THE PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
Abou Bekr Belkaid Tlemcen University
Faculty of Sciences
Computer Sciences Departement

Doctoral Thesis

Presented by

Fayssal Bendaoud

Entitled :

Management of joint radio resources in heterogeneous networks Beyond 3G

Jury members:

President	A. Chikh	Professor at University of Tlemcen.
Examiner	B. Benmammar	Associate Prof at University of Tlemcen.
Examiner	N. Achir	Associate Prof at University of Paris 13.
Examiner	M. KADDOUR	Associate Prof at University Oran 1 .
Supervisor	F. Didi	Associate Prof at University of Tlemcen.
Co-supervisor	M. Abdennebi	Associate Prof at University of Paris 13.

Abou Bekr Belkaid Tlemcen University

Abstract

Faculty of Sciences
Computer Sciences Departement

Doctor of Philosophy

Management of joint radio resources in heterogeneous networks Beyond 3G

By Bendaoud Fayssal

Heterogeneous Wireless Networks (HWNs) is a term referring to mixed networks made of different Radio Access Technologies (RATs) which aims to achieve the Always Best Connected concept (ABC). Users with multi-mode terminals can connect to different radio access technologies such as 802.16, 802.11, HSPA and LTE at the same time. The challenge consists of the choice of the suitable access network from the available RATs list; the decision process is called the Network Selection (NS). The network selection process depends on several parameters such as QoS, mobility, cost, energy battery etc. Several methods and approaches have been proposed to address this issue. Two main contributions are presented in this thesis, the first one called modified-SAW function to deal with the drawbacks of the legacy SAW, the proposed solution outperforms almost all the famous Multi Attribute Decision Making methods and eliminates the problem of rank reversal. The second contribution consists of presenting a framework that allows the users to select the best networks for the whole call session, especially from a mobility perspective. The framework consists of several steps, starting by the path prediction which is done using the Markov model order 2. The second step is to make the network selection on the zones of each predicted path, while the third step is to get the good RAT configuration for each predicted path. We use another function to select one of the best configurations to be used for all the possible used paths. The results show that our proposal performs very well by eliminating the unnecessary vertical handovers while maintaining a good Quality of Service (QoS).

To the beloved one Muhammad who is praised in the heavens and earth, may abundant peace and blessings of the Creator shower upon him, his family and companions forever.

To my parents, for all their love, kindness and sacrifice,

and

my sisters, uncles and aunties.

Acknowledgement

The quest for a PhD has indeed been a long one. I thank ALLAH first for His abundant blessings, for giving me in particular with the opportunity to pursue a doctorate. I thank Him for supporting me in the face of challenges with His Mercy and Guidance, and for facilitating the completion of this thesis. The Prophet (peace be upon him) has said " Whoever has not thanked people has not thanked ALLAH." I would like to thank a number of people whose support has been crucial in this success.

First of all I would like to thank my director of thesis Dr. FEDOUA DIDI and my supervisor Dr. MARWEN ABDENNEBI for their valuable guidance and support over the years. Their faith in my ability and their encouragement to explore new horizons helped bring out the best in me and i remain indebted to them for this. I thank MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH for awarding me a PhD scholarship to continue my researches in good conditions.

I would like to express my profound gratitude to Mr Azzedine Chikh, Professor of the University of Tlemcen, who has honoured me by accepting to preside the jury of this thesis. I would also like to express my sincere thanks to Dr Badr Benmammar, Associate Professor of the University of Tlemcen, Dr Nadjib Achir, Associate professor at Paris 13 University and Mejdi Kaddour Associate professor at Oran University for the honour they have given me by accepting to be rapporteurs and jury members of the thesis. I thank them for the interest they have shown in my work.

Among family and friends, I first would like to thank my parents for being so patient and supportive in every manner possible, my sisters and my friends for their moral support. I feel a deep gratitude towards all the teachers in the Informatics department of Tlemcen who have contributed on my studies.

Finally I thank ALLAH for gifting me with precious friends like Dr. Jean Thierry Stephen AVOCANH, Dr. Mohamed Amine KAFI, Dr. Mustapha BEKHTI, Dr. Meriem KHELIFI, Ms. Amel ZERZOURI, Ms. Lynda CHAMI and Miss Wiam Benhaddouche. Thank you all for ensuring my smile even during the rough times.

Contents

List of Figures

List of Tables

General Introduction

0.1 Heterogeneous wireless networks HWN

The field of telecommunications has witnessed an explosive evolution in the last decade. It is taking an important and crucial place in the today's societies. The most important innovation of the field of telecommunication is without a doubt the wireless technologies. These technologies are succeeding one to another and offering higher performances such as more bandwidth, less delay and packet loss rate ...etc. At the same time, we are witnessing a convergence of core networking infrastructure on the Internet Protocol IP. From a TCP/IP model perspective, the layer 3 is the important one by being merged with the layers 1 and 2, becoming one layer called IP layer. The result is that IP must support key features associated with wireless networking including handover, location, and mobility management. Besides, due to the variety of choices available from many cellular/wireless network providers, different access technologies, and disparate application's QoS requirements, there is a significant need to come up with a single united approach. This impressive evolution has forced the developers to create a new concept of wireless network called the heterogeneous wireless network or the next-generation network to deal with the heterogeneity between the different technologies.

A Heterogeneous wireless network HWN is a wireless network which consists of devices using different underlying radio access technologies RATs. There are several benefits for a HWN as opposed to a traditional homogeneous wireless network including increased reliability, improved spectrum efficiency, and increased coverage. Reliability is improved because when one particular RAT within the HWN fails, it may still be possible to maintain a connection by falling back to another RAT. Spectrum efficiency is improved by making use of RATs which may have few users through the use of load balancing across RATs and coverage may be improved because different RATs may fill holes in coverage that no single network would be able to fill. However, there are several problems still to be solved in heterogeneous wireless networks such as:

- Determining the theoretical capacity of HWNs.
- Interoperability of technology which means the network selection.
- Handover.
- Mobility.
- Quality of Service / Quality of Experience.
- Interference between RATs.

