necessary condition for the strategy profile (l, r, p^*) to be a Nash equilibrium.
In order to show sufficiency, we need to show that pure strategy players as well, cannot improve their expected utility through unilateral deviation from the equilibrium profile. Without loss of generality, suppose that there is at least one player using pure strategy T, we have

$$\begin{aligned} v_T(l, r, p^*) &\geq v_T(l+1, r, p^*) = v_S(l, r+1, p^*) \\ &\geq v_S(l-1, r+1, p^*) \\ &\geq p^* v_T(l, r, p^*) + (1-p^*) v_S(l-1, r+1, p^*) \end{aligned}$$

This last relation, states that an active user cannot improve its expected utility by unilaterally deviating from the strategy profile (l, r, p^*) using any strategy $p^* \in [0, 1)$, given relation (6.9). As done for **Scenario 1**, in **Scenario 1**, we have, $v_S(l, r+1, p^*) = -\alpha$, let $v_S(l+1, r, p^*) = -\alpha$ then:

$$\begin{aligned} v_T(l, r, p^*) \geq\ & v_T(l+1, r, p^*) = -\alpha \geq v_S(l-1, r+1, p^*) \\ \geq\ & p^* v_T(l, r, p^*) + (1-p^*) v_S(l-1, r+1, p^*) \end{aligned}$$

moreover,

$$v_S(l+1, r-1, p^*) \leq v_T(l+1, r, p^*) = -\alpha = v_S(l, r, p^*).$$

This completes the proof. □

Discussion on existence of (l, r, p^*) type equilibria
It is possible to isolate several cases where the relation (6.9) that characterizes a Nash Equilibrium of type (l, r, p^*), cannot be satisfied.
We denote by, $p = 0^+$(resp. $p = 1^-$) the mixed strategy infinitely close to zero(resp. to one), with which at least one mixer select to be active. Since, $v_T(l, r, p^*)$ is strictly decreasing with l and p^*, we have, $v_T(l+1, r, p^*) = v_S(l, r+1, p^*)$

$$\iff \begin{cases} v_T(l+1, r, 0^+) & > & -v_T(l, r+1, 0^+) \\ v_T(l+1, r, 1^-) & \leq & -v_T(l, r+1, 1^-), \end{cases}$$

(1) If $l \geq \Psi$, then there is no Nash equilibrium of the desired type. Indeed, $l > \Psi$, then $v_T(l, r+1, 0^+) \leq 0$ and

$$v_T(l+1, r, 0^+) \leq 0 \leq -v_T(l, r+1, 0^+).$$

Then there is no possible Nash Equilibrium according to relation (6.9).

(2) If $l + r + 1 > N - 1$, then there is no Nash equilibrium. We already have $l < \Psi$, let $l + r + 1 = N$ then,

$$v_T(l+1, r, p) = C_1 \geq 0 \; \forall \; p \text{ and}$$

$$v_S(l, r+1, p) = C_2 > 0 \; \forall \; p.$$

Since v_T is decreasing with l, we have, $0 \leq C_1 < C_2$ which contradicts relation (6.9).

A Nash Equilibrium of type (l, r, p^*) exists then only for $l < \Psi$ and for $l + r \leq N - 2$, thus there are exactly $\Psi(N-2) - \dfrac{\Psi(\Psi-1)}{2}$ Nash equilibria. In the following proposition we go further and decline some properties of the symmetric mixed strategy p^* at the equilibrium.
Proposition 6.3.10. *The mixed strategy p^* at the equilibrium increases as r increase and reversely decreases as l increase.*

Proof. For a fixed number l of nodes playing pure strategy T, the utility of a mixer when there are less nodes playing pure strategy S, decreases faster than when there are more nodes playing pure strategy S. For example we have,

$$\frac{\partial v_T(l+1, 0, p)}{\partial p} > \frac{\partial v_T(l+1, 1, p)}{\partial p}$$

Similarly, we will have

$$\frac{\partial v_T(l, 1, p)}{\partial p} > \frac{\partial v_T(l, 2, p)}{\partial p}.$$

Since, $v_T(l+1, 0, 0^+) = v_T(l+1, 1, 0^+)$ and $v_T(l, 1, 0^+) = v_T(l, 2, 0^+)$ then if p_1^*, p_2^* are such that $v_T(l+1, 0, p_1^*) = -v_T(l, 1, p_1^*)$ and $v_T(l+1, 1, p_2^*) = -v_T(l, 2, p_2^*)$, it follows that $p_1^* < p_2^*$.
The same reasoning holds for every $k < k'$ and p_1^*, p_2^* *s.t.* $v_T(l+1, k, p_1^*) = -v_T(l, k+$

$1, p_1^*)$ and $v_T(l+1, k', p_2^*) = -v_T(l, k'+1, p_2^*)$ then $p_1^* < p_2^*$.
We apply a similar reasoning reversely and conclude that for a fixed number r of nodes playing pure strategy S, for every $k < k'$ and p_1^*, p_2^* *s.t.* $v_T(k+1, r, p_1^*) = -v_T(k, r+1, p_1^*)$ and
$v_T(k'+1, r, p_2^*) = -v_T(k', r+1, p_2^*)$ then $p_1^* > p_2^*$. □

Summary on characterization of equilibria Throughout this section we have characterized the following different equilibria : Under assumption A we have

1. *pure equilibrium* : We shown that for the Zero-sum utility there exists a unique pure N.E. that sets at exactly Ψ active relay nodes. For the Fixed regret utility scenario, there exists two possible N.E. for a number of active nodes $N_T \in \{N_T^\alpha, N_T^\alpha - 1\}$.
2. *fully mixed equilibrium* : For both scenarios we shown that any fully mixed equilibrium $\mathbf{p}$ with $p_i \notin \{0, 1\}, \forall i$, is symmetric. Moreover, the mixed N.E. of our game is unique and characterized by : $A(N, p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1} (p^*)^{k-1} (1-p^*)^{N-k} U(T, k) = 0$ for the zero-sum utility scenario and characterized by $A'(N, p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1} (p^*)^{k-1} (1-p^*)^{N-k} [U(T, k) + \alpha] = 0$ for the fixed regret scenario.
3. *equilibrium with mixers and non-mixers*: The last characterized type of equilibrium is related to a population of relays composed of mixers and non-mixers. Here we shown that such type of equilibrium is characterized by a specific relation, namely relation (6.9). Moreover, we established that a Nash Equilibrium of this type exists only for $l < \Psi$ and for $l + r \leq N - 2$, thus there are exactly $\Psi(N-2) - \frac{\Psi(\Psi-1)}{2}$ Nash equilibria.

6.4 The multi-class case

In the first part we adhered to a common simplifying assumption in many earlier works on modeling performances of DTNs, i.e., we assumed that DTN nodes have all similar physical characteristics, e.g. transmission range, mobility patterns, energy capacities etc., i.e., the DTN is *homogeneous*. In this section we will design a model to allow a fairness between mobiles relays based on their capacities. We extend our results to DTNs with several classes of nodes. In fact, DTNs nodes may belong to different categories, e.g., mobile, laptop, PDA and/or have related communication/energy-autonomy features depending on transmission range, mobility, memory, energy capacity and active radio interface such as WiFi and Bluetooth. A DTN with different types of nodes is classified as *heterogeneous* [24, 29].

To this respect, we assume nodes to fall into classes according to their physical characteristics: the aspect we focus on is the heterogeneity energy budget/consumption of nodes. For example, devices using Bluetooth radio instead of WiFi consume between

10 to 50 times less power [40]. More precisely, WiFi interface's power consumption in an active data transfer state is of the order of 890 mW, compared to only 120 mW for Bluetooth due to a limited range and a simpler radio architecture. For small devices such as cell-phones and PDAs, with limited power budgets, the power consumption of a WiFi radio represents a significant proportion of the overall system power [87][94][8].

The extension of the game is done by devising a class-dependent reward mechanism. In fact, nodes of classes with larger battery capacity might choose to be more active to collect the reward, while nodes of classes with a limited battery capacity may participate less in order to save energy. As before, the sources wish to satisfy performance requirements in a way that conserve energy consumption and achieve consumption fairness.

6.4.1 The model

Heterogeneous DTNs considered in this section are composed of M classes of relay nodes: class j, $1 \leq j \leq M$, contains N_j nodes with inter-meeting intensity $\lambda_j > 0$, and $N = \sum_j N_j$. We let each class j has its own threshold Ψ_j that defines the majority/minority of nodes from class j. We will often refer to the case $M = 2$ for the sake of clarity; results shown later easily extend to hold in general unless otherwise stated.

The energy consumed by nodes, when active, i.e., playing T, has a large impact on the lifetime of the battery-operated mobile nodes due to limited energy budget in DTNs. This depletion of energy depends not only on the wireless technology used by each class's nodes but also on the type of these nodes (the rate at which energy is consumed by PDA-based phones is very high compared to laptops, thus, these devices can quickly drain their own batteries). We let g_j the energy cost for a relay node of class j when it remains active during a unit of time and we consider the inter-meeting intensity is the same for all classes, i.e. $\lambda_j = \lambda, \forall 1 \leq j \leq M$.[1] For the case $M = 2$ we assume that $g_1 > g_2$ such that nodes of class 1 has higher energy requirements than nodes from class 2 to be active.

The utility function for an active node of class j is:

$$U_j(T, N_T) = r_j P_{succ}(T, N_T) - g_j \tau$$

while the utility for a silent node is:

$$U_j(S, N_T) = -r_j P_{succ}(T, N_T) + g_j \tau$$

The thresholds Ψ_j as previously defined satisfies the following relation:

$$\forall 1 \leq j \leq M : \quad r_j P_{succ}(T, \Psi_j) = g_j \tau \tag{6.10}$$

[1] Future extensions of the model will account for heterogeneity in the inter-meeting intensities [38, 24].

6.4.2 Characterizing the equilibria

Proposition 6.4.3. *In the multi-class framework: There exists a unique pure NE attained when* $(\Psi_j)_{j\in\{1,...,M\}}$ *nodes among the total population select to be active for relays of each class* j.

Proof. The Nash Equilibrium is obtained when the following two conditions are satisfied:

$$\forall 1 \leq j \leq M : \begin{cases} U_j(S, N_T) & \geq & U_j(T, N_T + 1) \\ U_j(S, N_T - 1) & \leq & U_j(T, N_T) \end{cases} \tag{6.11}$$

Assume that for any class j exactly Ψ_j nodes are active, then we have:

$$U_j(S, \Psi_j) = U_j(T, \Psi_j) = 0 \geq U_j(T, \Psi_j + 1),$$

in the same way we have:

$$U_j(S, \Psi_j - 1) = -U_j(T, \Psi_j - 1) \leq 0 = U_j(T, \Psi_j),$$

then we have the conditions in (6.11) satisfied.

We now show that there are no other pure Nash equilibria. Let, for a class j, $\Psi_j^{'} \neq \Psi_j$, without loss of generality, let $\Psi_j^{'} > \Psi_j$ then $U(S, \Psi_j^{'}) \geq U(T, \Psi_j^{'} + 1)$: first condition of (6.11). However,

$$U(S, \Psi_j^{'} - 1) = -U(T, \Psi_j^{'} - 1) \geq 0 > U(T, \Psi_j^{'})$$

and the second relation is not satisfied. Continuing with the same reasoning used in the proof of proposition (6.3.2), we obtain that at the equilibrium there are exactly Ψ_j active nodes hence the proof. □

As in the case of homogeneous DTNs, we can extend the result to mixed strategies.

Proposition 6.4.4. *Let the fully mixed strategy profile of our game in the multi-class framework* $\mathbf{p} = (p_{11}, ..., p_{N_1 1}, ..., p_{1j}, ...p_{N_i j}, ..., p_{1M}, ..., p_{N_M M})$. *At the equilibrium, all players of the same class must use the same fully mixed strategy:* $p_{ij} = p_j$, $\forall i; \forall 1 \leq j \leq M$; *the result holds both Scenarios 1 and 2.*

Proof. We denote by $(p_{ij}, \mathbf{p}_{-i})$ the fully mixed strategy profile of the game when relay i of class j uses strategy p_{ij} and others use $p_{-i} = (p_{11}, ..., p_{N_1 1}, ..., p_{1j}, ..., p_{i-1j}, p_{i+1j}, ..., p_{N_j j}, ..., p_{1M}, ..., p_{N_M M})$
Scenario 1: The utility perceived by a given player i of class j when the strategy profile is P is given by:

$$U_j^i(\mathbf{p}) = (2\tilde{p}_i - 1)F_i(p_{-i})$$

with

$$F_i = \prod_{k\neq i}(1 - p_k)U_j(T, 1) + \sum_{k\neq i} p_{km} \prod_{k'\notin\{i,k\}} (1 - p_{k'm})U_j(T, 2) \sum_{k,k'\neq i} p_{km}p_{k'm} \prod_{k''\notin\{i,k,k'\}} (1 - p_{k''m})U_j(T, 3) + ... + \prod_{k\neq i} p_{km}U_j(T, N)$$

$\forall 1 \leq m \leq M$. Note about this function that:

- F_i is strictly decreasing by any unilateral increase of p_{km} by player k of class m.
- For any two $k \neq k'$ of the same class m, the mixed strategies $p_{km}, p_{k'm}$ are indifferently interchangeable variables in F_i.

At the equilibrium we have, $\forall$ player $i, \forall 1 \leq j \leq M$, $\dfrac{\partial U_j^i(\mathbf{p})}{\partial p_{ij}} = 0$. This implies that : $F_i = 0$. Moreover, the strategy profile $\mathbf{p} = (p_{11}^*, ..., p_{N_1 1}^*, ..., p_{1j}^*, ..., p_{N_j j}^*, ..., p_{1M}^*, ..., p_{N_M M}^*)$ is a Nash equilibrium if no user can increase its utility by any unilateral deviation. Now suppose that there exists i, k of class j, such that, $p_{ij}^* \neq p_{kj}^*$. Without lost of generality assume that $p_{ij}^* < p_{kj}^*$, we have,

$$\begin{aligned} 0 &= F_i(..., p_{1j}^*, ..., p_{i-1j}^*, p_{i+1j}^*, ..., p_{kj}^*, ..., p_{N_j^* j}, ..., p_{N_M M}^*) \\ &> F_i(..., p_{1j}^*, ..., p_{i-1j}^*, p_{i+1j}^*, ..., p_{ij}^*, ..., p_{N_j^* j}, ..., p_{N_M M}^*) \\ &= F_i(..., p_{1j}^*, ..., p_{k-1j}^*, p_{k+1j}^*, ..., p_{N_j^* j}, ..., p_{N_M M}^*) \\ &> 0 \end{aligned}$$

which is absurd. Thus $p_{ij}^* = p_{kj}^*$, $\forall\ i, k$ of class j.
Scenario 2: As in scenario 1, the utility perceived by a given player i when the strategy profile is $\mathbf{P} = (p_1, p_2, \ldots, p_N)$ is given by:

$$\begin{aligned} U_j^i(\mathbf{P}) &= p_{ij} * F_i(p_{-i}) - \alpha(1 - p_{ij})\Big[\prod_{k \neq i}(1 - p_{km}) + \sum_{k \neq i} p_{km} \prod_{k' \notin \{i,k\}} (1 - p_{k'm}) \\ &\quad + \sum_{k,k' \neq i} p_{km} p_{k'm} \prod_{k'' \notin \{i,k,k'\}} (1 - p_{k''m}) + ... + \prod_{k \neq i} p_{km}\Big] \end{aligned}$$

At the equilibrium we have, $\forall$ player i, $\dfrac{\partial U_j^i(\mathbf{P})}{\partial p_{ij}} = F_i'(p_{-i}) = 0$, where F_i' has exactly the same shape as F_i with $U_j(T, k)$ replaced by $U_j(T, k) + \alpha$, $k \in \{1, \ldots, N\}$. We then use the same reasoning as done with function F_i and conclude that, $p_{ij}^* = p_{kj}^*$, $\forall\ i, k$ of class j. □

Let p_j the symmetric mixed strategy adopted by every node of class j, $p_{ij} = p_j$, $\forall i, j$. For reasons of clarity, we characterize the mixed strategy p_j^* in a two-class scenario without any loss of generality ($M = 2$).
Proposition 6.4.5. *There exists a unique fully mixed Nash equilibrium (p_1^*, p_2^*) for the multi-class case. Moreover it is the solution of, $A_1(N, p_1^*, p_2^*) = A_2(N, p_1^*, p_2^*) = 0$ where:*

$$A_1(N, p_1^*, p_2^*) = \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2} (C_{k_1}^{N_1-1} p_1^{*k_1} (1 - p_1^*)^{N_1-k_1-1} C_{k_2}^{N_2} p_2^{*k_2} (1 - p_2^*)^{N_2-k_2}) U_1(T, k_1 + k_2)$$

and

$$A_2(N, p_1^*, p_2^*) = \sum_{k_1=0}^{N_1} \sum_{k_2=0}^{N_2-1} (C_{k_2}^{N_2-1} p_2^{*k_2} (1 - p_2^*)^{N_2-k_2-1} C_{k_1}^{N_1} p_1^{*k_1} (1 - p_1^*)^{N_1-k_1}) U_2(T, k_1 + k_2)$$

Moreover,

(i) if $\frac{g_1}{r_1} = \frac{g_2}{r_2}$ *then we have* $p_1 = p_2$.