The next generation of wireless networks are characterized by the existence and the deployment of various wireless access network, for instance, the WLAN family basically the

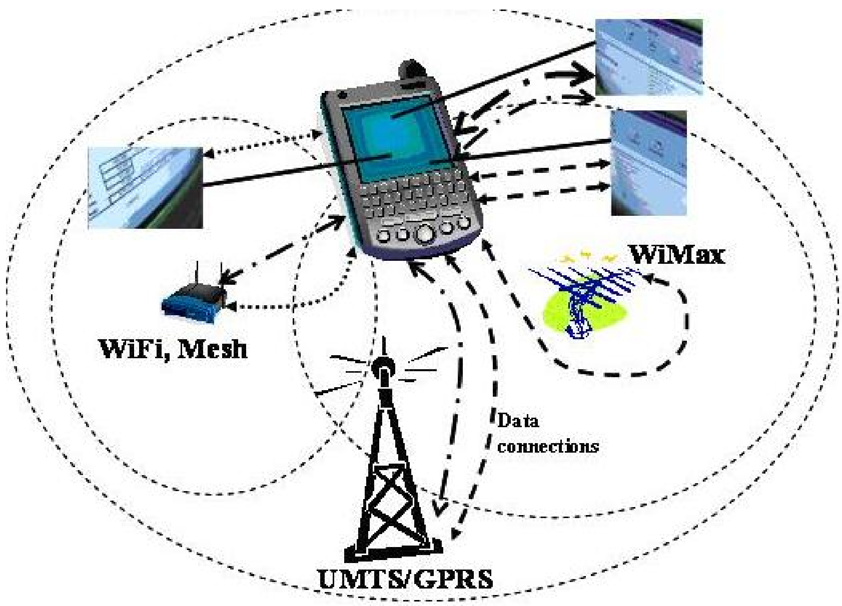

Figure 1: Heterogeneous wireless network HWN

IEEE802.11 standard [1], and the cellular family UMTS, HSPA and the LTE [2]. The variety in radio access creates the elements of the heterogeneous environment. This heterogeneous system has different characteristics such as, network availability, coverage and QoS. The differences between these technologies are basically relative to the speed and movements of the mobile node. One of the fundamental ideas of the heterogeneous systems is to give the users the ability to choose the best network among multiple existing RATs based on several criteria, this choice is called the Network Selection NS, [1] [2] and this thesis will focus basically on this problem.

0.2 Motivations

In the context of heterogeneous networks, users with multi mode terminals can connect to different radio access technologies such as 802.11, HSPA and LTE in the same time. The challenge here is to be best served. This is the primary concern of the mobile users nowadays as they, all of the users try to achieve the Always Best Connected ABC paradigm. To attain the ABC, users must swap between the radio access to maintain the best connectivity, this swap between the radio technologies is called the network selection procedure.

Up to the third generation network, the radio access network was mainly homogeneous. At the time, the network selection concept did not exist because people were bounded by a contract with only one network. After that, emerged the impressive growth of the internet applications and services that led to the development of the network technologies, and a significant expenditure in the mobile user industry. Customers are now equipped with Smartphones and are seeking for the Always Best Connected (ABC concept. It is obvious that no single network technology can sustain that, therefore, it was necessary to change the design, i.e. switching from the homogeneous systems to the heterogeneous one. The aim of the

[1] http://standards.ieee.org/
[2] http://www.3gpp.org/

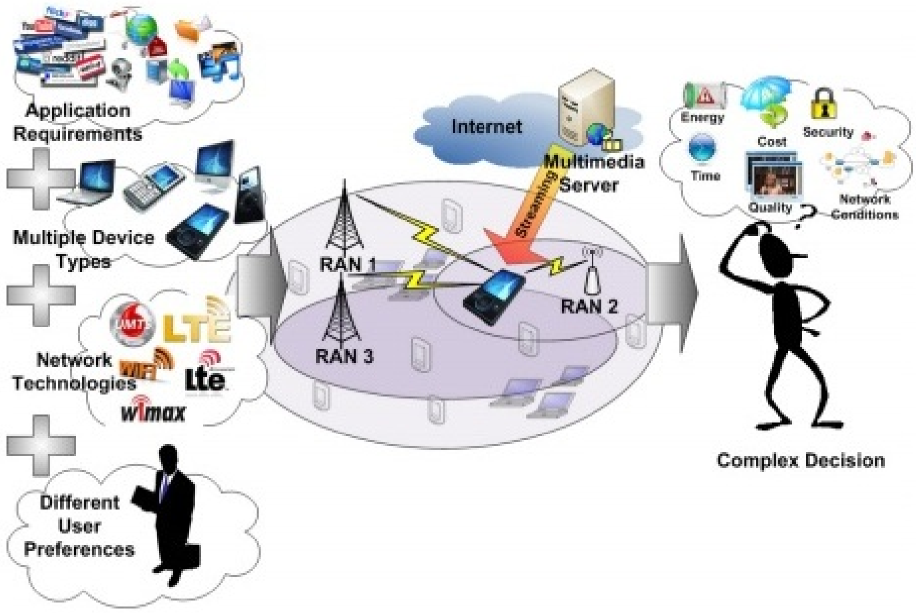

Figure 2: Network selection problem

fourth generation network is achieving the ABC concept by offering mobile users the ability to take advantage of networks having different architectures and performances.

The network selection procedure consists of the selecting of the best available network amongst the available ones [3]. The best network differs from one user to another due to the various preferences involved such as, cost, QoS, energy consumed and even mobility, [4] [5]. The main objective of our work consists on allowing users to achieve the ABC concept by choosing the RAT which provides the best QoS.

0.3 Approaches and contributions

So, as we said earlier, the network selection procedure consists of selecting the best available network among the available ones; this process is complicated because of the large number of parameters involved, cost, QoS, energy consumed, etc. To allow users to achieve the ABC concept by choosing the RAT which provides the best QoS, we started by making a review on method, approaches and strategies used to solve the network selection decision problem.

The network selection procedure can be reported to a Multiple Attribute Decision Making MADM problem. Since it involves multiple parameters (i.e. criteria), so this approach has been widely used to solve the NS problem [5], [1], [2]. Other methods like fuzzy logic and game theory have also been used to solve the network selection problem in [6] [7] [8]. Many studies in the literature propose algorithms that select a network among a set of candidate networks. These algorithms try to find and connect to the best network at time t_0. Selecting the most resourceful network by the user might not ideal because, numerous users can simultaneously choose the same network simultaneously; thus, the network becomes loaded and possibly unreachable. So, it would be interesting to have the optimal rank of the networks considering as many as we can of criteria to have the correct ranking of the available RATs.

In this thesis, two major contributions are presented

- The first contribution consists on the proposition of an algorithm based on the local criterion ranking to deal with the drawbacks of the MADM solutions. Indeed, the existing MADM methods suffer from the famous problem of rank reversal once an alternative is added or removed. Other problems occur in the legacy MADM such as the sensitiveness to the user preference in TOPSIS; the penalizing of the alternatives with poor attributes values in WPM. Our simulations expose the drawbacks of the traditional MADM approach in the context of network selection. The simulations show that our proposal outperforms other existing methods.

- The second contribution consists on proposition of a framework that allows the users to select the best networks for the whole call session especially in the mobility situation. The framework consists of several steps, starting by the path predictions. The latter is done using the Markov model order 2. The second step is to make the network selection on the zones of each predicted path, while the third step is to get the good RAT configuration for each predicted path and finally we use the mean value function to select one of the best configurations to be used for all the possible used paths. The results show that our proposed methodology performs very well by eliminating the unnecessary handover while maintaining a good QoS.

0.4 Organization of the manuscript

This thesis is composed of four chapters, which are divided as follows:

- **The first chapter** called also chapter of generalities, introduces the wireless networks, we present the most important families of wireless network, in this chapter also we make a distinction between traffic flows and their QoS requirements and the last section present the handover concept.

- **The second chapter** presents the state the art of the methods and approaches used to solve of the network selection NS problem. We present the widely used Multiple Attribute Decision Making methods. We also present the application of the game theory to solve the NS problem and we talk about the use of the Fuzzy logic in this case study. We give our critics and remarks of each of the used methods.

- **The third chapter**, the first contribution. It consists of proposing an algorithm called modified-SAW to make the selection of network; the simulations show that this algorithm outperforms the existing ones.

- **The fourth chapter**, we present in this chapter a framework consisting of different steps that allows the users to select the best network for the entire call session especially in the mobility situation. In this chapter, we try to minimize the number of vertical handover which degrades the QoS of the users..

Chapter 1

Radio Access Technologies RATs: An overview

Contents

1.1 Introduction

This chapter introduces the reader to different relevant background concepts in heterogeneous networks which are needed to clearly understand the objectives of this research study. It also provides an overview of some recent and important technological achievements in the wireless communications field which have been adapted by this research study in order to fulfil the research goals outlined earlier in chapter 1.
We start with an introduction of the wireless communication networks in section 1. Section 2 discusses the WLAN networks especially the 802.11 standard. In the section 3, we present the cellular networks family such as UMTS, HSPA and LTE, while in section four, we introduce the QoS and the types of traffic flows. Finally, we explain the handover concept.

1.2 Wireless communication and networks

In 1887, Marconi has demonstrated the ability of the radio wave to provide a continuous contact with the boats passing through the English canal. It was a huge innovation which evolved the ability of people to communicate with moving people. Since then, wireless communication systems have been developing and evolving at an accelerating pace. During the last ten years, the mobile radio communications industry has noticed a huge growth by orders of magnitude. Starting by the digital and Radio Frequency circuit fabrication improvements, the new large-scale circuit integration, and other miniaturization technologies which make the radio equipment smaller, portable, cheaper, and more reliable, [9].

The wireless communication system is considered as one of the biggest innovations in the history of mankind. It involves the transmission of information over a distance without help of wires, cables or any other forms of electrical conductors. The transmitted distance can be anywhere between a few meters (for example, a television's remote control) and thousands of kilometres (for example, radio communication). Nowadays, the wireless communication systems become an essential part of various types of wireless communication devices in which the user is able to communicate even from the remote operated areas. There are many devices used for wireless communication such as mobile phones, the cordless telephones, Zigbee wireless technology, GPS, satellite television and wireless computer parts. Current wireless phones include 3 and 4G networks, Blue-tooth and Wi-Fi technologies; which make them capable to use any kind of wireless communications, [10], [11].

Nowadays, we have two networks families, the IEEE [1] and the 3GPP [2]. For two decades, the IEEE family had dominated the internet market with the IEEE 802.11 standard; it was a big innovation for two things: (a) it was affordable in terms of cost; (b) it offered a high throughput. This standard was and still a milestone. At that time, the 3GPP's cellular technology was the Global System of Telecommunication Mobile GSM. GSM is a good technology for calling and sending SMS, but it cannot provide internet access, so, they develop an updated version of GSM called GPRS, unfortunately the access to internet with higher throughput was not possible either. At this moment the 3GPP's developers have realized that the problem resides in the internal process of GSM, they change the circuit-switched functionality to the packet-switched and they called this release as Universal Mobile Telecommunications System UMTS or the third generation of networks 3G. In the circuit

[1] http://standards.ieee.org/
[2] http://www.3gpp.org/

switched technology, the user billed relative to the antenna occupation' time's rather than the amount of data transmitted since that bandwidth is reserved for the user. However, the packet switched technology allows each user's packets to compete for available bandwidth, GSM is considered as a good technology for calling and sending Short Messages SMS, but it can't provide internet access. This shortcoming of GSM has pushed the 3GPP's developers to update GSM to the Global Packet Radio Service GPRS. Unfortunately, the internet access with higher data throughput was not possible either. At this moment the 3GPP's developers have realized that the problem resides in the internal process of GSM and GPRS. Subsequently, a change was made to the circuit-switched functionality to the packet-switched and called this release as Universal Mobile Telecommunications System UMTS or the third generation of networks 3G. In the circuit-switched technology, the user is billed depending on the antenna occupation time rather than the amount of data transmitted since the bandwidth is reserved for the user. However, the packet-switched technology allows each user transmissions to compete for the available bandwidth, and the user is billed for the amount of data transmitted, thus it utilizes the bandwidth much more efficiently. Moreover, the 3GPP has been upgraded with several enhancements called releases from UMTS until the LTE-A, [12].

1.2.1 Wireless networks

A wireless network is a network in which, at least two devices can communicate without wired connections. It uses the Infra-red or the radio wave to carry the information. Indeed, the radio wave transmission is more widespread due to its broader geographical convergence and higher data rate.