(ii) if $\frac{r_1}{g_1} < \frac{r_2}{g_2}$ *then* $g_1 > g_2 \Rightarrow p_1 < p_2$. *As a consequence we have* $\Psi_1 < \Psi_2$.

Proof. **Scenario 1** : The utility of one relay i of class 1 when the strategy profile (p_{i1}, p_{-i}) is played is given by:

$$\begin{aligned} V_1^i(p_{i1}, p_{-i}) &= p_i \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2} (C_{k_1}^{N_1-1} p_1^{k_1} (1-p_1)^{N_1-k_1-1} C_{k_2}^{N_2} p_2^{k_2} (1-p_2)^{N_2-k_2}) U_1(T, k_1+k_2) \\ &\quad + (1-p_i) \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2} (C_{k_1}^{N_1-1} p_1^{k_1} (1-p_1)^{N_1-k_1-1} C_{k_2}^{N_2} p_2^{k_2} (1-p_2)^{N_2-k_2}) U_1(S, k_1+k_2) \\ &= (2p_i - 1) * A_1(N, p_1, p_2) \end{aligned}$$

In the same way we can write the utility of a relay i of class 2 as:

$$V_2^i(p_{i2}, p_{-i}) = (2p_i - 1) * A_2(N, p_1, p_2)$$

where $A_1(N, p_1, p_2), A_2(N, p_1, p_2)$ are defined as follows:

$$A_1(N, p_1, p_2) = \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2} (C_{k_1}^{N_1-1} p_1^{k_1} (1-p_1)^{N_1-k_1-1} C_{k_2}^{N_2} p_2^{k_2} (1-p_2)^{N_2-k_2}) U_1(T, k_1+k_2),$$

and

$$A_2(N, p_1, p_2) = \sum_{k_1=0}^{N_1} \sum_{k_2=0}^{N_2-1} (C_{k_2}^{N_2-1} p_2^{k_2} (1-p_2)^{N_2-k_2-1} C_{k_1}^{N_1} p_1^{k_1} (1-p_1)^{N_1-k_1}) U_2(T, k_1+k_2).$$

As motivated in the proof of proposition (6.3.7), a mixed Nash equilibrium(p_1^*, p_2^*) is obtained here when

$$A_1(N, p_1^*, p_2^*) = A_2(N, p_1^*, p_2^*) = 0. \tag{6.12}$$

Scenario 2: The utility of an active user of Class 1 is given by:

$$\begin{aligned} V_1^i(p_{i1}, p_{-i}) &= p_i \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2} (C_{k_1}^{N_1-1} p_1^{k_1} (1-p_1)^{N_1-k_1-1} C_{k_2}^{N_2} p_2^{k_2} (1-p_2)^{N_2-k_2}) U_1(T, k_1+k_2) - (1-p_i)\alpha \\ &= p_i \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2} (C_{k_1}^{N_1-1} p_1^{k_1} (1-p_1)^{N_1-k_1-1} C_{k_2}^{N_2} p_2^{k_2} (1-p_2)^{N_2-k_2}) [U_1(T, k_1+k_2) + \alpha] - \alpha \end{aligned}$$

and utility of user i from Class 2 writes

$$V_2^i(p_{i2}, p_{-i}) = p_i \sum_{k_1=0}^{N_1} \sum_{k_2=0}^{N_2-1} (C_{k_2}^{N_2-1} p_2^{k_2} (1-p_2)^{N_2-k_2-1} C_{k_1}^{N_1} p_1^{k_1} (1-p_1)^{N_1-k_1}) [U_2(T, k_1+k_2) + \alpha] - \alpha$$

By replacing, $U_j(T, k_1 + k_2)$ by $U_j(T, k_1 + k_2) + \alpha$ in the first case we obtain the same conclusions. The proof is thus similar to the **Scenario 1**.Hence the existence of a mixed Nash equilibrium. Now let

$$C(i) = \sum_{k_1=0}^{N_1-2} \sum_{k_2=0}^{N_2-2} P(K_1 = k_1, K_2 = k_2) r_i P_{succ}(T, k_1 + k_2 + e_i + 1)$$

for user i, where $e_i = 1$ if user i is active and $e_i = 0$ otherwise. We can thus rewrite the expressions of $A_1(N, p_1^*, p_2^*)$ and $A_2(N, p_1^*, p_2^*)$ as follows:

$$A_1(N, p_1^*, p_2^*) = r_1 p_2 C(1) - r_1(1 - p_2)C(0) - g\tau \tag{6.13}$$
$$A_2(N, p_1^*, p_2^*) = r_2 p_1 C(1) - r_2(1 - p_1)C(0) - g\tau \tag{6.14}$$

It follows that, $A_1(N, p_1^*, p_2^*) = A_2(N, p_1^*, p_2^*) = 0 \implies$

$$p_2 C(1) - (1 - p_2)C(0) = \frac{g_1\tau}{r_1} \tag{6.15}$$
$$p_1 C(1) - (1 - p_1)C(0) = \frac{g_2\tau}{r_2} \tag{6.16}$$

letting $\frac{g_1\tau}{r_1} = \frac{g_2\tau}{r_2}$ we have, $p_1 = p_2$. This completes the proof of (i).
Now, let $\gamma_1 = \frac{g_1\tau}{r_1}, \gamma_2 = \frac{g_2\tau}{r_2}$ then from (6.15) and (6.16) we have:

$$\begin{aligned} (p_2 - p_1)C(1) + (p_2 - p_1)C(0) &= \gamma_1 - \gamma_2 \\ \Rightarrow (p_2 - p_1)(C(0) + C(1)) &= \gamma_1 - \gamma_2 \end{aligned}$$

Since, $C(0) > C(1) > 0$[2], then, $\gamma_1 > \gamma_2 \Rightarrow p_2 > p_1$. This tells that in order to have fewer nodes active in class 1 we should allocate smaller reward. However, if we come back to the definition of Ψ_1 we have,

$$\begin{aligned} r_1 P_{succ}(T, \Psi_1) - g_1\tau = 0 &\Rightarrow P_{succ}(T, \Psi_1) = \frac{g_1\tau}{r_1} > \frac{g_2\tau}{r_2} \\ \Rightarrow r_2 P_{succ}(T, \Psi_1) > g_2\tau &\Rightarrow P_{succ}(T, \Psi_1) > P_{succ}(T, \Psi_2) \end{aligned}$$

Under assumption A we have, $\Psi_2 > \Psi_1$.Hence the proof of (ii). □

The last result allow us to extend the the minority game with only one threshold to a minority game with several thresholds allowing to control the average number of active users in each class at equilibrium. Due to the complexity of the expressions, it's in general difficult to obtain an explicit solution of (6.12). We are able however to obtain numerical solution as shown in Fig. 6.2.

[2]This comes from the fact that the more number of active nodes, the less is the probability of obtaining the reward for a tagged node.

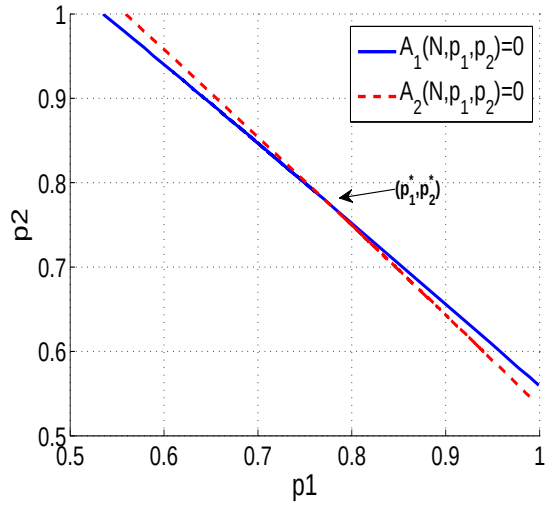


Figure 6.2: *The mixed Nash equilibrium: multi-class, where* $g1 = 0.8 \times 10^{-4}, g2 = 0.5 \times 10^{-4}, r2 = 0.15, \lambda = 0.03, \tau = 100, N_1 = 20, N_2 = 20$

6.5 Distributed reinforcement learning algorithm

In this section we introduce a distributed reinforcement learning algorithm: it permits to relays to adjust strategies they play over time in the framework of the DTN MG designed in section 6.2. The analysis of convergence of the algorithm relies on a stochastic model that gives rise to an associated continuous time deterministic dynamic system. It will be proved that this process converges almost surely towards a stationary state which is characterized as ϵ-approximate Nash equilibrium.

In DTNs, limited computational power and low energy budget of relays requires adaptive and energy-efficient mechanisms letting relays adapt to operating conditions at low cost. The learning algorithm proposed here matches this reality of DTNs since, as we shall see, it has the following attractive features:

- It is genuinely distributed: strategy updating decision is local to relays;
- It depends uniquely on the realized payoffs: nodes utilize local observations to estimate their own payoffs;
- It uses simple behavioral rule in the form of logit rule.

We assume that each relay node i has a prior perception x_i of the payoff performance for each action (To be active, or not), and makes a decision based on this piece of information using a random choice rule. The payoff of the chosen action is then observed and is used to update the perception for that particular action. This procedure is repeated round after round, each round of duration τ generating a discrete time stochastic process which is the learning process.

For notation's sake, denote $A = \{T, S\}$ the set of pure strategies, and Δ_i is the set of

mixed strategies for player i with $i \in \{1, ..., N\}$. Let $V^i(.)$ the payoff function for player i. The algorithm works in rounds of duration τ, at round k, each relay node i takes an action a_i^k according to a mixed strategy $\pi_i^k = \sigma_i(x_i^k) \in \Delta_i$. The fully mixed strategy is generated according to the vector $x_i^k = (x_{ia}^k)_{a \in A}$ which represents its perceptions about the payoffs of the available pure strategies. In particular, relay node i's fully mixed strategies are mapped from the perceptions based on the logit rule:

$$\sigma_{ia}(x_i) = \frac{e^{\beta x_{ia}}}{e^{\beta x_{iT}} + e^{\beta x_{iS}}} \tag{6.17}$$

where β is commonly called the temperature of the logit. The temperature has a smoothing effect: when $\beta \to 0$ it leads to the uniform choice of strategies, while for $\beta \to \infty$ the probability concentrates on the pure strategy with the largest perception. We assume throughout that σ_{ia} is strictly positive for all $a \in A$.

At round k, the perceptions x_{ia}^k will determine the mixed strategies $\pi_i^k = \sigma_i(x_i^k)$ that are used by each player i to choose at random action T (to be active) or S (to be silent). Then each player estimates his own payoff $\tilde{u}_i^k$, with no information about the actions or the payoffs of the other players, and uses this value ($\tilde{u}_i^k$) to update its perceptions as:

$$x_{ia}^{k+1} = \begin{cases} (1-\gamma^k)x_{ia}^k + \gamma^k \tilde{u}_i^k & \text{if } a_i^k = a \\ x_{ia}^k & \text{otherwise,} \end{cases} \tag{6.18}$$

where $\gamma^k \in (0,1)$ is a sequence of averaging factors that satisfy $\sum_k \gamma^k = \infty$ and $\sum_k (\gamma^k)^2 < \infty$ (examples of such factor are $\gamma^k = \frac{1}{k}$ or $\gamma^k = \frac{1}{1 + k \log k}$). A player only changes the perception of the strategy just used in the current round and keeps other perceptions unchanged. Algorithm (3) summarizes the learning process. The discrete time stochastic process expressed in (6.18) represents the evolution of relay node perceptions and can be written in the following equivalent form:

$$x_{ia}^{k+1} - x_{ia}^k = \gamma^k [w_{ia}^k - x_{ia}^k], \forall i \in \{1, .., N\}, a \in A \tag{6.19}$$

with

$$w_{ia}^k = \begin{cases} \tilde{u}_i^k & \text{if } a_i^k = a \\ x_{ia}^k & \text{otherwise.} \end{cases} \tag{6.20}$$

In what follows we will prove that this algorithm can attain a steady state for the coordination process among players. Also, the information it needs to operate is minimal.

6.5.1 Convergence of the Learning Process

Based on the theory of stochastic algorithms, the asymptotic behavior of (6.19) can be analyzed through the corresponding continuous dynamics [16]:

$$\frac{dx}{dt} = E(w|x) - x, \tag{6.21}$$

Algorithm 3 Distributed reinforcement Learning Algorithm

1: **input:** $k = 1$, each relay node i chooses its action (T or S) according to distribution p_i and set its initial perception value $x_i^0 = 0$.
2: **while** $max(|x_{iT}^{k+1} - x_{iT}^k|, |x_{iS}^{k+1} - x_{iS}^k|) > \epsilon$ **do**
3: Each relay node i updates its fully mixed strategy profile at iteration k according to (6.17).
4: Relay node i selects its actions using its updated fully mixed strategy profile.
5: Relay node i estimates its payoff $\bar{u}_i^k$.
6: Relay node i updates its perception value according to (6.19).
7: $k \leftarrow k + 1$
8: **end while**

where $x = (x_{ia}, \forall i \in \{1, .., N\}, a \in A)$ and $w = (w_{ia}, \forall i \in \{1, .., N\}, a \in A)$.

Let us make equation (6.21) more explicit by defining the mapping from the perceptions x to the expected payoff of user i choosing action a as $G_{ia}(x) = E(V^i|x, a_i = a)$.
Proposition 6.5.2. *The continuous dynamics (6.21) may be expressed as*

$$\frac{dx_{ia}}{dt} = \sigma_{ia}(G_{ia}(x) - x_{ia}) \tag{6.22}$$

Proof. Using the definition of the vector w, the expected value $E(w|x)$ can be computed by conditioning on player i's action as:

$$\begin{aligned} E(w_{ia}|x_{ia}) &= \pi_{ia}U(a, \pi_{-i}) + (1 - \pi_{ia})x_{ia} \\ &= \sigma_{ia}G_{ia}(x) + (1 - \sigma_{ia})x_{ia} \end{aligned} \tag{6.23}$$

which with (6.21) yields (6.22). □

This can be interpreted as follows: when the difference between the expected payoff and the perception value is large, the perception value, from (6.19), will be updated with a large expected value $w_{ia}^k - x_{ia}^k$ and this difference will be reduced.

In the following theorem, we prove that the learning process admits a contraction structure with a proper choice of the temperature β .

Theorem 6.5.2.1. *Under the logit decision rule (6.17), if the temperature satisfies* $\beta < \frac{1}{n_s r}$, *then the mapping from the perceptions to the expected payoffs* $G(x) = [G_{ia}(x), \forall i \in \{1, .., N\}, a \in A)]$ *is a maximum-norm contraction.*

Proof. Recall that $G_{ia}(x)$ is the expected payoff of relay node i choosing action a given the perceptions for all players x. Assume the chosen action is to be active (T), then $G_{iT}(x)$ can be written as:

$$G_{iT}(x) = \sum_{j=0}^{N} n_s r P_{succ}(T, j) C_j^N (\sigma_{iT}(x_i))^j (1 - \sigma_{iT}(x_i))^{N-j} - g\tau$$

Now consider the difference $G_{iT}(x_i) - G_{iT}(\hat{x}_i)$ given two arbitrary perceptions x_i and $\hat{x}_i$ of a relay node i :

$$\begin{aligned}
|G_{iT}(x_i) - G_{iT}(\hat{x}_i)| &= |\sigma_{iT}(x_i) \sum_{j=1}^{N-1} C_{j-1}^{N-1} (\sigma_{iT}(x_i))^{j-1} (1-\sigma_{iT}(x_i))^{N-j} U(T,j) \\
&\quad -\hat{\sigma}_{iT}(\hat{x}_i) \sum_{j=1}^{N-1} C_{j-1}^{N-1} (\hat{\sigma}_{iT}(\hat{x}_i))^{j-1} (1-\hat{\sigma}_{iT}(\hat{x}_i))^{N-j} U(T,j)| \\
&\leq |\sigma_{iT}(x_i) \sum_{j=0}^{N-1} n_s r \left(C_j^{N-1} (\sigma_{iT}(x_i))^{j} (1-\sigma_{iT}(x_i))^{N-j} \right) \\
&\quad -\hat{\sigma}_{iT}(\hat{x}_i) \sum_{j=0}^{N-1} n_s r \left(C_j^{N-1} (\hat{\sigma}_{iT}(\hat{x}_i))^{j} (1-\hat{\sigma}_{iT}(\hat{x}_i))^{N-j} \right) | \\
&\leq |\sigma_{iT}(x_i) n_s r - \hat{\sigma}_{iT}(\hat{x}_i) n_s r| \\
&\leq n_s r |\sigma_{iT}(x_i) - \hat{\sigma}_{iT}(\hat{x}_i)|
\end{aligned}$$

We know that $\sigma_{ia}(x_i)$ is continuously differentiable, then by the mean value theorem, there exists $\bar{x}_{ia} = \delta(x_{ia} - \hat{x_{ia}})$ with $0 < \delta < 1$ such that:

$$\begin{aligned}
\sigma_{iT}(x_i) - \hat{\sigma}_{iT}(\hat{x}_i) &= \frac{e^{\beta x_{iT}}}{\sum_{a\in A} e^{\beta x_{ia}}} - \frac{e^{\beta \hat{x}_{iT}}}{\sum_{a\in A} e^{\beta \hat{x}_{ia}}} \\
&= \beta \big[\frac{e^{\beta \bar{x}_{iT}} (\sum_{a\in A} e^{\beta \bar{x}_{ia}}) - e^{2\beta \bar{x}_{iT}}}{(\sum_{a\in A} e^{\beta x_{ia}})^2} (x_{iT} - \hat{x}_{iT}) - \sum_{a'\in A, a'\neq T} \beta \frac{e^{\beta \bar{x}_{ia'}} e^{\beta \bar{x}_{iT}}}{(\sum_{a\in A} e^{\beta x_{ia}})^2} (x_{ia'} - \hat{x}_{ia'}) \big] \\
&= \beta \big[C_T (x_{iT} - \hat{x}_{iT}) - \sum_{a'\in A, a'\neq T} \beta C_{a'} (x_{ia'} - \hat{x}_{ia'}) \big]
\end{aligned}$$

where $C_T = \frac{e^{\beta \bar{x}_{iT}} (\sum_{a\in A} e^{\beta \bar{x}_{ia}}) - e^{2\beta \bar{x}_{iT}}}{(\sum_{a\in A} e^{\beta x_{ia}})^2}$ and $C_{a'} = \frac{e^{\beta \bar{x}_{ia'}} e^{\beta \bar{x}_{iT}}}{(\sum_{a\in A} e^{\beta x_{ia}})^2}$. We can easily observe $C_T = \sum_{a'\in A, a'\neq a} C_{a'}$ and $2C_a \leq 1$. Then:

$$\begin{aligned}
|\sigma_{iT}(x_i) - \hat{\sigma}_{iT}(\hat{x}_i)| &\leq \beta C_T |x_{iT} - \hat{x}_{iT}| + \sum_{a'\in A, a'\neq T} \beta C_{a'} |x_{ia'} - \hat{x}_{ia'}| \\
&\leq \beta (C_T + \sum_{a'\in A, a'\neq T} C_{a'}) ||x_i - \hat{x}_i||_\infty \\
&\leq \beta ||x - \hat{x}||_\infty. \qquad (6.24)
\end{aligned}$$

Combining (6.24) and (6.24), we obtain

$$|G_{iT}(x) - G_{iT}(\hat{x})| \leq \beta n_s r ||x - \hat{x}||_\infty$$

We obtain the same result when player i chooses to be silent (S). Observing that since by the minority game rule $G_{iT}(\cdot) G_{iS}(\cdot) \leq 0$, then if $\beta < \frac{1}{n_s r}$, indeed $G(x)$ is a maximum-norm contraction. □

Based on the property of contraction mapping, there exists a fixed point x^* such that $G(x^*) = x^*$. In the following theorem we show that the distributed learning algorithm also converges to the same limit point x^*.