The wireless radio networks are used to connect the remote devices easily in distances ranging from meters to several kilometers. Also the installation process does not require a heavy infrastructure such as the case with the wired networks, which resulted in a rapid development of such technologies. The radio transmissions are used for several types of applications, but they are susceptible to interference; that is why a regulation is needed in each country to define the frequency ranges and power in which it is possible to transmit for each usage category [13].

Each communication between two devices (computer, PDA, wireless printers) without a physical connection (coaxial cable, twisted pair, optical fiber ...) is called a wireless network where each terminal must have a device (network adapter, antenna) to access this network and exchange data with other devices. This network allows users to stay connected while moving in the coverage area.

The wireless local area networks WLAN are the famous form of the wireless network. Their goal is purely economic. Actually, they allow sharing hardware and software resources between posts and users in a distance less than 1 kilometer. The advantage of shared hardware resources is obvious, reducing the needs of printers, drives and even processors. Other types of local wireless networks exist as the difference resides in the served distance and the data rate offered, we cite, the wireless metropolitan area networks WMAN that permit to connect the devices on a city. Another type of the wireless local network is the wireless wide area network WWAN which can serve the whole country [13].

The following features are common to all local networks:

- Symmetrical multipoint connection : Multipoint means that there are several entries on the link. Symmetrical means that the connected devices can talk directly to each other. This notion of symmetry implies developing methods allowing all devices to access medium and sending data.

- Each type of wireless local network has a limited distance that must be sufficient to connect the different devices.

- The networks must provide a high data rate, it depends on the type of network.

1.2.2 Mobile networks

A mobile network is a wireless network that allows simultaneous use of millions of wireless terminals, whether stationary or moving, including when travelling at high speed over a long distance. While, mobile computing is defined as the possibility for users equipped with mobile devices to access services and advanced applications, through a shared network infrastructure, regardless of the physical location or movement of these users.

To maintain communication with moving terminals, all radio access technologies must solve the same problem: distribute as efficiently as possible, a single radio spectrum between numerous users. For this, various multiplexing techniques are used for coexistence and separation of users and radio cells: time division multiplexing, the frequency multiplexing and the code multiplexing, or more often a combination of these techniques.

In the next sections, we are going to provide more details concerning the standards of mobile networks; we will detail the two kinds of mobile networks, the data networks or the Wireless Local Area Networks (WLAN) and cellular networks. In addition, to this we will present the two biggest great existing categories of the wireless networks that are IEEE family of networks; in particular, we will focus on the WLAN, and the 3GPP standard representing the cellular networks.

1.3 Institute of Electrical and Electronics Engineers: IEEE standard

Pronounced I triple E, it is an American organization which regroups together engineers, researchers and students. The goal of the IEEE is developing standards in the field of electricity and electronics to ensure compatibility between the different materials produced by the manufacturers.

The IEEE standards are developed by the technical committees of IEEE societies and by the coordination committees of IEEE Standards Board. Volunteers, the committee members are not remunerated; they must necessarily be members of the Institute. All standards developed in IEEE represent a consensus of the broad expertise on the subject gathered at the Institute, as well as the activities outside the IEEE which wished to participate in the development of the standard.

The existence of an IEEE Standard does not imply that there is no other way to produce, test, measure, purchase and provide the services encompassed by the IEEE standard. The use of IEEE standards is voluntary. Additionally, the point of view expressed at the time of

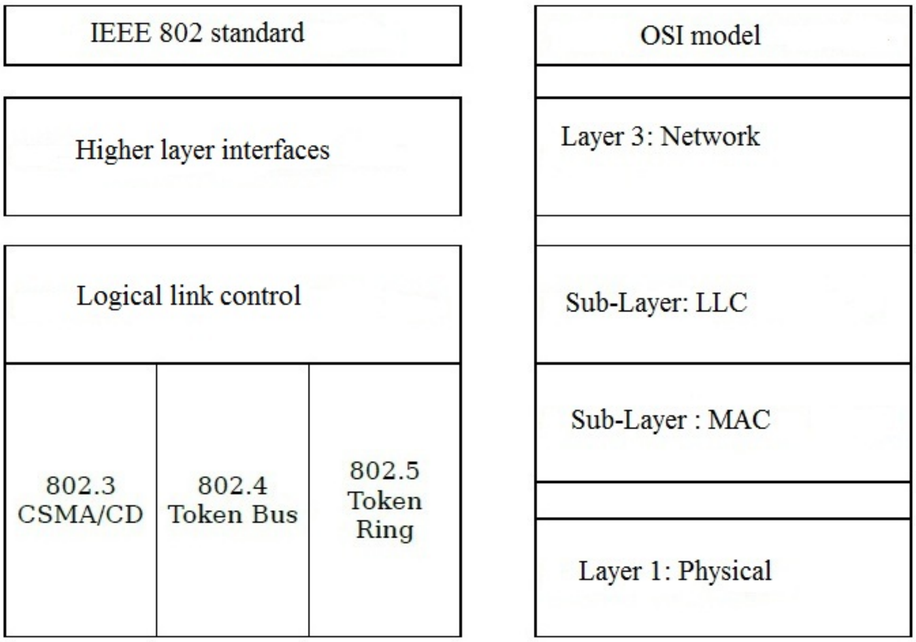

Figure 1.1: IEEE standard with the OSI model

the approval and publication of a standard could change as a result of the advancement of knowledge and / or the feedback received from users of the standard. Each IEEE standard is revised or ameliorated at least every five years. Generally, when a document dating back more than five years and it has not been renewed, it is reasonable to conclude that its contents even if it still has a certain value, do not reflect the latest knowledge. The users should ensure that they have on hand the latest version of the standard [14].

So, for the purposes of renewing and revising its standards, the Institute accepts the comments from any interested party, regardless of its affiliation with IEEE. The suggested changes to a document should take the form of a proposal, accompanied by the relevant comments [15].

Among the IEEE, a committee is responsible for the standardization of local networks. Mainly, the 802 Committee was formed in 1980 to develop standards for the emerging technologies. Their work is limited to the layers 1 and 2 of reference model for the Open Systems Interconnection OSI of the International Standardization Organization ISO. Moreover, the layer 2 (data link) of the OSI model has been divided into two sub-layers which are The medium access control layer MAC and Logical Link Control LLC, Figure 1.1.