Theorem 6.5.2.2. *If $G(x)$ is a $||.||_\infty$-contraction, its unique fixed point x^* is a global attractor for the adaptive dynamics (6.22), and the learning process (6.19) converges almost surely towards x^*. Moreover the limit point x^* is globally asymptotically stable.*

Proof. Since $G(x)$ is a $||.||_\infty$-contraction, it admits a unique fixed point x^*. According to general results on stochastic algorithms the rest points of the continuous dynamic (6.22) are natural candidates to be limit point for the stochastic process (6.19). All together with ([16], corollary 6.6), we have the almost sure convergence of (6.19), given that we exhibit a strict Lyaponuv function ϕ.
Now let $\phi(x) = ||x_{ia} - x^*||_\infty$, then $\phi(x^*) = 0, \phi(x) > 0, \forall x \neq x^*$. Let $i \in \{1, ..., N\}, a \in A$ be such that $\phi(x) = |x_{ia} - x^*_{ia}|$. If $x_{ia} \geq x^*_{ia}$, then $\phi(x) = x_{ia} - x^*_{ia}$. Since $G_{ia}(x)$ is a maximum norm contraction, there exist a Lipschitz constant ξ such that $G_{ia}(x) - G_{ia}(x^*) \leq \xi(x_{ia} - x^*_{ia})$, and $G_{ia}(x^*) = x^*_{ia}$. All together combined with equation (6.22), we can write:

$$\begin{aligned}
\frac{d\phi(x)}{dt} &= \frac{d(x_{ia} - x^*_{ia})}{dt} = \frac{dx_{ia}}{dt} \\
&= \sigma_{ia}(G_{ia}(x) - x_{ia}) = \sigma_{ia}(G_{ia}(x) - G_{ia}(x^*) + x^*_{ia} - x_{ia}) \\
&\leq \sigma_{ia}\xi(x_{ia} - x^*_{ia}) + x^*_{ia} - x_{ia} = -(1 - \sigma_{ia}\xi)\phi(x) < 0, \forall x \neq x^*.
\end{aligned}$$

and a similar argument for the case $x_{ia} \leq x^*_{ia}$ also shows that $\frac{d\phi(x)}{dt} < 0, \forall x \neq x^*$. Thus the function $\phi(x)$ is a strict Lyaponuv function and x^* is globally asymptotically stable, hence the proof. □

6.5.3 Approximate Nash Equilibrium

From lemma (6.5.2.1) and theorem (6.5.2.2), we have:

$$G_{ia}(x^*) = E(V^i | x^*, a_i = a) = x^*_{ia}.$$

This is a property of the equilibrium (x^*) of the distributed learning algorithm: its value x^*_{ia} is an accurate estimation of the expected payoff in the equilibrium. Moreover we show that the fully mixed strategy

$$p^* = (\sigma^*_{ia} = \frac{e^{\beta x^*_{ia}}}{e^{\beta x^*_{iT}} + e^{\beta x^*_{iS}}}, \forall a \in A, i \in \{1...N\})$$

is an approximate Nash equilibrium.

Proposition 6.5.4. *Under the Logit decision rule (6.17), the fully mixed strategy $p^* = \sigma^*(x^*)$ at the equilibrium x^* is a ϵ-approximate Nash equilibrium for our game (proposition 6.3.7) with*

$$\epsilon = -\frac{1}{\beta}\sum_{a\in A}\sigma^*_{ia}(ln(\sigma^*_{ia}) - 1).$$

Proof. A well-known characterization of the logit probabilities gives:

$$\begin{aligned}\sigma_{ia}(x^*) &= \arg\max_{\sigma_i=[\sigma_{iT},\sigma_{iS}]}\sum_{a\in A}\sigma_{ia}E(V^i|x^*,a_i=a) - \frac{1}{\beta}\sum_{a\in A}\sigma_{ia}(ln(\sigma_{ia})-1)\\ &= \frac{e^{\beta E(V^i|x^*,a_i=a)}}{e^{\beta E(V^i|x^*,a_i=T)}+e^{\beta E(V^i|x^*,a_i=S)}} = \frac{e^{\beta x^*_{ia}}}{e^{\beta x^*_{iT}}+e^{\beta x^*_{iS}}},\end{aligned}$$

and since ([21], pp.93)

$$\max_{\sigma_i}\sum_{a\in A}\sigma_{ia}E(V^i|x^*,a_i=a) - \frac{1}{\beta}\sum_{a\in A}\sigma_{ia}(ln(\sigma_{ia})-1) \le \max_{\sigma_i}\sum_{a\in A}\sigma_{ia}E(V^i|x^*,a_i=a)$$

then, we have:

$$\sum_{a\in A}\sigma^*_{ia}E(V^i|x^*,a_i=a) \ge \max_{\sigma_i}\sum_{a\in A}\sigma_{ia}E(V^i|x^*,a_i=a) - \epsilon$$

where $\epsilon = \max_{i\in\{1\ldots N\}}\{-\frac{1}{\beta}\sum_{a\in A}\sigma_{ia}(ln(\sigma_{ia})-1)\}$.

Hence the fully mixed strategy $p^* = \sigma^*(x^*)$ in the equilibrium x^* is a ϵ-approximate Nash equilibrium. □

Observe that the parameter ϵ illustrates the effect of the temperature β. A larger ϵ (smaller β) means worse learning performance.

6.6 Application : Two-hops routing and exponential inter-contacts

In the previous sections we presented under a general context of DTN how a controlled minority game can be used to induce a stable cooperative behavior among the relays without actual cooperation. So far we assumed that the inter-contact time between nodes follows a random distribution and relay nodes can adopt any relaying policy.
In this section and for the numerical analysis, we will assume that relay nodes use the two hop routing scheme, in which any mobile that receives a copy of the packet from the source can only forward it to the destination. The time between subsequent contacts of a node with any other node in the network is now assumed to follow an exponential distribution with parameter $\lambda > 0$. The validity of this model for synthetic mobility models has also been discussed in [4]. In particular, regarding the rewarding policy adopted by the source nodes, we assume that upon successful delivery of a message, the relay node receives a positive reward R if and only if it is the first one to deliver the

message to the corresponding destination.
Under those assumptions, we can obtain the expressions of different quantities: in particular the probability that an active node relays a copy of a received packet to destination within time τ is $1 - Q_\tau$ where the expression of Q_τ is given by [11]:

$$Q_\tau = (1 + \lambda\tau)e^{-\lambda\tau}. \tag{6.25}$$

Now, the probability of successful delivery of the message for an active node is:

$$\begin{aligned} P_{succ}(T, N_T) &= (1 - Q_\tau) \sum_{k=1}^{N_T - 1} C_{k-1}^{N_T - 1} \frac{(1 - Q_\tau)^{k-1} Q_\tau^{N_T - k}}{k} \\ &= \frac{1 - Q_\tau^{N_T}}{N_T} \end{aligned} \tag{6.26}$$

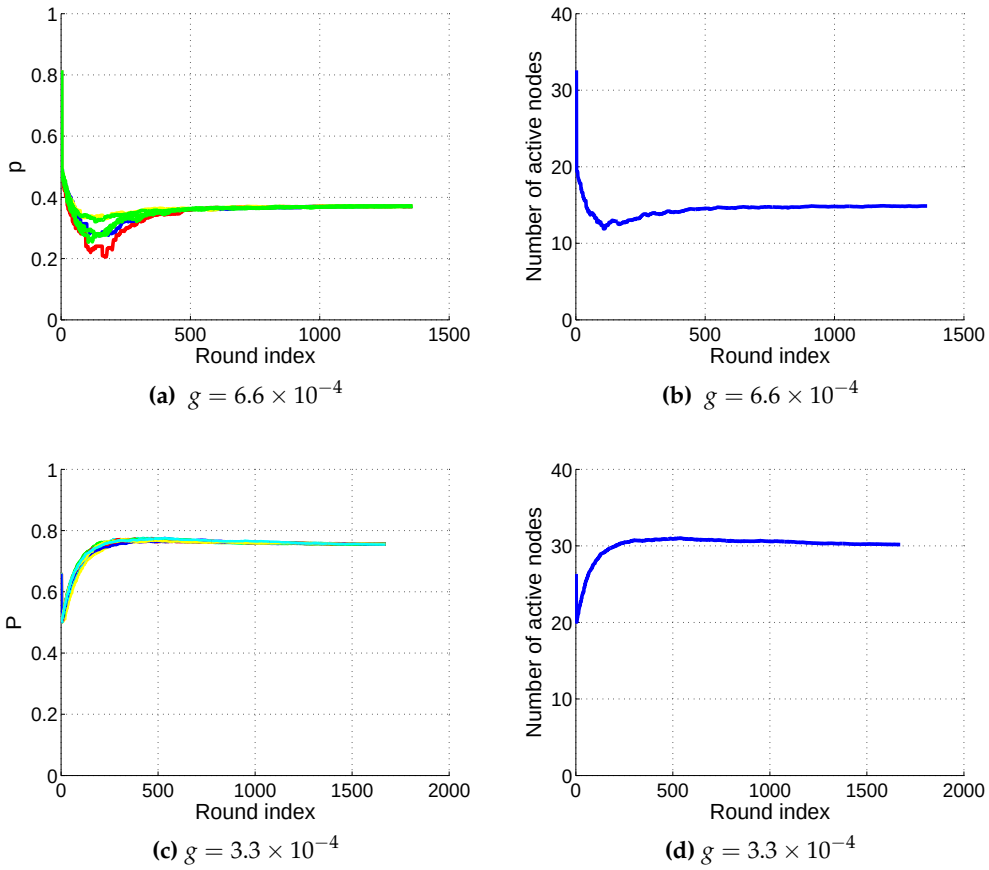
where $C_h^k = \binom{k}{h}$, such that each node seeks to be the first to deliver a given message to its destination.

6.7 Numerical Results

In this section, we provide a numerical analysis of the performance achieved by DTN nodes following the distributed reinforcement learning mechanism proposed in section 6.5. First, we focus on the achieved performance in a homogeneous network where all nodes have the same energy constraint (g). Second, we examine the performance of our algorithm in a multi-class framework (heterogeneous DTN), where we consider the existence of two classes of nodes. Then we will verify the intuitive result obtained in proposition (6.4.5) which states that by allocating smaller reward to a class, fewer nodes of this class will choose to be active. The results presented here take into account the utility functions defined in Scenario 1. The parameters $\lambda = 0.03, \tau = 100$ are used through out the numerical analysis.

Homogeneous DTN The performance of our learning algorithm in the homogeneous case is shown in Fig. 6.3. In this case we consider $g = 6.6 \times 10^{-4}, N = 40$. We set the sequence $\gamma^k = \frac{1}{k}$ for all iterations k, and the temperature $\beta \to \infty$, note that this choice of β is a good deal since it allows our algorithm to attain the Nash equilibrium (proposition (6.5.4)).

In Fig. 6.3(a) we observe that the probability to be active for a node i ($p_i, \forall i \in \{1...N\}$) converges to the symmetric equilibrium ($p^* = 0.35$) which is the solution of (6.6). Moreover, it is interesting to notice that the average number of active nodes at the equilibrium approaches the value of ($\Psi = 15$) where Ψ defines the comfort level of the minority game in pure strategy (Fig. 6.3(b)). Such behavior is, in fact, a convergence to the strictly mixed Nash equilibrium discussed in proposition (6.3.7). The same observation is recorded in Fig. 6.3(c,d) where a smaller energy consumption parameter



(a) $g = 6.6 \times 10^{-4}$ **(b)** $g = 6.6 \times 10^{-4}$

(c) $g = 3.3 \times 10^{-4}$ **(d)** $g = 3.3 \times 10^{-4}$

Figure 6.3: *Learning the mixed strategy: homogeneous case.*

$g = 3.3 \times 10^{-4}$ yields a larger activation rate which can be noticed in the convergence of p_i to the mixed Nash equilibrium ($p^* = 0.75$). As a result, there is in average more active nodes ($\Psi = 30$) at the equilibrium.

Heterogeneous DTN The performance of the learning algorithm in the heterogeneous DTN is investigated in two cases, symmetric (i.e. when $\frac{g_1}{r_1} = \frac{g_2}{r_2}$) and asymmetric ($\frac{g_1}{r_1} \neq \frac{g_2}{r_2}$). We consider first the symmetric case. We consider $g_1 = 0.8 \times 10^{-4}, g_2 = 0.5 \times 10^{-4}, N_1 = 20, N_2 = 20$ then setting $r2 = 0.15$ we obtain $r_1 = 0.24$. In Fig. (6.4)(a) we observe that the probability of being active of nodes of both classes (p_1, p_2) converges to the symmetric Nash equilibrium discussed in proposition (6.4.5), and the value it converges to ($p_1^* = p_2^* = 0.78$) is the solution of the equation ($A_1(N, p_1^*, p_2^*) = A_2(N, p_1^*, p_2^*) = 0$) as shown in Fig(6.2). The average number of active nodes, depicted in Fig (6.4)(b), converges to $\Psi = 30$ that satisfies the relation (6.10).

In Fig(6.5), we depict the asymmetric case, when $g_1 > g_2$ and $r_1 < \frac{g_1 r_2}{g_2}$. In

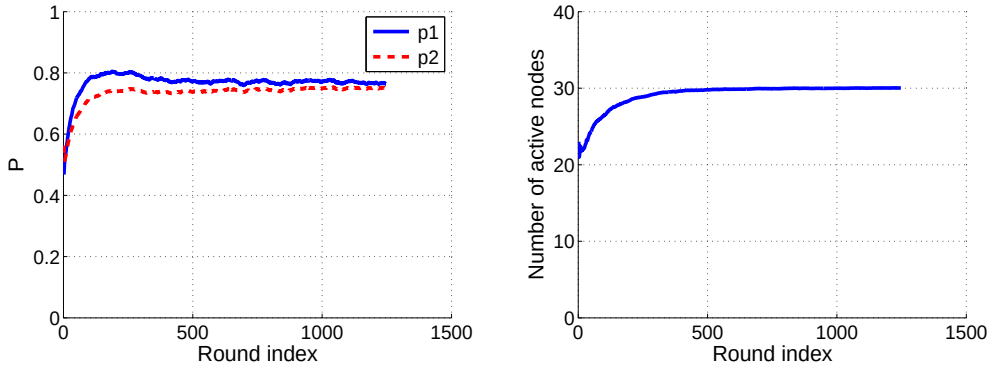


Figure 6.4: *Learning the mixed strategy: heterogeneous symmetric case, where:* $g1 = 0.8 \times 10^{-4}, g2 = 0.5 \times 10^{-4}, r2 = 0.15$.

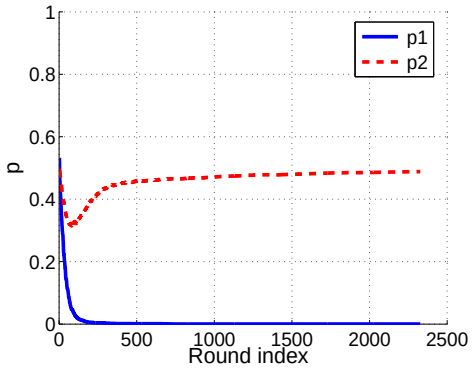
Fig(6.5)(a,c,e) we observe that ($p_2 > p_1$), in other words, the nodes with high energy constraint (class 1)are less active, thus by allocating smaller reward (r_1), fewer nodes of class 1 are active. Notice in Fig(6.5)(b,d,f) that the average number of active nodes $\Psi_1 \leq N_T < \Psi_2$.