1.3.1 WLAN: 802.11

IEEE 802.11 is a set of standards for wireless local area networks that were developed by the Working Group 11 of the standards committee LAN/MAN IEEE 802. The term "802.11x" is also used to refer to this set of standards and not some standard of this set as might suggest the letter "x" usually used as a variable. So there is not, either, only standard 802.11x designated by the term. The term IEEE 802.11 is also used to designate the original 802.11 standard, which is sometimes called the legacy 802.11 [16].

In 1985 and by the United States, The Federal Communications Commission (FCC) had decided to open several bands of the wireless spectrum for use without a government license.

IEEE Standard	Definition and goals
802.1	The management of the local area network and the Virtual LAN VLAN.
802.2	It defines Logical Link Control LLC as the upper portion of the data link layer of the OSI Model.
802.3	It defines the physical layer and data link layer's Media Access Control MAC of the wired Ethernet, CSMA/CD.
802.4	Definition of : Token bus protocol and the CSMA/CA protocol. (disbanded)
802.5	Definition of Token ring protocol
802.6	Advises and specifications on the wide area networks (disbanded).
802.7	Advises and specifications on broadband networks (disbanded).
802.8	Advises and specifications on optical fibre networks (disbanded).
802.9	Integrated Services Digital Network ISDN (disbanded).
802.10	Interoperability of the security mechanisms between LAN/MAN (disbanded).
802.11	Wireless Local Area Networks WLAN: Infra-red, radio waves.
802.12	100VG-AnyLAN, LAN using the concept of Demand Priority,
802.13	Not used.
802.14	Network and TV cables modems (disbanded).
802.15	Wireless Personal Area Networks WPAN: Bluetooth, Zigbee.
802.16	Wireless Metropolitan Area Networks WMAN: WIMAX.
802.17	Resilient Packet Ring RRR.
802.18	Advises and specifications of the radio communications.
802.19	Advises and specifications of the cohabitation with other standards.
802.20	Wireless broadband access.
802.21	handover between networks of the same type as well as handover between different network types.
802.22	Wireless Regional Area Networks WRAN

Table 1.1: IEEE 802.x standards

As a result of this decision, the 802.11 or "Wi-Fi" as it is popularly known has sprung into the existence. The unlicensed bands were mainly the bands allocated to equipment such as microwave ovens which use radio waves to heat food; they were called "garbage bands". To operate in these bands though, some devices were required to use "spread spectrum" technology. This technology spreads a radio signal over a wide range of frequencies, making the signal less susceptible to interference and difficult to intercept. Meantime, the term "IEEE 802.11b-compliant" was considered too long and hard to remember for consumers. As a result, the Wi-Fi Alliance has invented the term "Wi-Fi" that meant nothing at that time, it just sounded like the "hi-fi" which stands for "high-fidelity", used in consumer electronics. This term is meant to indicate that the equipment labelled with this qualification offers superior quality by comparison with common equipment. Later, the meaning "wireless fidelity" was attached to Wi-Fi.

The first IEEE committee called IEEE 802.11 was set up to look for the possibility of getting an open standard started in 1990. Meanwhile, the demand for wireless devices was so high, and in 1997 the standard was published, with regards that the Devices adhering to the new standard were already created.

The IEEE 802.11 standard is actually the original standard published in 1997 (Wi-Fi is the commercial name, and wrongly we talk about Wi-Fi "standards"). Meanwhile, the original standard defines an over-the-air interface between a wireless client and a base station called also access point, or between two or more wireless clients. It offered a throughput of 1 or 2 Mbit/s . Revisions were then made to the original standard in order to increase the throughput, to add other capabilities(in the case of 802.11a, 802.11b, 802.11g, 802.11n and 802.11ac, called physical 802.11) or specify functions safety or interoperability.

The first two published amendments were IEEE 802.11b which operate in the industrial, medical and scientific — ISM — band of 2.4 GHz, and IEEE 802.11a, functioning in the available 5 GHz bands (5.15-5.35 GHz, 5.47-5.725 GHz, and 5.725-5.825 GHz). The third variant, IEEE 802.11g, was approved in June 2003. Both IEEE 802.11a and IEEE 802.11g use a more advanced form of modulation called orthogonal frequency-division multiplexing (OFDM). Using OFDM in the 2.4 GHz band allows the IEEE 802.11g to achieve speeds of up to 54 Mbps. In the following, we present the major revisions of the 802.11 standard and their significance.

- 802.11b called the Wi-Fi, it was considered as the most widespread WLAN standard and it was installed in the early 2000. It offers a theoretical throughput of 11 Mbit/s (6 Mbit/s effective) with a theoretical range of up to 300 meters in an open environment. The frequency range used is the 2.4GHz in ISM band, only 13 radio channels available with maximum 3 non-overlapping. The modulation used is either: CCK, DBPSK or QPSK.

- 802.11a or Wi-Fi 5, it provides a high throughput within a radius of about 10 meters, the offered throughput is in the range of 54 Mbps in theory and an effective throughput of 27 Mbps in the radio frequency band 5 GHz U-NII Unlicensed - National Information Infrastructure. The 802.11a standard specifies 52 sub-carriers, eight combinations not overlapped, are used for the main channel. It uses an adaptive modulation depending on radio conditions: 16QAM, 64QAM, QPSK or BPSK.

- 802.11g standard was released in 2003; it uses the same OFDM modulation method as the IEEE 802.11a standard so as to achieve the theoretical speed of 54 Mbit/s. It works at the same frequency band specified by IEEE 802.11b.

- 802.11n is an amendment that improves upon the previous 802.11 standards by adding multiple inputs and output antennas (MIMO). 802.11n operates on both the 2.4 GHz and the 5 GHz bands. Support for 5 GHz bands is optional. It operates at a maximum net data rate from 54 Mbit/s to 600 Mbit/s. Thus, The IEEE has approved the amendment, and it was published in October 2009. Prior to the final ratification, enterprises were already migrating to 802.11n networks based on the Wi-Fi Alliance's certification of products conforming to a 2007 draft of the 802.11n proposal.

- 802.11ac, published in December 2013, it builds on 802.11n. Changes compared to 802.11n include wider channels (80 or 160 MHz for 802.11ac versus 40 MHz for 802.11n) in the 5 GHz band, more spatial streams (up to eight for 802.11ac versus four for 802.11n), higher-order modulation (up to 256-QAM vs. 64-QAM), and the addition of

Multi-user MIMO (MU-MIMO). In October 2013, high-end implementations support 80 MHz channels, three spatial streams, and 256-QAM, yielding a data rate of up to 433.3 Mbit/s per spatial stream, 1300 Mbit/s total, in 80 MHz channels in the 5 GHz band. [19] Vendors have announced plans to release the so-called "Wave 2" devices with support for 160 MHz channels, four spatial streams, and MU-MIMO in 2014 and 2015 [13].

1.4 3rd Generation Partnership Project: 3GPP standard

The 3rd Generation Partnership Project 3GPP is the cooperation between telecommunication standards which unites seven telecommunications standard development organizations known as Organizational Partners such as the ITU (International Telecommunication Union), ETSI (Europe), ARIB/TTC (Japan), the CCSA (China) ATIS (North America) and TTA (Korea). Basically, the aim of the 3GPP project was to develop globally acceptable specifications for third generation 3G mobile systems to produce and publish the technical specifications for 3G mobile networks (3G) and fourth generation (4G), Figure 2.2 [17] .

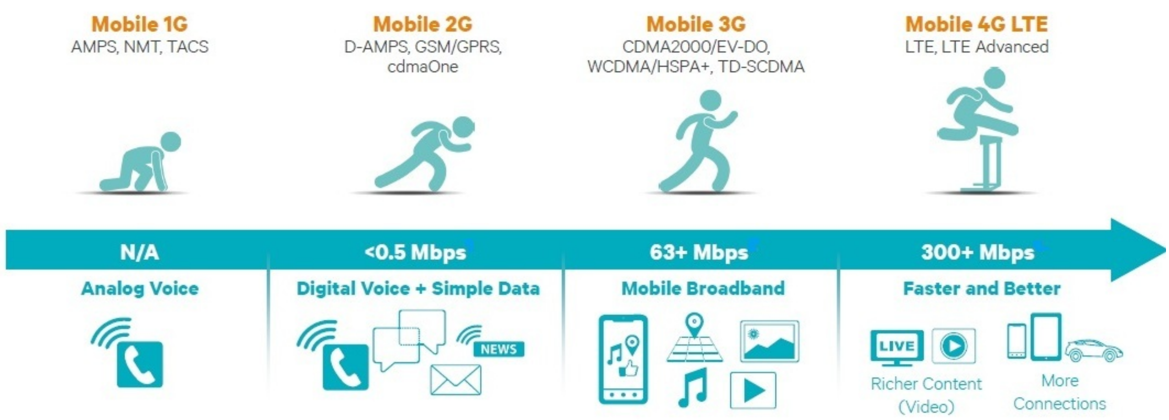

Figure 1.2: From 1G to 4G

The first operational cellular communication system was deployed in the Norway in 1981 and was followed by similar systems in the US and UK. In 1982, the AMPS Advanced Mobile Phone System was invented by Bell Labs and it was installed in the United States. It was also used in England, where it was called TACS Total Access Communication System, and in Japan, it was known as MCS-L1. These first generation systems provided voice transmissions by using frequencies around 900 MHz and analogue modulation.

In AMPS, the cells are typically 10 to 20 km across, whereas in digital systems, the cells are smaller. Each cell uses some set of frequencies not used by any of its neighbours. The key idea that gives cellular systems far more capacity than previous systems is the use of relatively small cells and the reuse of transmission frequencies in nearby (but not adjacent) cells.
The AMPS system has these characteristics:

- It covered 2,100 square miles with only ten base stations.

- Each base station with an antenna tower height between 150 and 550 ft.
- The system was designed for a carrier-to-interference ratio (CIR) of 18dB to be able to satisfy the voice call quality.
- The system was deployed to reuse the frequencies in a 7-cell pattern with 3 sectors per cell.

This standard was aborted in the early of 2000s and it was replaced in the United States by a digital version named the Digital-AMPS 2G D-AMPS which uses the Time Division Multiple Access TDMA. This latter retains the same principle of mobiles' identification by their ESN series number. Other mobile operators prefer migration to the 2G CDMA standard and then to 3GCDMA EVDO at "Verizon Wireless" or UMTS/WCDMA at "AT&T Mobility"; Afterword, early in 2010, towards the 4G LTE, the standard was chosen by all North American operators [18].

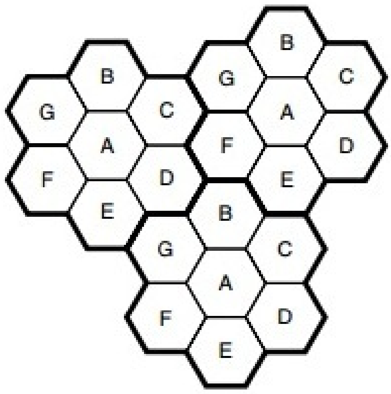

Figure 1.3: Frequencies are not reused in adjacent cells

1.4.1 2G systems: Case study GSM

Examples of 2G digital cellular systems include the Global System for Mobile Communications GSM, IS-95 CDMA, and IS-136 TDMA systems. GSM is by far the most widely deployed of these systems, it was mainly used in Europe; IS-95 is deployed in North America and parts of Asia; IS-54 enhanced to IS-136 that is TDMA-based system that was designed as a digital evolution of AMPS was initially deployed in North America but was later discontinued and replaced mostly by GSM [19].

In 1982, the Conference of European Post and telegraph has regrouped many European countries to standardize and specified a European system for mobile communication. The group was called "Group Special Mobile" GSM. Its aim was to create a mobile system that delivers the voice service seamlessly across all of Europe with an inexpensive price. Because, standards of communications in Europe before GSM were variable and mutually incompatible. In 1989, the European Telecommunication Standard Institute ETSI took over the development of GSM and the first version was released in 1990. At the time, several operators in Europe deployed GSM and gained acceptance beyond Europe. It then was renamed as The Global System of Mobile Communication.

The GSM air-interface is based on a TDMA/FDMA scheme in which eight users are multiplexed on a single frequency channel of 200kHz width by assigning different time slots to each one and the Gaussian Minimum Shift Keying GMSK is used for modulation. The latter was chosen because it provides a good power and spectral efficiency characteristics. It was called the "second generation" 2G telephony standard as the communications use an entirely digital mode, unlike the first generation of mobile phones. It uses the 900 MHz and 1800 MHz frequency bands. However, in the United States, the used frequency bands are the 850 MHz and 1900 MHz bands. Thus, the mobile phones referred as tri-band can operate in Europe, the United States, and dual-band phones operate only in Europe. The GSM standard allows a maximum bit rate of 9.6 kbps, which permits the voice to be transmitted as well as the low volume digital data such as text messages for Short Message Service SMS or multimedia messages for Multimedia Message Service MMS.