6.8 Conclusions

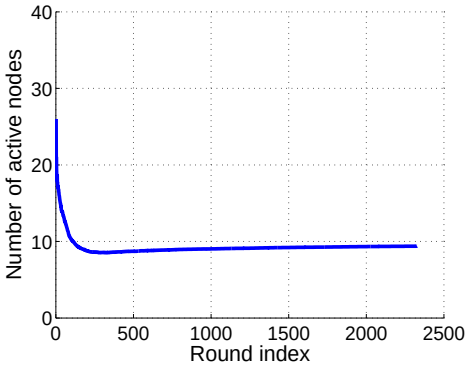
Coordination of mobiles which are part of a DTN is a difficult task due to lack of permanent connectivity. Operations in DTNs, in fact, do not support the usage of timely feedback to enforce cooperative schemes which may be implemented on mobile nodes. Nevertheless, coordination is worth indeed in order to attain efficient usage of resources. Moreover, selfish behavior and activation control becomes core when owners of relay devices may need incentive to spend memory and battery.

To this respect, our approach provides a novel mechanism designed using the theory of Minority Games (MGs). MGs are non-cooperative games which apply to contexts where the payoff of players decreases with the number of those who compete. We could design a reward mechanism for two hop routing protocols that runs fully distributed and with no need for any dedicated coordination protocol. I.e., the source controls how many nodes to activate in order to attain a target message delivery probability. It does so by setting the reward for nodes who deliver first and such in a way to avoid over provisioning of activated relays. Finally, we developed a distributed stochastic learning algorithm able to converge to the optimal solution.

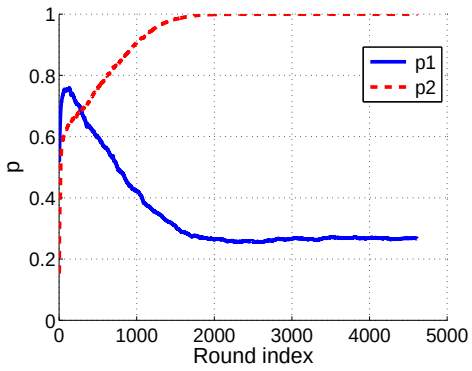
Future works will investigate how to extend the models and the properties of convergence of our algorithm to other types of networks such as cognitive radios and peer-to-peer networks.



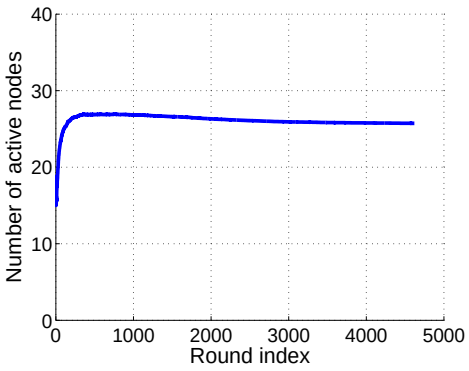
(a) $r_1 = 0.056, r_2 = 0.05, \Psi_1 = 7, \Psi_2 = 10.$



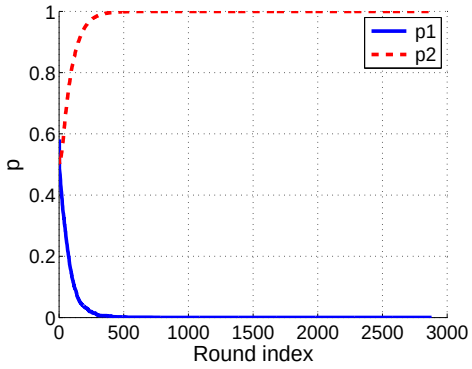
(b) $r_1 = 0.056, r_2 = 0.05, \Psi_1 = 7, \Psi_2 = 10.$



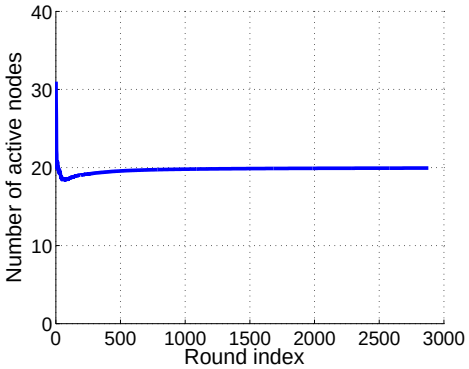
(c) $r_1 = 0.2, r_2 = 0.14, \Psi_1 = 26, \Psi_2 = 28.$



(d) $r_1 = 0.2, r_2 = 0.14, \Psi_1 = 26, \Psi_2 = 28.$



(e) $r_1 = 0.12, r_2 = 0.125, \Psi_1 = 15, \Psi_2 = 25.$



(f) $r_1 = 0.12, r_2 = 0.125, \Psi_1 = 15, \Psi_2 = 25.$

Figure 6.5: *Learning the mixed strategy: heterogeneous asymmetric case, where:* $g1 = 0.8 \times 10^{-4}, g2 = 0.5 \times 10^{-4}$.

Part IV

Implementations and design of experimental platforms

Chapter 7

Evolutions and integrations in matlab LTE simulator

7.1 Introduction

This chapter is dedicated to the development of a simulation platform. In networking research, it is a common issue to provide reliable and repeatable experimentations for the purpose of validations or explorations. On one hand physical experiments on devices are liable to provide realistic measurement data, given some measurement errors and bias depending on the environment. Besides it is not always possible to afford having the appropriate equipment for every single test and even when it is possible attrition of repetition is yet another limitation. A regular way of proceeding is then to first rely on simulators, which will support tests without any risk of damage errors and/or severe crashes, this before implementations on board of real devices. However, simulators are not necessarily standardized and there always exists a gap between the results outputted by a simulator and results of same experiments that would be obtained on real devices. It is thus necessary, for the validation of the developed approaches throughout a research project to use a reliable and well-constructed simulator rather than self-developed codes. A reliable simulator needs however to be inspired from standards and specifications documents. There are several such simulators in the literature (for example: NS-2 or 3, Omnet++, etc) that are usually employed to simulate realistic experiments in a networking area. The most popular simulators like, NS 2 or even Omnet ++ however do not implement all the necessary features to simulate for example a system level scenario for the latest mobile communication standards (in our case LTE). This fact has motivated the development of new simulators and the evolution of existing ones(for example: Although NS-3 is the third version of the NS-2, it presents totally differently and bases on C++ object oriented logic rather than TCL scenario description codes in NS-2).

In the scope of the ECOSCELLS project, which includes the major part of the work achieved in this thesis, it has been question for validation purpose, to select a simu-

lator. Indeed, as we mentioned before, simulation is a crucial point in validation of algorithms or heuristics which are generally based on constrained mathematical models. We were interested in validation of different algorithms developed by the partners in the project, that will operate at the system level. The objective of system level simulations is to allow for the analysis of several aspects of network performances observed from the global functioning of the network. For its flexibility and for the specific context of the ECOSCELLS project, we selected the matlab LTE simulator from the University of Vienna. The simulator was chosen for different other reasons and in conjunction with the partners. This simulator thoroughly developed by a network team from university of Vienna, directed by Josep Colom Ikuno, Martin Wrulich and Markus Rupp, yet does not include all the necessary features needed for integration and simulation of, for example, FFR, ICIC or MRO algorithms. Our approach has then been to develop, adapt and integrate new features in the simulator in order to support simulations of such algorithms in a realistic environment.

7.1.1 Main contributions

Throughout this chapter we will present the simulator with its building blocks and highlight its lacks in necessary features. We will mention our development procedure and the contributions brought to the enhancement of the simulator. As contributor to algorithm design and integration in the scope of the ECOSCELLS project, we put in front the different analysis made to the simulator and present a diagram of implementation. The rest of the chapter is organized as follow. In section 7.2 we precise our specific context of implementations, we define the concept of system level simulations and give a description of the simulator with its functional analysis. In section 7.3 we develop the different integrations and development to to be done on the system level simulator for its evolution with new features. Section 7.4 presents the simulation and experimentation settings and also decline the results obtained through simulations. Section 7.5 eventually concludes this chapter.

7.2 LTE System level simulator from university of Vienna

In this section we detail the structure, composition and goals of the simulator. First of all we define what we understand by system level simulations.

7.2.1 System level simulations (concerns and metrics)

The system level simulations in the mobile wireless network are dedicated to observation of phenomenon such as scheduling, interference, mobility and propagation. This includes the evaluation of the performances of the network entities at the layers 2 and 3 in addition to the environment of propagation. At the UE, some pre-simulated outputs map from the mac layer are used to make the correspondence between variables

such as channel conditions (some propagation models are generally used to simulate the channel conditions ex: Okumura model, the COST-231 Walfish-Ikegami model,...), mobility and the obtained BLER. Referring to the LTE stack architecture (figure 7.1) the operations are concerned mainly with the RLC layer at layer2. At the eNodeBs, the physical layer is also abstracted by simplified models that capture its essential characteristics. The concerned mechanisms are essentially Inter cell RRM(for example FFR algorithms operating over the X2 interface), Resource Bearer Control, CMC, RAC, Measurements and scheduling on top of RRC mechanisms working with simulated outputs of the RLC layer. The LTE simulator from university of Vienna develops all the aforementioned structure that we detail in the following.

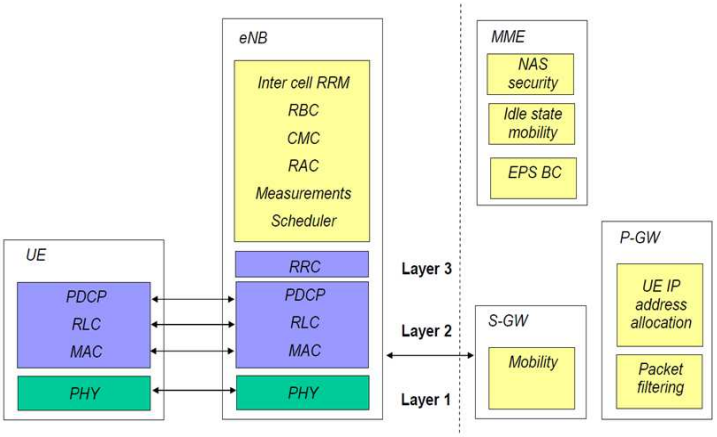


Figure 7.1: *LTE Stack architecture and system level entities*

7.2.2 Description of the simulator

As previously announced, the purpose of the LTE system layer simulator is to observe effects of issues such as cell planning, scheduling, or interference in the cellular network, in our case the LTE network. A run of simulation operates over a Region Of Interest (ROI) in which the eNodeBs and UEs are positioned for a simulation length in Transmission Time Intervals (TTIs) which is defined at the beginning of the simulation. The simulator operates in three distinct phases:

- Input parameter description.
- Initializations, loading of physical layer traces and main loop.
- Traces saving and outputs generation.

Those different phases of the simulator operation are represented on figure 7.2. Each phase is also represented by a given matlab script in the execution hierarchy:

- **LTE_sim_launcher.m** : This is a batch file that allows loading a configuration file of choice of the form "LTE_load_params_xxx.m" where 'xxx' represents the specific scenario to be launched. A call to the execution of the LTE_sim_main.m main simulation file is also done.
- **LTE_sim_main.m** : It is the main simulation file that contains the initialization and the main loop of the simulator.
- **LTE_load_params.m** : Holds the list of the different parameters that can be configured depending on the scenario.

If we go back to the building block structure, the simulator is composed of two main building blocks, namely the link measurement model and link performance model. A detailed picture of the composition of each building block can be found in [55]. Essentially, the link measurement model defines the network layout that can be either generated or loaded from some network planning tool and the mobility management model which influences the interference structure. The traffic model is also defined here as well as the resource scheduling strategy adopted at the eNodeBs. Power allocation strategy on resource blocks, propagation models and precoding scheme are set in this block. On the other hand, the link performance model, which is subsequent to the link measurement model, includes features such as link adaptation strategy (CQIs, SINR and MCS mapping) and SINR averaging methods (EESM or MIESM). The execution flow of the simulator is presented in figure 7.3 where $\rightarrow$ represents the data flow in and out of the simulator's link abstraction model. For the scope of the ECOSCELLS project some necessary features are thus already implemented allowing essentially to simulate multi-cell, multi users scenarios, frequency reuse mechanisms light mobility and handover procedure, and scheduling.

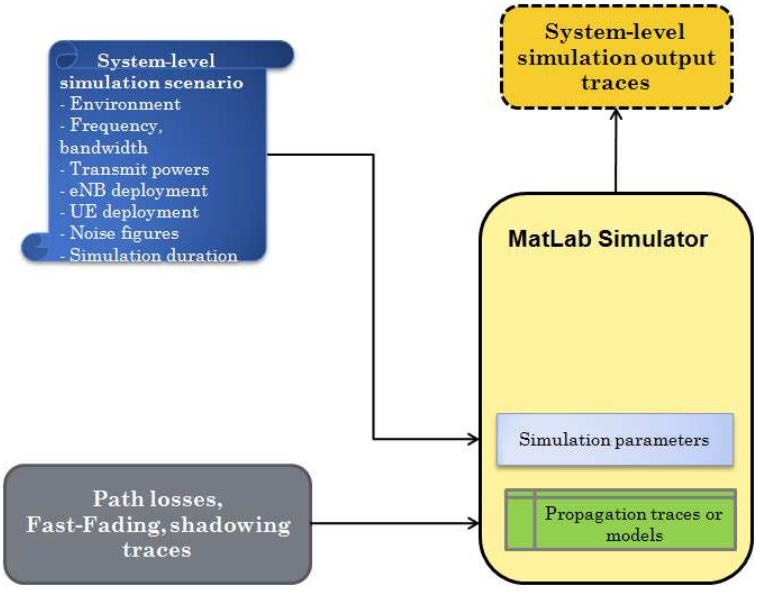


Figure 7.2: *Operational structure and organization of the simulator*

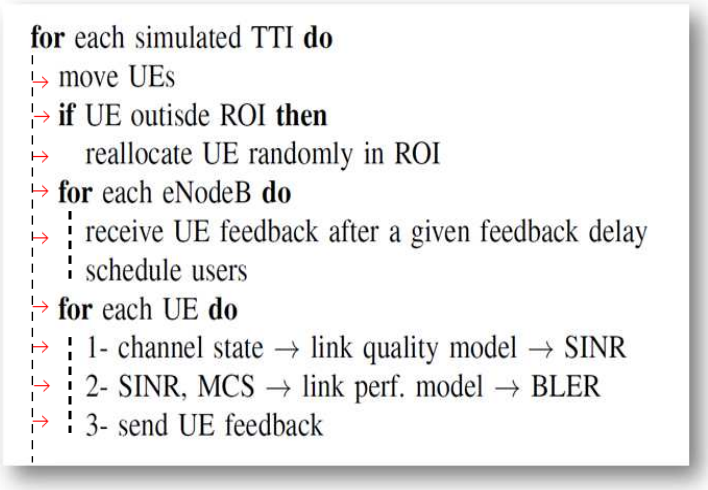


Figure 7.3: *Execution flow of the main loop of the simulator*

7.2.3 Functional analysis

For our implementations purpose it was necessary to identify the lacking features for the algorithms that are going to be implemented. A thorough functional analysis of the simulator has allowed to identify the absence of essentially the X2 interface, dynamic power allocation, minimum bit rate requirement for QoS provision, exchange of specific control messages between UEs and eNodeBs instead of only classical feedbacks(PMI, CQIs, Layers, quantized SINR...), channel models with non-hexagonal structure and complete implementation of handover procedures. Particularly as contributor to the project we focused on the following implementations :

- Specific exchanges in the uplink : This new feature is dedicated to additional information to be forwarded to the eNodeB by the UE; it includes exchange of specific neighbors interferences and received signal power on allocated resources blocks.
- Generation of neighbors lists on non-hexagonal network structure : The original implementation of the simulator allows only simulation of an hexagonal network. Moreover, only the seven closest neighbors are illegible to appear in the node neighbor list. Here we added some flexibility to this feature using a pathloss respective neighbor list. A maximum number of neighbors have also been set to avoid situations of handover to unknown neighbors.
- Integration of X2 API developed by Bell Labs Alcatel lucent : We have integrated in the main loop some functions of the X2 interface implemented by partners from Bell Labs Alcatel lucent.
- Definition of new private X2 message for ICIC : An additional ICIC specific X2 private message was created for utilization in integrated algorithms.

7.3 Development and integrations

Recall that the objective of using the LTE system level simulator is to simulate and evaluate the performances of our algorithms. Specifically we have focused of the implementation of the Bi-levels optimization algorithm for self-organization in the small-cells network. The algorithm is distributed and aims at performing fractional frequency reuse for small-cells networks (e.g. LTE) including users association. The proposed approach is therefore distributed and allows the bases stations, using a light collaboration, to achieve an efficient utilization of the frequencies, with the optic of maximizing the total system utility. The model includes both downlink and uplink transmissions and assume a resource allocation scheme with a single user per RB. The algorithm is presented in chapter 2 and is based, at the higher layer on the gradient descent algorithm and at the second layer on a learning algorithm namely the pursuit algorithm. In figure 7.4 we depicted the exchange flow of information which have motivated the implementation of additional features in the simulator. The exchange of information works as follow :

(1) The UE collects interference informations from each neighbors over every allocated RBs and forward the obtained vector to its serving eNodeB.