The GSM system supports the circuit-switched data CSD that is a method of implementing telecommunication networks in which two network nodes establish a dedicated communication channel called circuit through the network before the nodes may communicate. The circuit guarantees the full bandwidth of the channel and remains connected for the duration of the communication session. The circuit functions as if the nodes were physically connected with an electrical circuit. In circuit switching, the delay is constant during a connection because the circuit cannot be degraded by competing users. Thus, the circuit is protected from use by other callers until the circuit is released and a new connection is set up. Even if no actual communication is taking place, the channel remains reserved and protected from other sessions [20].

By the mid-1990s along with the growth of the demands for a higher data rates, ETSI introduced the GSM Packet Radio Systems GPRS as an evolutionary step for GSM systems. The GPRS system has the same frequency bands, time slots and signaling links as GSM. This enables the coexistence between the two systems. Additionally, GPRS defined four different channel coding schemes supporting 8 kbps to 20 kbps per slot. Under favorable channel conditions, the 20 kbps rate can be used, and if all eight slots in the GSM TDM frame were used for data transmission, in theory, GPRS could provide a maximum data rate of 160 kbps. The typical implementations of GPRS provided a user data rate of 20-40 kbps which is much higher by comparison with the original version of GSM.

Figure 3.4 provides the architecture of the GSM/GPRS system with details of the element of each sub-system.

- **The base Station Subsystem BSS:** It is the traditional section of the cellular networks which is responsible for signaling and traffic management between a mobile phone and the network switching subsystem. The BSS carries out the coding of speech channels, allocation of radio channels to mobile phones, paging, transmission and reception over the air interface and many other tasks related to the radio network.

  It regroups the Base Station Transceivers BTS that the Mobile Station MS connects using the air-interface. The Base Station Controller BSC, which manages the traffic, mobility across the BTS and insures the transportation of packet to the core networks. The Radio Network Controller RNS is the updated version of the BSC used in the 3G systems.

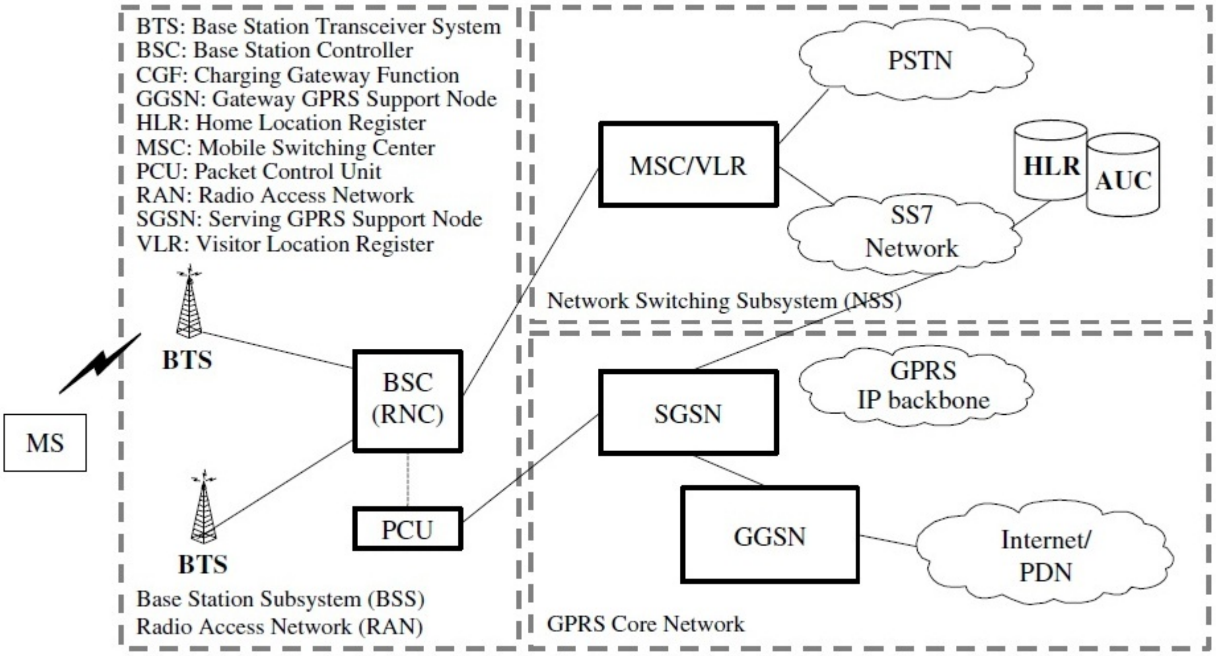

Figure 1.4: GSM architecture

- **The Networks Switching Subsystem NSS :** This component carries out call switching and mobility management functions for mobile phones roaming in the network of base stations. It is owned and deployed by mobile phone operators and allows mobile devices to communicate with each other in the wider public switched telephone network (PSTN). The architecture contains specific features and functions which are needed since the phones are not fixed in one location.
  It is also called the GSM Core Network; it contains the Mobile Switching Center MSC which provides the necessary switching to connect the calling parties. It also determines the location of the callers using the Home Location Register HLR and the Visitor Location Register VLR.

The illustration shows also that the GSM system is compatible with the GPRS that is an updated version, which inserts some new components like: a) the Serving GPRS Support Node SGSN that provides location and mobility management and may be thought of as the packet data equivalent of MSC. b) and the Gateway GPRS Support Node GGSN which provides the IP access router functionality and connects the GPRS network to the Internet and other IP networks; in addition to an upgrade of the existing network elements such as the BTS with a packet control unit PCU to handle data.

At the beginning of 1997, with the aim to boost the GSM standard to provide higher data handling capabilities, the last upgrade of the GSM was introduced; it was called the Enhanced Data Rate for GSM Evolution EDGE. The latter added a support of 8PSK modulation to boost the data rate. This allowed for a maximum per slot data rate of 59.2 kbps, a three-fold increase from GPRS speed. The increase in data speed to 384 Kbps theoretically placed EDGE as an early pre-taste of 3G, although it was labelled 2.75G by industry experts. Typical user rates for EDGE varied from 80 to 120 kbps.