(2) Once information is received at the eNodeB from all the attached users, an interference matrix is built that contains interferences information from each neighbors (M neighbors on the figure).

(3) Interferences load are then put on format for each neighbors

(4) Finally, each neighbor receives on the X2 interface the interference informations concerning all the covered mobiles.

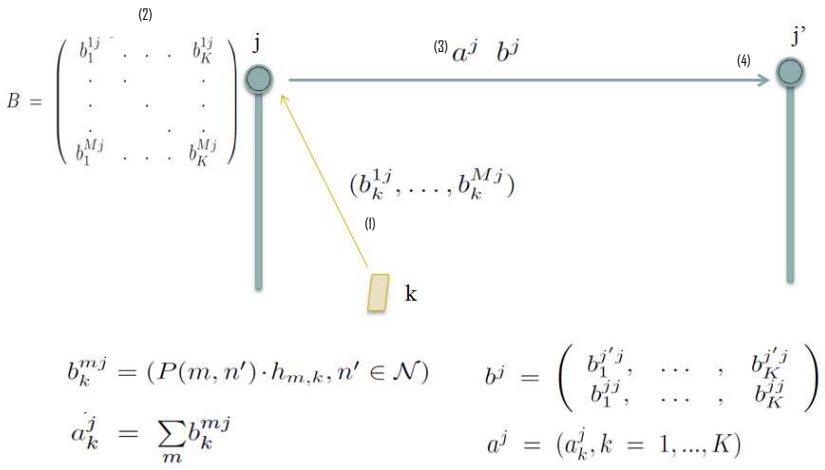


Figure 7.4: *Exchange of information sequence*

The integration of our algorithm holds in two different parts. We design in a first part a new scheduler based on the existing *Round Robin*(RR) scheduler of the simulator. The new scheduler works in the following steps :

1- Collect users feedbacks for each of the "*k*" RBs: using the new feedbacks structure.

2- Make the interference data on format as a matrix : *Double[nrNeighbors][k] B*, for future exchange with neighbors.

3- Share local ICIC data on the X2 channel with interfering neighbors, *Double[k] a and Double[2][k] b*.

4- Receive ICIC data from interfering neighbors. Here, since users are handled by the eNodeBs sectors and there exists an X2 interface only between eNodeBs, a new procedure was initiated at the integration of X2 API that allows to collect received ICIC messages at eNodeBs and forward to the destination sector.

5- Express the next powers vector using the gradient descent algorithm. We have defined the following function :

function Vector<double> P = dynamicPowerOpti (Vector<double> currentP, Double[nrNeighbors][k] a', Double[nrNeighbors][2][k] b')

6- Make optimized power assignation

7- Perform RBs allocation using RR scheduling functions.

The second stage of our integrations consists in the implementation of the pursuit Algorithm inside the simulator. In figure 7.5 we present the work flow diagram of the algorithm also described in chapter 2.

The implementation is done in the following steps :

1- After calculation of the current SINR and expression of the corresponding utilities, UEs must update their association strategies

2- The following function is then triggered:

function [int eNodeB_id, integer sub-band_id] = dynamicAssocOpti(Vector<Integer> neighbor_eNodeBs, double current_SINR)

The function allows to select the best eNodeB among neighbors, iteratively, with a common update process with other mobiles.

3- Trigger a handover request to the new access selected.

7.4 Simulations results

Next to the different integrations, we follow on with the simulation of our scheme in a nearly realistic environment. We use the new version of the simulator, including the

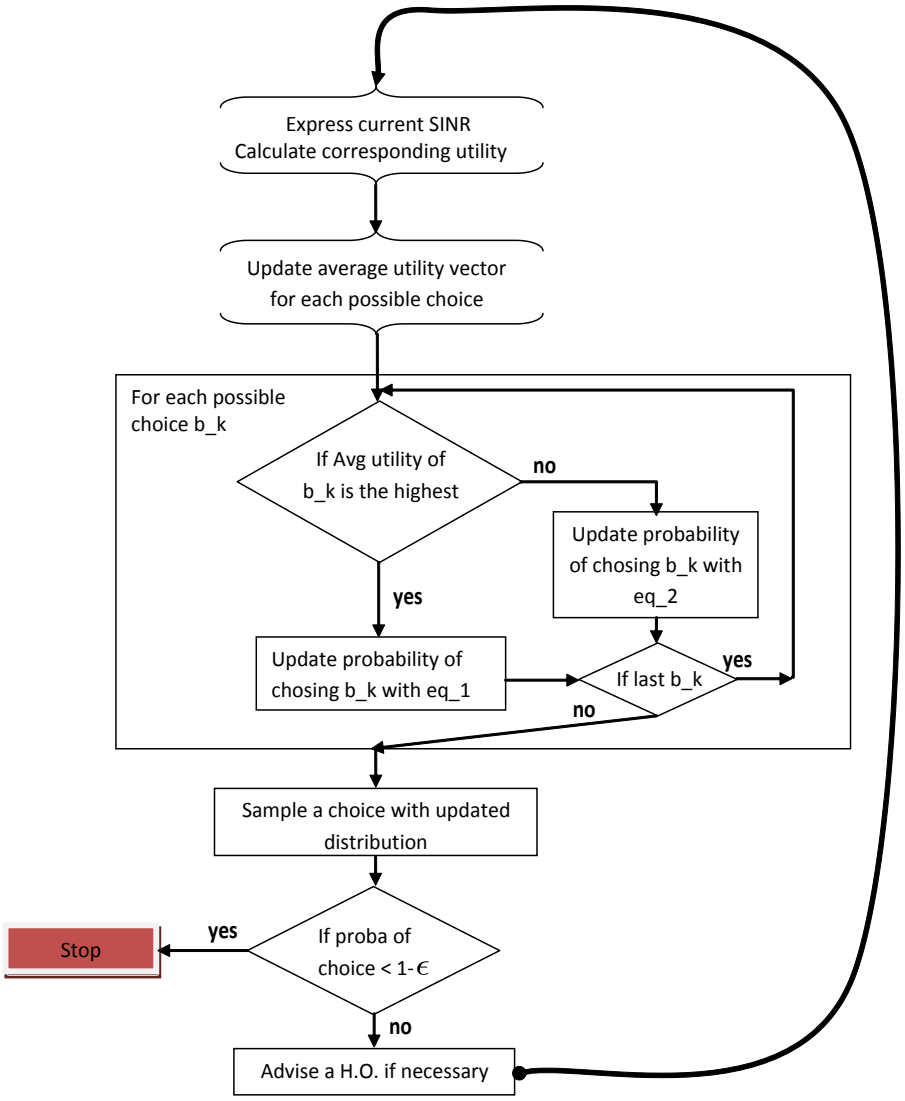


Figure 7.5: *Execution flow of the pursuit algorithm*

different implementations and integrations brought by the different partners and ourselves. Note that the simulator is still in progress of development and at this stage we cannot include all the settings of the simulation due to privacy terms. The simulations here will then just compare the results obtained after simulation of our approach with results obtained using the former implementation of round robin scheduling and homogeneous power allocation in the simulator. The simulations framework includes pedestrian straightforward mobility model, and the propagation environment is a pregenerated volcano model from partners in the project(SIRADEL). Other settings are :

- System Frequency = 2.14Ghz
- Operating Bandwidth = 10Mhz

- UEs are placed over the whole ROI (density is divided however by 5)
- Target sector: Center sector
- Number of TXs = 2
- Number of RXs = 2
- Transmission Mode = TxD(MIMO)
- Max Number of sector Neighbors = 21
- Simulation duration = 100 TTIs
- Latency time scale = 25 TTIs (used to calculate the average throughput)
- Network source = volcano (loaded)
- Shadow fading = not considered
- Micro-scale fading = volcano
- Antenna gain pattern = TS36.942
- Scheduler = round robin or stochastic gradient

By lack of time for full integration, the results presented in the following do not include actual execution of handovers but the handovers advices are counted. In figure 7.6 we plot the mean BLER by RB for several eNodeBs and small cells of the network. If we compare with figure 7.7 we can see that the power control applied at the scheduler doesn't affect ostensibly the resource allocation policy and the RBs allocation patterns are very similar for both figures. This sets a common basis for comparison of obtained performances. Note that the fluctuations of the BLER are essentially due to the randomization of the wireless channel. In figure 7.8, we can see that for several eNodeBs the dynamic power allocation applies effectively on the allocated resource block. Compared to figure 7.9 where all the eNodeBs use homogeneous power allocation policy over all the resources blocks, the power levels are set accordingly to reduce the level of inter-cell interferences. Indeed, as we can notice from the figure 7.8 only small cells bases stations that are more exposed to the inter-cell interferences phenomenon have their power allocation vectors modified over time while homogeneous allocation is optimal at the Macro eNodeB. On the basis of the convergence properties exposed in chapter 2, we expect the power allocation vectors to stabilize over time. From figures 7.10 and 7.11 one can notice that the average data traffic on each site remain the same with however a deep degradation of transmission in figure 7.11 that can be imputed to stochasticity of the optimization method. Since users traffic model is full buffer, and the allocated resource blocks are similar for both scenarios, we should expect the amount of transmitted data to also be similar. Eventually we plot in figure 7.13 the throughput obtained by each user using our ICIC mechanism, we also plot the average throughput that easily compare with the average throughput obtained by users without any optimization and homogeneous power allocation in figure 7.12. Both approach seems to achieve the same average level of throughput for user in the ROI at the last iteration despite some bounces for our iterative approach. However we can notice that the capacity

is more fairly distributed to some users that receive some more capacity when applying our ICIC scheme. The gap between the two simulated scenarios can also become larger with the number of iterations. The statistics maintained on the number of handovers allows to highlight the adaptivity of our algorithm which reduces the number of advised HO procedure from 100% to less than 10% within a few iterations. Although some users are still incited to proceed with handovers, we expect those remaining users to stabilize with iterations. However since users are not actually tacking the handovers advises, this metric only give some incite but does not really reflect the performances of our approach.

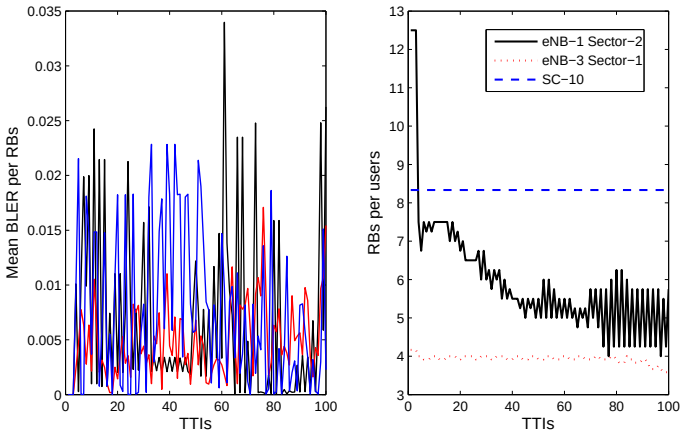


Figure 7.6: *Schedulers traces over TTIs for several BSs using the classical round robin scheduler*

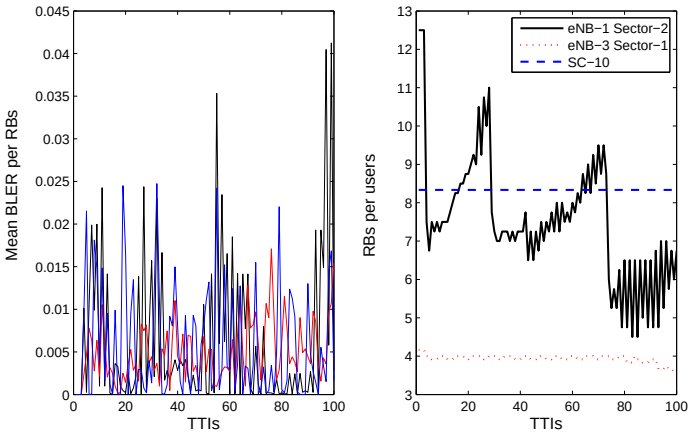


Figure 7.7: *Schedulers traces over TTIs for several BSs using the stochastic gradient scheduler*

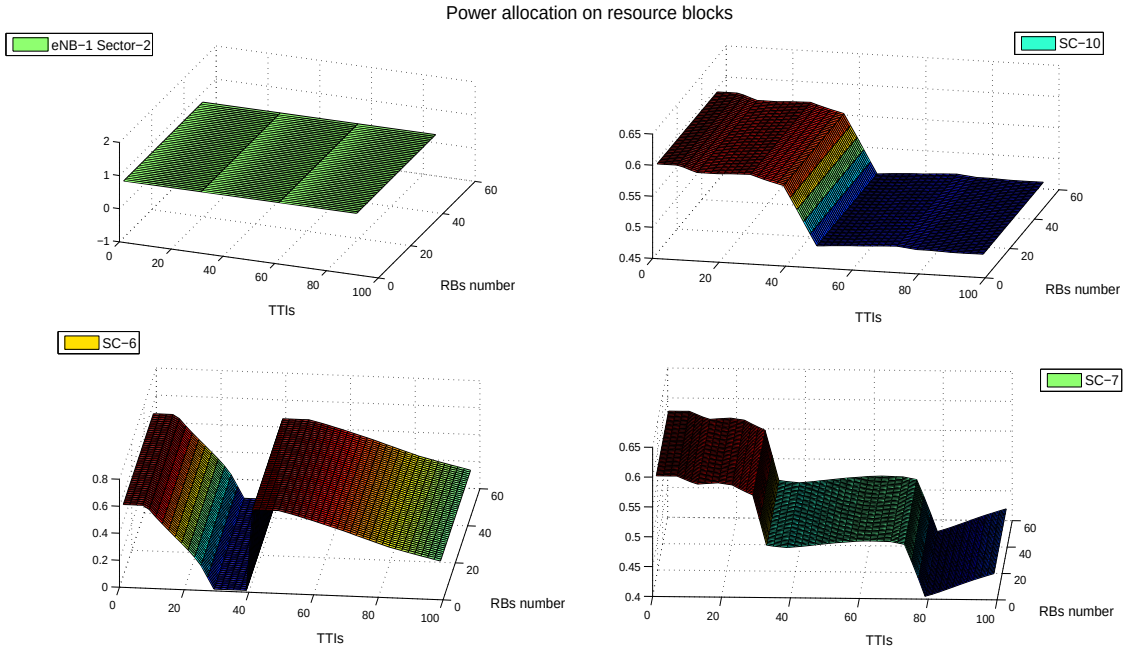


Figure 7.8: *Power allocation vector over TTIs obtained using application of the stochastic gradient scheduler*

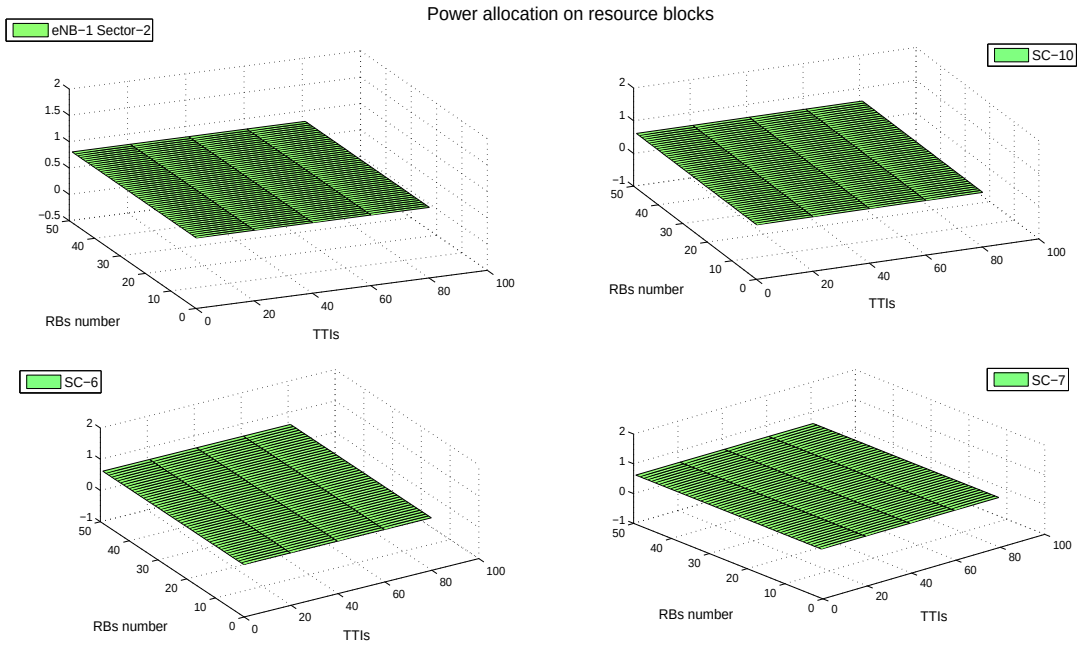


Figure 7.9: *Power allocation vector over TTIs obtained using application of the round robin scheduler*

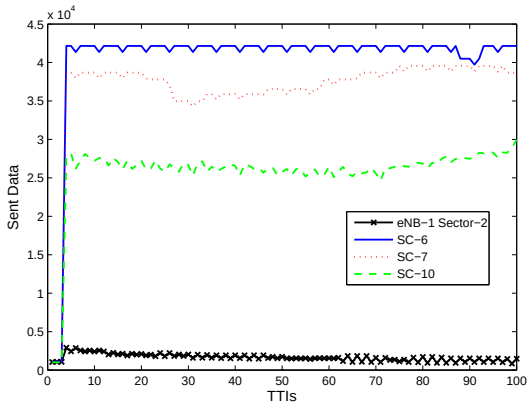


Figure 7.10: *Proportion of data transmitted by several eNodeBs over TTIs when using RR scheduler*

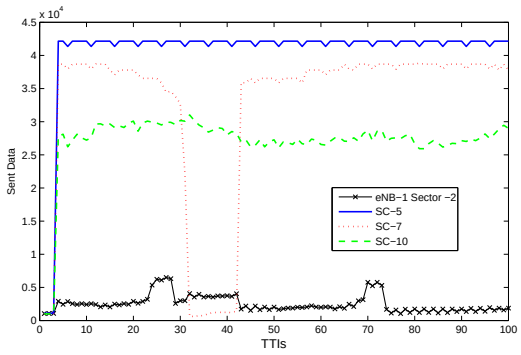


Figure 7.11: *Proportion of data transmitted by several eNodeBs over TTIs when using the stochastic gradient scheduler*

7.5 Conclusion

In this chapter we have presented different aspects of integration and evolution of a simulation plateform for the LTE network. We based on a public version for academic use of the LTE system level simulator from university of Vienna. We have presented the different evolutions brought to the simulator in order to integrate and simulate the self-organizing scheme proposed in the firsts chapters of the thesis. The list of evolutions have been presented as well as our integration policy of the proposed scheme. Eventually we have tried to compare using the classical Round Robin scheduler as a reference case, the performances of the algorithms in a particular mobile network scenario. This has allowed to confirm the announced performances of the bi-level self-organization algorithm. However, since all the integration are not effectively implemented in the simulator, the results obtained in this chapter are not exhaustive and full integration of the algorithm will effectively present the out performances of our approach on existing FFR mechanisms.

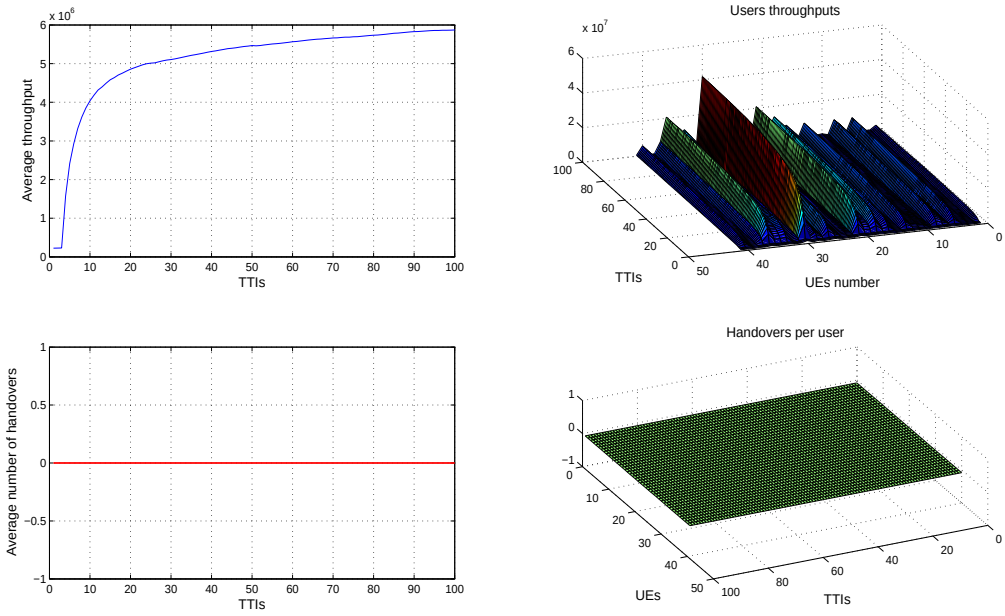


Figure 7.12: *Per user and average throughput and number of HO over TTIs obtained using the round robin scheduler*

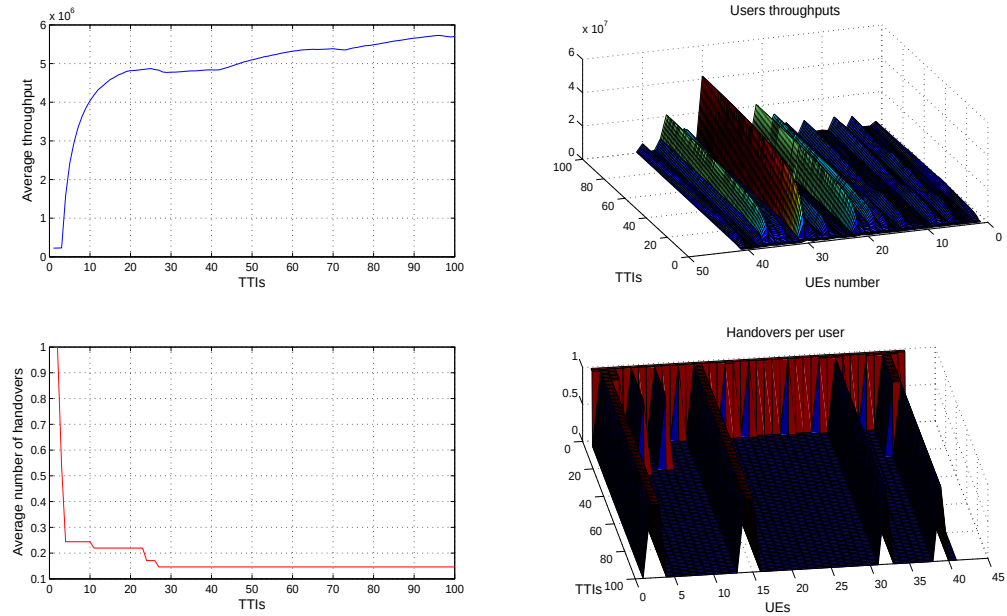


Figure 7.13: *Per user and average throughput and number of HO over TTIs obtained using the stochastic gradient scheduler*

Conclusion

Summary and general discussion

In this thesis we have focused on application of Game theoretical tools such as learning mechanisms, especially reinforcement learning algorithms to address system's performance evaluation in communication networks. With the growing density of current networks, composed of several technologies and an increasing diversity of terminals at the end users, a centralized approach of the network management has become particularly difficult. On one hand thanks to the improvements of modulation and coding techniques for transmission, as well as radio access methods, the new technologies of communications are now capable of offering very high bit rates to the end users. Several technologies with considerable capacities have therefore emerged from telecommunication networks as well as computer science networks and are now merged inside the overall structure of the new Next Generation Networks. On another hand, the improvement of processing capabilities of the user equipment has motivated design of non-hierarchical networks where the mobile terminal plays a central role in distributed and ad-hoc networks using proximity communications. This fuliginous progress in the system capacity obeys to the well known law of Moore who predicted a period often quoted as "18 months" for a doubling in electronic systems performances. The scalability of the network, due to the popularization of modern communications, has also incited the research community in finding some new ways of addressing the management of the network. Communications networks have thus become distributed by nature and have lead to exploration of decentralized ways of conceiving the network management.

Our approach in this thesis has been to take advantage of this distributed architecture of the network environment to introduce some level of autonomy, self-configuration and self-organization at the network agents. In a first place, we have addressed the problem of user-network association in which mobile users of the network have to select among a pool of access nodes, the one providing the best conditions of transmission, given that other nodes seek for the same outcome. The problem has been modeled as a non-cooperative game where the players are the mobile users and for which we have defined the possible Nash equilibria. The notion of Nash equilibrium was very important to us since it defines a stable state of the system in which no player can improve its utility by unilateral deviation from the equilibrium. Different type of utility functions have been modeled and we have shown that some fully distributed reinforcement learning algorithms can be implemented on board of the mobile agent to achieve the Nash equilibrium yielding better capacities for the user and thus improving the network performances. In another work, we have combined access selection procedures with distributed optimization in a framework of mobile communication technology namely LTE-advanced networks, to address the inter-cells interferences cancellation(ICIC) issue. We used the notion of stackelberg game and defined a bi-levels algorithm performing joint dynamic power control for FFR using stochastic power control, and attachment selection by means of a powerful distributed reinforcement learning mechanism. Here also the approach was proved to be conclusive and relatively efficient depending on appropriate time-scales selection and effectiveness of exchanges of

data. Moving from hierarchical to non-hierarchical networks, the same technique of distributed reinforcement learning mechanism has been employed to address a problem of routing configuration selection in DTNs. Indeed the control of the routing policy is a crucial matter for performance improvement in DTNs in order to establish a tradeoff between energy expenditure and high successful delivery probability. In this type of networks, the more the mobile relays are active, the higher the delivery probability but indeed the more the energy cost for delivery is increased. By defining a specific rewarding mechanism for successful delivery, we set an incentive policy allowing the mobile agents to select the routing configuration that maximizes their personal reward in trade of energy expenditure. Again our objective was to find a stable configuration of the network in the form of Nash equilibrium. The computation of the Price of Anarchy allowed us to compute the maximal gap to the global optimum solution of our problem. In DTNs the scalability of the population of users of the networks while being a major strength for an increased successful delivery probability, however limits the impact of a single user on the outcome of the overall population. It is then difficult to stick to the only concept of Nash equilibrium. Thus we have extended our analysis to the notion of evolutionary stable equiliria which defines an evolutionary stable strategy (ESS) profile of the population. We used the notion of ESS to study how a control of message forwarding at the source node in the DTN can help to improve successful delivery probability while satisfying a constraint on the total energy spent. Under this framework of large population of competing relays, we analyzed the activation control problem which defines the strategy by which a relay in the DTN can decide to participate or not to message relaying. We have computed the ESS (population profile) for our activation control game and study the influence of some systems parameters on the ESS as well as its behavior depending on the source forwarding control. So far the consideration of energy constraint has focused on the number of relays involved in the relaying but does include energetic level of relays and its evolution in time. To consider the battery state in the decisions of relays, we use the concept of MDEG(Markov Decision Evolutionary Games) for which the decision of mobile depends on the energy state in which their are currently operating. This adds some level of complexity to our model of activation control game while considering several assumptions of a more realistic environment. In the same context of DTNs, we focused on a new model for coordination without cooperation in DTNs. This model bases on the concept of Minority games for which we computed different type of pure and mixed Nash equilibria. We have presented a two scale approach allowing the operator of the network to tune the rest point of the defined competition in order to insure a QoS level or efficiency of the network. We ended up by devising a distributed reinforcement learning algorithm using a logit rule based updating mechanism. The performance evaluation of our approach by numerical investigations has shown fast convergence and stability properties. As the major parts of our work include testing and validations by simulations, we have been involved during our thesis in several implementations and integrations of simulation platforms. The last chapter of the thesis has then been dedicated to the different integrations and development brought to a system level simulator for mobile networks. We have presented our different contributions and the performances obtained in result of the different integrations. Overall, we are convinced that diversity and auto reconfigu-

rations can be exploited in the heterogeneous networks to design new architectures of cooperation between the agents of the network, where autonomous learning agents can interact to improve the overall and/or individual capacity of users of the network. The application of the different distributed learning techniques developed throughout this thesis are then liable to be applied in many other scenarios for network performance optimization.

Some perspectives

Several directions are available in the following of this work in the area of wireless communication networks.

Learning and hierarchical learning : Throughout our work we used some learning algorithms for reinforcement learning, mainly classical L_{R-I} mechanism by Sastry, and the pursuit algorithm from the same authors, for convergence to pure Nash equilibrium. However, other reinforcement learning algorithms can be used in the same framework developed in this thesis such as fictitious play, Q-learning, no-regret learning and other reinforcement mechanisms. As exposed in the theoretical background section of the introduction, each learning algorithm can be used to match a specific network scenario. For example, when the pure Nash equilibrium does not exist and there always exists a mixed Nash equilibrium, the algorithm as it is converges to one of the absorbing state, i.e., unit vector. Yiping et all [129], proposed a new algorithm, called the linear reward-penalty algorithm, which is useful especially when a pure Nash equilibrium does not exist and convergence to mixed Nash Equilibrium can be achieved. Such algorithm is usually desirable when users consider load sharing between several servers in case of multi-homing for example. The reinforcement learning algorithms are also generally constrained from different aspects such as : appropriate parameters calibrations(step size...), number of state observations(switching from states to states) or accuracy of the received feed-back. An approach to those limitations can be to consider introduction of hierarchical learning algorithms, which allow to define an additional layer where costs tuning and appropriate parameters calibration can be achieved on top of the learning mechanism.

Robustness and seamless dynamic: We have studied different network architecture mainly coexistence of different technologies with overlapping areas from one side and DTN networks architectures from the other side. In our works we explored cases when the population of users is homogeneous so that they have similar constraints and updating patterns. We can assume different updating rules(step sizes for example) in order to introduce some heterogeneity between users. Another aspect is related to repartition of users over the coverages of coexisting technologies. We noticed that having some users outside the area of coexistence, where the competition for better access holds, leads to better performances in average, this because it can compensate the defaults of the strategies used in the area of coexistence. In an extension of this dynamic,

we can study the impact of users mobility in this specific context, where users will alternately enter an exit the area of competition. In DTN configurations from another side, it would be interesting to study forwarding scheme with recovering capabilities for nodes in the network. By allowing arrivals and departures of users in the networks, we can simulate nodes recovering from infections, as arrivals of fresh nodes at the following of departure of some infected nodes. This configuration makes the system dynamic in the number of mobiles, a more realistic approach to a DTN case. However this adds a higher complexity to the infection rate model as a fluid propagation model in a tube with perforated holes. In the same context of DTNs, considering live time control games, we restricted activation control in our analysis for the design of our learning algorithm. It will be interesting to investigate the general case of live-time control with actual partial activation, and study the conditions for stability of the learning process with increased dynamic.

Appendix A : Proof of optimal source forwarding control policy in DTN

Proof of theorem 5.4.0.1 In the proof of this theorem we require the two following lemma on the costs function $H(y,u)$, and other general results.
Lemma 7.5.1. *For a fixed number of users N in a local interaction, the costs function H is concave in u for every given value of the population profile $\alpha(y)$.*

Proof. For N fixed and a given value of the population profile $\alpha(y)$, let

$$h^N(u) = (Q^u_{\tau'})^N \frac{1-(1-\beta\alpha(y))^N}{N\alpha(y)} A - \frac{1-(Q^u_{\tau'})^N}{N}\eta$$

with $A = (1 - P_A(\tau'))$ and $\eta = (P_A(\tau') - P_A(\tau))$, from the expression of $H(\alpha(y),u)$. If h is concave in u for any N then $H(\alpha(y),u)$ is also concave. Using $\beta(u) = 1 - \frac{Q^u_\tau}{Q^u_{\tau'}}$ we have,

$$\begin{aligned}
h^N(u) &= (Q^u_{\tau'})^N \frac{1-\left[1-(1-\frac{Q^u_\tau}{Q^u_{\tau'}})\alpha(y)\right]^N}{N\alpha(y)} A - \frac{1-(Q^u_{\tau'})^N}{N}\eta \\
&= \frac{(Q^u_{\tau'})^N - \left[Q^u_{\tau'} - (Q^u_{\tau'} - Q^u_\tau)\alpha(y)\right]^N}{N\alpha(y)} A - \frac{1-(Q^u_{\tau'})^N}{N}\eta \\
&= \frac{-1}{N\alpha(y)}\left[(Q^u_{\tau'}(1-\alpha(y)) + Q^u_\tau\alpha(y))^N A - (Q^u_{\tau'})^N(A+\alpha(y)\eta)\right] - \frac{\eta}{N}.
\end{aligned}$$

Let $B = A + \alpha(y)\eta$; we express the derivative of $h^N(u)$:

$$\frac{dh^N(u)}{du} = \frac{-1}{\alpha(y)}\left[A[Q^u_{\tau'}(1-\alpha(y)) + Q^u_\tau\alpha(y)]^{N-1}[(\dot{Q^u_{\tau'}})(1-\alpha(y)) + (\dot{Q^u_\tau})\alpha(y)] - (Q^u_{\tau'})^{N-1}(\dot{Q^u_{\tau'}})\right]$$

Let $f(\alpha(y),u) = Q^u_{\tau'}(1-\alpha(y)) + Q^u_\tau\alpha(y)$ then $f(0,u) = Q^u_{\tau'}$.

$$\begin{aligned}
\frac{dh^N(u)}{du} &= \frac{1}{\alpha(y)}\left[Af^{N-1}(\alpha(y),u)(-f'(\alpha(y),u)) - B(f(0,u))^{N-1}(-f'(0,u))\right] \\
&= \frac{1}{\alpha(y)}\Big((f(0,u))^{N-1}[A(-f'(\alpha(y),u)) - B(-f'(0,u))] + \\
&\quad A(-f'(\alpha(y),u))[f^{N-1}(\alpha(y),u) - f^{N-1}(0,u)]\Big)
\end{aligned}$$

We know that $(\dot{Q^u_\tau}) \leq (\dot{Q^u_{\tau'}})$, f is decreasing in u and $(\dot{Q^u_{\tau'}}) = (\dot{Q^u_{\tau'}})(1-\alpha(y)) + (\dot{Q^u_{\tau'}})\alpha(y)$ then $(\dot{Q^u_{\tau'}}) \geq (\dot{Q^u_{\tau'}})(1-\alpha(y)) + (\dot{Q^u_\tau})\alpha(y)$ this implies that $f'(0,u) \geq f'(\alpha(y),u)$ and $f^{N-1}(\alpha(y),u) - f^{N-1}(0,u)$ is decreasing. Let's show that $A(-f'(\alpha(y),u)) - B(-f'(0,u))$ is also decreasing.

$$A(-f'(\alpha(y),u)) - B(-f'(0,u)) = A[f'(0,u) - f'(\alpha(y),u)] + \alpha(y)\eta f'(0,u)$$

$\Rightarrow$

$$\begin{aligned}\frac{d(A(-f'(\alpha(y),u))-B(-f'(0,u)))}{du} &= A\alpha(y)((\ddot{Q}^u_{\tau'})-(\ddot{Q}^u_{\tau}))+\alpha(y)\eta(\ddot{Q}^u_{\tau'})\\ &= \frac{B\left[e^{-\lambda u\tau'}((\lambda\tau')^2(1-u)^2+2)-2ue^{-\lambda\tau'}\right]}{(1-u)^3}\\ &\quad -\frac{A\left[e^{-\lambda u\tau}((\lambda\tau)^2(1-u)^2+2)-2ue^{-\lambda\tau}\right]}{(1-u)^3}\end{aligned}$$

Though the negativity of this last expression has been observed for all the several set of values we experimented, it is not obvious to see here. We will then assume that it is negative and conclude that the function H is concave in u.
This completes the proof. □

Lemma 7.5.2. *Let $\tilde{y}^*$ the solution of the equation $H(y)=0$. The following assertions are verified:*

- *$Q^u_{\tau'}$ is a decreasing function in u.*
- *Q^u_{τ} is a decreasing function in u.*
- *$\tilde{y}^*$ is a decreasing function in u .*
- *Under some specific conditions we have $\tilde{y}^*(\epsilon)\geq 1$ otherwise $\tilde{y}^*(\epsilon)\leq 0$ with ϵ a very small positive number.*

Proof. The proof of the two first points derive intuitively from the fact that the higher the value of the source control the higher the chance for a given relay to deliver the message to the destination. Thus $Q^u_{\tau'}$ and Q^u_{τ} are decreasing functions in u. We prove here the two last points.

- To show that $\tilde{y}^*$ is a decreasing function in u, we first show that H is a non-increasing function of y. Indeed, we have,

  $$H(y)=\frac{\sum_{N=1}^{\infty}P(K=N)\left[(Q_{\tau'})^N\frac{1-(1-\beta y)^N}{Ny}(1-P_A(\tau'))-\frac{1-(Q_{\tau'})^N}{N}(P_A(\tau')-P_A(\tau))\right]}{(1-P_A(\tau'))(1-P_A(\tau))}.$$

  Only the term $f(y)=\dfrac{1-(1-\beta y)^N}{Ny}$, in the expression of H is dependent on y.
  For every parameters (other than y) fixed, we have:

  $$\begin{aligned}\frac{df(y)}{dy} &= \frac{N\beta(1-\beta y)^{N-1}Ny-N\left(1-(1-\beta y)^N\right)}{(Ny)^2}\\ &= \frac{N\beta y(1-\beta y)^{N-1}-1+(1-\beta y)^N}{Ny^2}=\frac{(1-\beta y)^{N-1}(N\beta y+1-\beta y)-1}{Ny^2}\\ &= \frac{(1-\beta y)^{N-1}((N-1)\beta y+1)-1}{Ny^2}\end{aligned}$$

Let's show that $(1-\beta y)^{N-1}((N-1)\beta y+1)-1$ is negative. $(1-\beta y)^{N-1}((N-1)\beta y+1)-1 \leq 0 \Rightarrow (N-1)\beta y+1 \leq \frac{1}{(1-\beta y)^{N-1}}$ which is true since $\frac{d((N-1)\beta y+1)}{dy}\,|_{(y=0)}= \frac{d\left(\frac{1}{(1-\beta y)^{N-1}}\right)}{dy}\,|_{(y=0)}= \beta(N-1)$ and $\frac{d^2\left(\frac{1}{(1-\beta y)^{N-1}}\right)}{d^2y} = \frac{\beta^2(N-1)(N+1)}{(1-\beta y)^{N+2}} > 0$.
Thus H is a non-increasing function of y.
Using lemma 7.5.1 we have $if\ \exists\ u\ s.t\ H(u)=0$ for a given y then, u is unique in $[0,1]$.
$\forall y_1 \leq y_2,\ H(y_1,u) \geq H(y_2,u)$, then $if\ \exists\ u_1, u_2\ s.t.\ H(y_1,u_1)=H(y_2,u_2)=0$ then $u_1 \geq u_2$ and $\tilde{y}^*$ is a decreasing funtion of u.

- We have, $H(y,u)=V_\tau(F,y)-V_\tau(F,y)$, let's find a condition on τ and τ' so that $H(y,u)\geq 0$,
$\Rightarrow\ V_\tau(F,y) \geq V_\tau(F,y),\ \iff$

$$\begin{aligned} U(\tau,y)\left(\frac{1}{1-P_F}+\frac{1}{1-P_A(\tau)}\right) &\geq \frac{U(\tau,y)}{1-P_F}+\frac{U(\tau',y)}{1-P_A(\tau')} \\ \frac{U(\tau,y)}{1-P_A(\tau)} &\geq \frac{U(\tau',y)}{1-P_A(\tau')} \\ \frac{1-P_A(\tau')}{1-P_A(\tau)}\frac{U(\tau,y)}{U(\tau',y)} &\geq 1 \\ T\frac{U(\tau,y)}{U(\tau',y)} &\geq 1 \end{aligned}$$

with $T=\frac{1-P_A(\tau')}{1-P_A(\tau)} \leq 1$. When u is taken very small, we have,

$$\begin{aligned} \frac{U(\tau,y)}{U(\tau',y)} &= \frac{\sum_{N=1}^{\infty}P(K=N)(P_{succ}(\tau',N,y)+(Q_{\tau'})^N\frac{1-(1-\beta y)^N}{Ny})}{\sum_{N=1}^{\infty}P(K=N)P_{succ}(\tau',N,y)} \\ &= 1+\frac{\sum_{N=1}^{\infty}P(K=N)(Q_{\tau'})^N\frac{1-(1-\beta y)^N}{Ny}}{\sum_{N=1}^{\infty}P(K=N)\frac{1-(Q_{\tau'})^N}{N}} \\ &= 1+\frac{\sum_{N=1}^{\infty}P(K=N)\frac{d}{du}\frac{1-(1-\beta y)^N}{Ny}}{\sum_{N=1}^{\infty}P(K=N)\frac{d}{du}(-(Q_{\tau'}))} \end{aligned}$$

Considering u small, we have, $\frac{d}{du}(-(Q_{\tau'}))=\lambda\tau'+e^{-\lambda\tau'}-1$ and

$$\begin{aligned} \frac{d}{du}\left(\frac{1-(1-\beta y)^N}{Ny}\right) &= \frac{\frac{d}{du}[1-(1-\beta y)^N](Ny-(N\dot{y}(1-(1-\beta y))))}{Ny^2} \\ &= \frac{(1-\beta y)^{N-1}\frac{d}{du}(\beta y)}{y} \end{aligned}$$

assuming that $\dot{y}$ is bounded

$$\frac{d}{du}\left(\frac{1-(1-\beta y)^N}{Ny}\right) = \frac{\beta\dot{y}+\dot{\beta}y}{y} = \dot{\beta}$$

Thus,

$$\frac{U(\tau,y)}{U(\tau',y)} = 1 + \frac{\dot{\beta}}{\lambda\tau' + e^{-\lambda\tau'} - 1} = \frac{\lambda\tau + e^{-\lambda\tau} - 1}{\lambda\tau' + e^{-\lambda\tau'} - 1}$$

and $H(y,u) \geq 0 \iff \frac{1}{T} \leq \frac{\lambda\tau + e^{-\lambda\tau} - 1}{\lambda\tau' + e^{-\lambda\tau'} - 1}$.

- If $\frac{1}{T} \leq \frac{\lambda\tau + e^{-\lambda\tau} - 1}{\lambda\tau' + e^{-\lambda\tau'} - 1}$ then $H(y,u) > 0$ and $y* = 1$.
- If $\frac{1}{T} > \frac{\lambda\tau + e^{-\lambda\tau} - 1}{\lambda\tau' + e^{-\lambda\tau'} - 1}$ then $H(y,u) \leq 0$ and $y* = 0$.

This completes the proof. □

We give now the proof of the theorem 5.4.0.1.

Proof. Given the expression of the probability of success, maximizing $P_s(u)$ comes down to minimize the expression $(Q^u_{\tau'})^{(1-y(u))} . (Q^u_{\tau})^{y(u)}$. Let

$$f(u) = (1 - y(u)) \log(Q^u_{\tau'}) + y(u) \log(Q^u_{\tau})$$

, we need to minimize $f(u)$.

$f'(u) = y'(u)\Big[\log(Q^u_{\tau}) - \log(Q^u_{\tau'})\Big] + (1 - y(u))\left(\frac{(Q^u_{\tau'})'}{(Q^u_{\tau'})} - \frac{(Q^u_{\tau})'}{(Q^u_{\tau})}\right) + \frac{(Q^u_{\tau})'}{(Q^u_{\tau})}$.

For u small, using lemma 7.5.2, we have: $y* = 1$ or $y* = 0$

If $y* = 0$, given that $y*$ is decreasing in u then $y* = 0 \ \forall \ u$. $f'(u) = \frac{(Q^u_{\tau'})'}{(Q^u_{\tau'})} \leq 0$ thus, f is decreasing and $P_s(u)$ is always increasing on $[0,1]$.

On the other hand, if $y* = 1$ for u small, we need to prove that if $P_s(u) = P_{max}$ then $u \in [u_0, 1]$. Since y^* is a decreasing function of u, we have, $\bar{y}^*(0) = 1 \implies \exists u_0$ s.t. $\bar{y}^*(u_0) = 1$ and $\bar{y}^*(u_0 + \delta) \leq 1$, $\delta > 0$. Where $\bar{y}^*$ is the projection of y^* on the interval $[0,1]$.$y* = 1$ for $u \in [0, u_0] \Rightarrow f$ is decreasing and P_s is always increasing $for\ u \in [0, u_0]$.

This completes the proof. □

List of publications

International Journals

1. Rachid El-Azouzi, Francesco De Pellegrini, Habib B.A. SIDI and Vijay Kamble "Evolutionary forwarding games in Delay Tolerant Networks: equilibria, mechanism design and stochastic approximation". Computer Networks, Available online 1 December 2012. url:http://dx.doi.org/10.1016/j.comnet.2012.11.014
2. Habib B. A. Sidi, Rachid El-Azouzi, Majed Haddad, "Fractional Frequency Reuse Stackelberg Model for Self-Organizing Networks", EURASIP Journal on Wireless Communications and Networking (Submitted 2012).

International Conferences

1. Rachid El-Azouzi, Habib B.A. Sidi, Francesco De Pellegriniy and Yezekael Hayel. "Markov Decision Evolutionary Game for Energy Management in Delay Tolerant Networks" NETGCOOP 2011, Paris France.
2. Habib B.A. Sidi, Rachid El-Azouzi et Majed Haddad. "Fractional Frequency Reuse Stackelberg Model for Self-Organizing Networks" Wireless Day 2011, Ontario Canada.
3. Rachid El-Azouzi, Habib B.A. Sidi, Julio Rojas-Mora, and Amar Prakash Azad. "Delay Tolerant Networks in partially overlapped networks: A non-cooperative game approach". BIONETICS 2009, Avignon France.

Workshops and National Communications

1. Habib B.A. Sidi, Rachid El-Azouzi, Yezekael Hayel, Julio Rojas-Mora. "Méthode distribuée de gestion dynamique des ressources radios dans les réseaux sans fils hétérogènes". ROADEF 2010, Toulouse France.
2. Julio Rojas-Mora, Habib B.A. Sidi, R. El-Azouzi, and Yezekael Hayel. "Distributed learning in a Congestion Poisson Game". AlgoGT2010, Bordeaux France.
3. Habib B.A. Sidi, Wissam Chahin, Rachid El-Azouzi and Francesco De Pellegrini. "Energy efficient minority game for delay tolerant networks". Technical Report,

url:http://arxiv.org/abs/1207.6760, 2012.

4. Julio Rojas-Mora., Habib B.A. Sidi., Rachid El-Azouzi and Yezekael Hayel (2010). "A Decentralized Algorithm for Radio Resource Management in Heterogeneous Wireless Networks with Dynamic Number of Mobiles". LIA Research Report.

Bibliography

[1] COST 231. *Urban Transmission Loss Models for Mobile Radio in the 900 and 1800 MHz Bands*. EURO-COST Std. 231, 1991.

[2] 3GPP Std. *Soft frequency reuse scheme for UTRAN LTE*. R1-050 507, May 2005.

[3] 3rd Generation Partnership Project (3GPP). *Evolved Universal Terrestrial Radio Access Network (EUTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions*. TR 36.902, Dec. 2008.

[4] P. Nain A. Al-Hanbali and E. Altman. Performance of ad hoc networks with two-hop relay routing and limited packet lifetime. In *First International Con- ference on Performance Evaluation Methodologies and Tools (Valuetools)*, Pisa, 2006.

[5] J. Crowcroft C. Diot R. Gass A. Chaintreau, P. Hui and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6:606–620, 2007.

[6] C. Barakat A. Krifa and T. Spyropoulos. Optimal buffer management policies for delay tolerant networks. In *Proceedings of IEEE SECON*, June 16-20 2008.

[7] L. L. abd T. Soderstrom. *Theory and practice of recursive identification*. in MIT Press, Cambridge, MA, 1983.

[8] Y. Agarwal, R. Gupta, and C. Schurgers. Dynamic power management using on demand paging for networked embedded systems. In *Design Automation Conference, 2005. Proceedings of the ASP-DAC 2005. Asia and South Pacific*, volume 2, pages 755 – 759 Vol. 2, jan. 2005.

[9] Syed Hussain Ali and Victor C. M. Leung. Dynamic frequency allocation in fractional frequency reused ofdma networks. In *In IEEE Transactions on Wireless Communications*, 2009.

[10] E. Altman, A. Prakash Azad, T. Basar, and Francesco De Pellegrini. Optimal activation and transmission control in delay tolerant networks. In *Proceedings of IEEE INFOCOM, San Diego*, 15–19 March, 2010.

[11] Eitan Altman. Competition and cooperation between nodes in delay tolerant networks with two hop routing. In *Proceedings of Netcoop*, Eindhoven, The Netherlands, Nov. 23-25 2009.

[12] Amar Prakash Azad. *Advances in Network Control and Optimization*. PhD thesis, Laboratoire d'Informatique d'Avignon (UPRES No 4128), 2010.

[13] R.S. B. A. Sutton and R. Williams. Reinforcement learning is direct adaptive control. *IEEE Control Systems Magazine*, pages 19–22, April, 1992.

[14] Q. Zhang B. L. H. Y. X. G. C. Long. Non-cooperative power control for wireless ad hoc networks with repeated games. 25(6), 2007.

[15] J. Baxter and P. L. Bartlett. Infinite-horizon policy-gradient estimation. *Journal of Artificial Intelligence Research*, 15:319–350, 2001.

[16] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999.

[17] M. Benaim and M. Hirsch. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. 29(1-2):36–72, October, 1999.

[18] Dimitri Bertsekas and Robert Gallager. *Data Networks*. Prentice Hall, 1987.

[19] John N. Bertsekas, Dimitri P.; Tsitsiklis. *Parallel and Distributed Computation:Numerical Methods*. [Online] URI: http://hdl.handle.net/1721.1/3719, 1989.

[20] T. Borgers and R. R. Sarin. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory, Elsevier*, 77(1):1–14, November, 1997.

[21] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004.

[22] J. Brown. *The nature of set to learn and of intra-material interference in immediate memory*, volume 6. in Quart. J. of exp. Psychol.

[23] N. B. Mandayam C. U. Saraydar and D. J. Goodman. Pricing and power control in a multi-cell wireless data network. *IEEE Journal on Selec. Areas in Comm*, 19:1883–1892, 2001.

[24] W. Chahin, R. El-Azouzi, F. De Pellegrini, and A.P. Azad. Blind online optimal forwarding in heterogeneous delay tolerant networks. pages 1 –6, oct. 2011.

[25] D. Challet and Y.-C. Zhang. Emergence of Cooperation and Organization in an Evolutionary Game. page 8006, August 1997.

[26] Chung Shue Chen and FranÃ§ois Baccelli. Self-optimization in mobile cellular networks: Power control and user association. In *in Proc. IEEE ICC*, 2010.

[27] R. Combes, Z. Altman, M. Haddad, and E. Altman. Self-optimizing strategies for interference coordination in OFDMA networks. In *IEEE ICC 2011 Workshop on Planning and Optimization of Wireless Communication Networks*, Kyoto, Japan, June 2011.

[28] L. Echabbi D. Barth and C. Hamlaoui. Optimal transit price negotiation: The distributed learning perspectives. *Journal of Universal Computer Science*, 14:745–765, 2008.

[29] F. De Pellegrini, E. Altman, and T. Basar. Optimal monotone forwarding policies in delay tolerant mobile ad hoc networks with multiple classes of nodes. In *proc. of WiOpt WDM Workshop*, Avignon, France, June 4 2010.

[30] Pauolo Teixeira de Sousa and Peter Stuckmann. *Telecommunication network interoperability*. Telecommunication systems and technologies Vol. II http://www.eolss.net/Sample-Chapters/C05/E6-108-22.pdf,.

[31] Brewer E. Fall K. Jain S. Ho M. Patra R. Demmer, M. *Implementing Delay Tolerant Networking*. Technical report, IRB-TR-04-020, Intel Corporation, December 2004.

[32] Arnaud Doucet, Nando de Freitas, and Neil (Eds.) Gordon. Sequential monte carlo methods in practice. XIV:581 p. 168 illus., 2001.

[33] T. Basar E. Altman and F. De Pellegrini. Optimal control in two hop relay routing. In *Proceedings of IEEE CDC*, Shanghai, China, December 16–18 2009.

[34] Y. Hayel E. Altman, R. El-Azouzi and H. Tembine. The evolution of transport protocols: An evolutionary game perspective. *Computer Networks Journal*, 53:1751–1759, 2010.

[35] Francesco De Pellegrini Daniele Miorandi Eitan altman, Giovanni Neglia. Decentralized stochastic control of delay tolerant networks. In *IEEE Infocom*, Rio de Janeiro, Brazil, April 19-25 2009.

[36] Tamer Basar Eitan Altman and Francesco De Pellegrini. Optimal monotone for- warding policies in delay tolerant mobile ad-hoc networks. In *Inter-Perf 2008:Work- shop on Interdisciplinary Systems Approach in Performance Evaluation and Design of Computer and Communication Systems*, Athens, Greece, October 2008.

[37] R. El-Azouzi, F. De Pellegrini, and V. Kamble. Evolutionary forwarding games in delay tolerant networks. In *Proceedings of WiOPT*, Avignon, 29 May- 5 June 2010.

[38] Rachid El-Azouzi, Habib B. A. Sidi, Francesco De Pellegrini, and Yezekael Hayel. Markov decision evolutionary game for energy management in delay tolerant networks. In *NETGCOOP 2011*, October, 2011.

[39] I. Erev and A. Roth. Prediction how people play games : Reinforcement learning in games with unique strategy equilibrium. *American Economic Review*, 88:848–881, 1998.

[40] E. Ferro, Erina Ferro, and F. Potorti. Bluetooth and wi-fi wireless protocols: A survey and a comparison. *IEEE Wireless Communications*, 12:12–26, 2004.

[41] Filippov. *Differential equations with discontinuous righthand sides*. in Kluwer Academic, Dordrecht, 1988.

[42] David Mittelstädt Florian Wamser and Dirk Staehle. Soft frequency reuse in the up-link of an ofdma netwok. In *IEEE VTC*, 2010.

[43] Doug Gray for WiMAX forum. Comparing mobile wimax with hspa+, lte, and meeting the goals of imtadvanced, February 2009.

[44] Fudenberg and Levine. Learning in games : Where do we stand ? *European Economic*, Review 42:631–639, 1998.

[45] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.

[46] Herbert Gintis. *Game Theory Evolving*. Princeton University Press, 2009.

[47] R. Giuliano, C. Monti, and P. Loreti. Wimax fractional frequency reuse for rural environments. *IEEE Commun. Mag.*, 15:60–65, Jun. 2008.

[48] J. Greifenberg and D. Kutschera. Rdtn: An agile dtn research platform and bundle protocol agent. In *7th International Conference, WWIC 2009*, Enschede, The Netherlands, May 27-29 2009.

[49] Robin Groenevelt and Philippe Nain. Message delay in manets. *In ACM SIGMETRICS, Banff, Canada*, pages 412–413, June 2005.

[50] Alessio Guerrieri, Iacopo Carreras, Francesco De Pellegrini, Daniele Miorandi, and Alberto Montresor. Distributed estimation of global parameters in delay-tolerant networks. *Computer Communications*, 33(13):1472–1482, 2010.

[51] E. R. Guthrie. *Psychological facts and psychological theory*, volume 43. in Psychological Bulletin.

[52] R. El-Azouzi Y. Hayel H. Tembine, E. Altman. Evolutionary games in wireless networks. *to appear in IEEE Transaction On systems, MAN and Cyberbetics*, 2010.

[53] Majed Haddad, Zwi Altman, Salah Eddine Elayoubi, and Eitan Altman. A nash-stackelberg fuzzy q-learning decision approach in heterogeneous cognitive networks. In *GLOBECOM*, Miami, USA, Dec. 2010.

[54] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica, Econometric Society*, 68(5):1127–1150, September, 2000.

[55] Josep Colom Ikuno, Martin Wrulich, and Markus Rupp. System level simulation of LTE networks. In *Proc. 2010 IEEE 71st Vehicular Technology Conference*, may 2010, address = Taipei, Taiwan, http://publik.tuwien.ac.at/files/PubDat_184908.pdf,.

[56] D. Jensen J. Burgess, B. Gallagher and B. N. Levine. Maxprop: Routing for vehicle-based disruption-tolerant networks. In *Proceedings of IEEE Infocom 2006*, 2006.

[57] T. Fruedman J. Leguay and V. Conan. Evaluating mobility pattern space routing for dtns. In *Proceedings of IEEE INFOCOM*, 2006.

[58] Aktul Kavas. Comparative analysis of wlan, wimax and umts technologies. In *PIERS Proceedings, 140 - 144*, Prague, Czech Repulic, August 27-30 2007.

[59] Kets, W., Voorneveld, and M. Congestion, equilibrium and learning: The minority game. (2007-61), 2007.

[60] Nevelson M.B Khasminskii, R.Z. Stochastic approximation and recursive estimation. *Translations of Mathematical Monographs*, 47, 1976.

[61] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. 23:462–466, 1952.

[62] S. Kiran and R. Chandramouli. An adaptive energy efficient link layer protocol for bursty transmissions over wireless data networks. In *Proceedings of ICC 2003*, 2003.

[63] V. R. Konda and J. T. Tsitsiklis. Actor-critic algorithms. *SIAM Journal on Control and Optimization*, 42(4):1143–1166, 2003.

[64] Vimal Kumar and Neeraj Tyagi. Media independent handover for seamless mobility in ieee 802.11 and utms based on ieee 802.11. In *International Conference on computer science and Information Technology*, Chengdu, China, July 9-11 2010.

[65] H. J. Kushner and G.G. Yin. Stochastic approximation algorithms and applications. *Applications of Mathematics. 35. Berlin: Springer.*, 1997.

[66] Harold J. Kushner and G. George Yin. *Stochastic Approximation Algorithms and Recursive Algorithms and Applications*. Springer-Verlag, New York, 2nd edition, 2003.

[67] G. Hiertz L. Berlemann, C. Hoymann and S. Mangold. Coexistence and interworking of ieee 802.16 and 802.11(e). In *Proceedings of VTC 2006*, 2006.

[68] J. Perez-Romero L. Giupponi, R. Agusti and O. Sallent. A framework for jrrm with resource reservation and multiservice provisioning in heterogeneous networks, mobile networks and applications. *Springer*, 11, 2006.

[69] J. Perez-Romero L. Giupponi, R. Agusti and O. Sallent. A novel approach for joint radio resource management based on fuzzy neural methodology. *to appear in IEEE Trans. on Vehicular Technology*, 2007.

[70] A. Passarella L. Pelusi and M. Conti. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *IEEE Communications Magazine*, 44:134–141, November 2006.

[71] R. P. M. G. Lagoudakis. Least-squares policy iteration. *Journal of Machine Learning Research*, 4:1107–1149, 2003.

[72] L. Ljung. Analysis of recursive stochastic algorithms. *in IEEE Trans. Automat. Control.*, AC-22:551–575, 1977.

[73] L. Ljung. *System identification theory for the user*. in Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

[74] Petri Mähönen and Marina Petrova. Minority game for cognitive radios: Cooperating without cooperation. *Physical Communication*, 1(2):94 – 102, 2008.

[75] P. Maille and B. Tuffin. The progressive second price mechanism in a stochastic environment. *Netnomics*, 5:119–147, May 2003.

[76] Joseph Mitola. *Cognitive radio: An integrated agent architecture for software defined radio*. PhD thesis, Stockholm, Sweden, 2000.

[77] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. In *International Symposium on Voice, Video and Data Communications*, 1998.

[78] Esteban Moro. The Minority Game: an introductory guide. *eprint arXiv:cond-mat/0402651*, February 2004.

[79] R. Myerson. Population uncertainty and poisson games. *International Journal of Game Theory*, 27, 1998.

[80] K. Najim and A. S. Poznak. Learning automata-theory and applications. New York: Pergamon, 1994.

[81] K. Narendra and M.A.L. Thatcher. *Learning automata: An introduction*. Prentice Hall, New Jersey, 1989.

[82] K. S. Narendra and M. A. L. Thathachar. On the behavior of learning automata in a changing environment with applications to telephone traffic routing. *IEEE Trans. on Syst., Man and Cybernetic*, 10:262–269, 1980.

[83] G. Neglia and X. Zhang. Optimal delay-power tradeoff in sparse delay tolerant networks: a preliminary study. In *Proceedings of ACM SIGCOMM 237–244*, CHANTS, 2006.

[84] NGMN Alliance. *NGMN Recommendation on SON and O&M Requirements*. Dec. 2008.

[85] Mark D. Corner Nilanjan Banerjee and Brian Neil Levine. An energy-efficient architecture for dtn throwboxes. *INFOCOM 2007. 26th IEEE International Conference on Computer Communications*, pages 776 – 784, 6-12 May 2007.

[86] Y. Xing P. Maille, B. Tuffin and R. Chandramouli. User stragegy learning when pricing a red buffer. *Simulation Modelling, Practice and Theory*, 17:548–557, March 2009.

[87] Trevor Pering, Yuvraj Agarwal, Rajesh Gupta, and Roy Want. Coolspots: reducing the power consumption of wireless mobile devices with multiple radio interfaces. In *Proceedings of the 4th international conference on Mobile systems, applications and services*, MobiSys '06, pages 220–232, New York, NY, USA, 2006. ACM.

[88] Gian Paolo Perrucci. *Energy Saving Strategies on Mobile Devices*. PhD thesis, 9220 Aalborg, Denmark, 2009.

[89] Corinne Touati Pierre Coucheney and Bruno Gaujal. Fair and efficient user-network association algorithm for multi-technology wireless networks. In *Proceedings of INFOCOM 2009 Mini Conference*, 2009.

[90] W. B. Powell. *Approximate dynamic programming, solving the curses of dimensionality*. in Wiley Sons, New Jersey, 2007.

[91] V.V. Phansalkar P.S. Sastry and M.A.L. Thathachar. Decentralized learning of nash equilibria in multi-person stochastic games with incomplete information. *IEEE Transactions on Systems, Man and Cybernetics*, 24:769–777, May 1994.

[92] J. Rojas-Mora R. El-Azouzi, B. A. Habib Sidi and Amar Azad. Game theory and learning algorithm in delay tolerant network. In *Proceedings of Bionetics*, Avignon, 10-11 December 2009.

[93] V. Raghunathan and P. Kumar. On delay-adaptive routing in wireless networks. In *Proceedings of CDC 2004*, 2004.

[94] V. Raghunathan, T. Pering, R. Want, A. Nguyen, and P. Jensen. Experience with a low power wireless mobile computing platform. In *Low Power Electronics and Design, 2004. ISLPED '04. Proceedings of the 2004 International Symposium on*, pages 363 – 368, aug. 2004.

[95] Herbert Robbins and Sutton Monro. A stochastic approximation method. *Ann. Math. Statist.*, 22(3):400–407, 1951.

[96] Philippe Nain Robin Groenevelt and and Ger Koole. The message delay in mobile ad hoc networks. In *Performance Evaluation 62 (2005) 210-228*, August 2005.

[97] Robert W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2, 1973.

[98] R.W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *J. Game Theory*, 2:65–67, June 2005.

[99] K. Fall V. Cerf B. Durst K. Scott S. Burleigh, L. Torgerson and H. Weiss. Delay-tolerant networking: an approach to interplanetary Internet. *IEEE Communications Magazine*, 41:128–136, June 2003.

[100] R. Patra S. Jain, K. Fall. Routing in a delay tolerant networking. In *Proceedings of SIGCOMM 2004*, Miami, Florida, 2004.

[101] W. Sandholm. Evolutionary implementation and congestion pricing. *Review of Economic Studies*, 69, 2002.

[102] W. H. Sandholm. *Population Games and Evolutionary Dynamics*. Forthcoming, MIT Press,Cambridge, 2008.

[103] Shang and Li Hui. Self-organized evolutionary minority game on networks. *2007 IEEE International Conference on Control and Automation*, 00:2186–2188, 2007.

[104] Li Hui Shang. Self-organized evolutionary minority game on networks. In *International Conference of Control and Automation*, May 30- June 1, 2007.

[105] B. Skinner. *Science and human behavior*. in New York : Macmillan, 1953.

[106] J. MAYNARD SMITH and G. R. Price. The logic of animal conflicts. *Nature*, 246, 11/02 1973.

[107] K. Son, S. Chong, and G. de Veciana. Dynamic association for load balancing and interference avoidance in multi-cell networks. *IEEE Trans. Wireless Commun.*, 8:3566–3576, Jul. 2009.

[108] T. Spyropoulos, K. Psounis, and C. Raghavendra. Efficient routing in intermittently connected mobile networks: the multi-copy case. *ACM/IEEE Transactions on Networking*, 16:77–90, Feb. 2008.

[109] D. Stahl. The inefficiency of first and second price auctions in dynamic stochastic environments. *Netnomics*, 4:1–18, 2002.

[110] A. L. Stolyar and H. Viswanathan. Self-organizing dynamic fractional frequency reuse in ofdma systems. In *In Proc. IEEE INFOCOM*, 2008.

[111] A. L. Stolyar and H. Viswanathan. Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordiantion. In *Proceedings of IEEE INFOCOM*, 2009.

[112] Illsoo Sohn R. K. Ganti T. Novlan, J. G. Andrews. Comparison of fractional frequency reuse approaches in the ofdma cellular downlink. In *In Proc. IEEE Globecom*, 2010.

[113] J. B. N. Tao and L. Weaver. A multi-agent policy gradient approach to network routing. In *Proceedings of ICML*, 2001.

[114] Muhammad Mukarram Bin Tariq, Mostafa Ammar, and Ellen Zegura. Message ferry route design for sparse ad hoc networks with mobile nodes. In *Proceedings of ACM MobiHoc*, Florence, Italy, May 22–25 2006.

[115] Peter D. Taylor and Troy Day. Stability in negotiation games and the emergence of cooperation. *Proceedings of the Royal Society*, 2007.

[116] Hamidou Tembine, Eitan Altman, Rachid El Azouzi, and Yezekael Hayel. Bio-inspired Delayed Evolutionary Game Dynamics with Networking Applications. *Telecommunication Systems*, vol. 47(no. 1-2):pp. 137–152, June 2011.

[117] A. L. Thathachar and P. S. Sastry. A class of rapidly converging algorithms for learning automata. In *Proc. IEEE Int. Conf. Cybern Soc.*, pages 602–606, 1984.

[118] M.A.L. Thathachar and P.S. Sastry. *Networks of learning automata, Techniques for Online Sthochastic Optimization*. Kluwer Academic Publishers Group, 2004.

[119] Wikipedia the free encyclopedia. Machine learning. http://en.wikipedia.org/wiki/Machine_learning.

[120] A. Vahdat and D. Becker. Epidemic routing for partially connected adhoc networks. *Technical report, CS-2000-06,Duke University*, 2000.

[121] Vincent and Brown. *Evolutionary Game Theory, Natural Selection and Darwinian Dynamics*. Cambridge University Press, 2005.

[122] E. Zegura W. Zhao, M. Ammar. Controlling the mobility of multiple data transport ferries in a delay-tolerant network. In *Proceedings of IEEE INFOCOM 2005*, Miami, Florida, March 13–17, 2005.

[123] J. B. Watson. *Behaviorism*. in University of Chicago Press, 1930.

[124] J Weibull. *Evolutionary Game Theory*. MIT Press, 1995.

[125] Fierce Broadband Wireless. Wimax, lte and hspa+: Comparing operators' 4g coverage areas. http://www.fiercebroadbandwireless.com/special-reports/wimax-lte-and-hspa-comparing-operators-4g-coverage-areas.

[126] S. Mau X. Jing and R. Matyas. Reactive cognitive radio algorithms for co-existence between ieee 802.11b and 802.16a networks. In *Proceedings of Globecom 2005*, 2005.

[127] J. Kurose X. Zhang, G. Neglia and D. Towsley. Performance modeling of epidemic routing. *Computer Networks*, 51:2867–2891, 2007.

[128] Y. Xing and R. Chandramouli. Qos constrained secondary spectrum sharing. In *Proceedings of DySPAN 2005*, 2005.

[129] Yiping Xing, R. Ch, and Senior Member. R.: Stochastic learning solution for distributed discrete power control game in wireless data networks. Technical report, 2004.

[130] Haojin Zhu, Xiaodong Lin, Rongxing Lu, Yanfei Fan, and Xuemin Shen. Smart: A secure multilayer credit-based incentive scheme for delay-tolerant networks. *Vehicular Technology, IEEE Transactions on*, 58(8):4628 –4639, oct. 2009.