All these upgrades from the original version of GSM to the last release called EDGE have one aim, enhancing the data rate to allow people making voice call and other services like exchanging images, and especially videos. Thus, the 3GPP's developers have realized that the problem resides in the internal process of GSM and GPRS, they change then the circuit-switched functionality to the packet-switched and they called this release as Universal Mobile Telecommunications System UMTS or the third generation of networks 3G. In the circuit-switched technology, the user is billed relative to the radio channel occupation's time rather than the amount of data transmitted since the bandwidth is reserved for the user. However, the packet-switched technology used since the 3rd generation of cellular network up to now allows the competition between packets of different users for the available bandwidth, and the user is billed for the amount of data transmitted; thus, it utilizes the bandwidth much more efficiently.

1.4.2 3G systems : Case study :UMTS and HSPA

The totalities of the 2G wireless systems are voice-centric and some of them support some data over their voice paths, but at painfully slow speeds limited usually between 9.6 Kb/s and 14.4 Kb/s. So, in the world of 2G, voice dominated the activities of these systems while data was already used in the wire-line and some wireless standards of communications. Concurrently, fixed or wireless, all networks were affected by the rapid growth of the Internet.

The expanding for the high data rate have forced the 3GPP's developers to make fundamental modifications in the entire 2G system. They changed the circuit-switched method and adopted the packet-switched technology. This latter represents the heart and the key feature of the 3G systems. Packet-switched networks transmit data in separate and small blocks packets based on the destination address. When received, packets are reassembled in the proper sequence to recreate the original message while the circuit-switched networks require dedicated point-to-point connections during the communication session.

In packet switching networks unlike circuit switching networks, it is not required to establish the connection initially. The channels are available for usage by many users. But when the capacity or the number of users increases then a situation of congestion appears in the network. The Packet switched networks are mainly used for data application and they keep the best performance of voice application as the 2G systems. Figures 1.5 and 1.6 display the difference between the two concepts of circuit and packet switching[21].

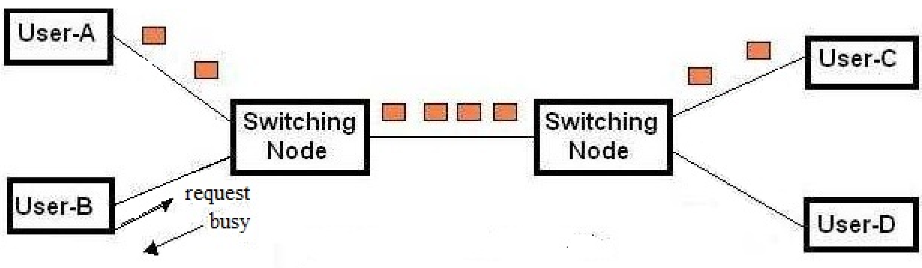

Figure 1.5: Circuit-switched network

So, providing higher data rate, making a significant increase in voice capacity, and supporting advanced services / applications, including the multimedia services were the main objectives of the work group of the 3GPP. The work for the third generation of network had begun in the early 1990s. The International Telecommunications Union ITU has developed a set of specifications for the 3G systems called at that time the IMT-2000 and started

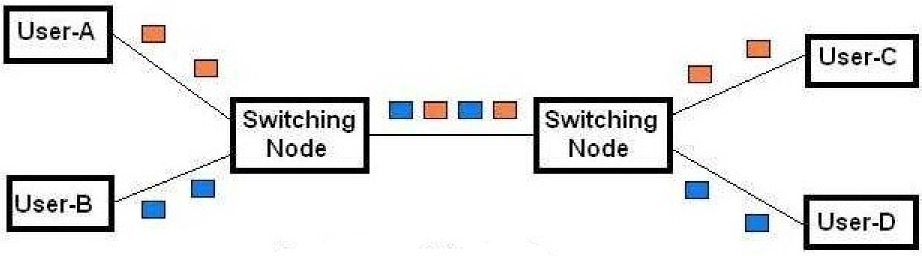

Figure 1.6: Packet-switched network

identifying spectrum for it. The objective was to create a global harmonized specification for mobile communication that facilitates the global interoperability and provides the scale to lower cost. The ITU laid out the following data rate requirements as the criterion for IMT-2000: 2 Mbps in fixed or in building environments, 384 kbps in pedestrian or urban environments and 144 kbps in a wide area or mobility environments.

The Universal Mobile Telecommunication System developed by ETSI as the 3G system for IMT-2000 based on the evolution of GSM; it is considered as the most well known standard of the 3G networks. Many proposals were submitted to the ITU and only few of them have been accepted to be 3G standards, the proposals were different, but the majority of these standards have selected the Code Division Multiple Access CDMA as the access technique because it is more suitable for the packet switched mode. The UMTS standards have proposed and used the Wideband CDMA WCDMA. The WCDMA allows sending all the data packets simultaneously in a disorder way on any frequency and it is for the mobile phone to receive the packets and gathers them in the correct order. This technology allows to transmit simultaneously more data and thus, offering a much higher data rate than those permitted by GSM and GPRS. In theory, the data rate is 2 Mbps from a fixed location and 384 kbps in mobility [22].

Each communication system implies that the data transmission must be possible towards the base station and from it. This can be achieved through two ways, transmitting data on one frequency and receiving it in another. The use of different frequencies between the two transmissions is made to avoid signal interferes. This is known as frequency division duplex FDD and it is one of the most commonly used schemes by cellular schemes such as UMTS-FDD. It is also possible to use a single frequency rather than using different frequencies allocations but we use different time allocations. The transmission times are split into slots, then transmissions in one direction that takes place in one time slot, and those in the opposite direction take place in another. This scheme is known as time division duplex TDD, and it is used for UMTS-TDD.

The UMTS network is composed of two sub-systems: the UMTS Terrestrial Radio Access Network UTRAN and the core network. The UTRAN is mainly responsible for transferring the data generated by the user. It is a gateway between the user equipment and the core network via the Uu and Iu interfaces while the core network provides the switching, routing, and subscriber management. The UTRAN is also responsible for security, mobility and the radio resource management. It is composed of base station called NodeB and Radio Network Controllers RNC which communicate using some communication interfaces.

- Uu: it allows communication between user equipment and the UTRAN.

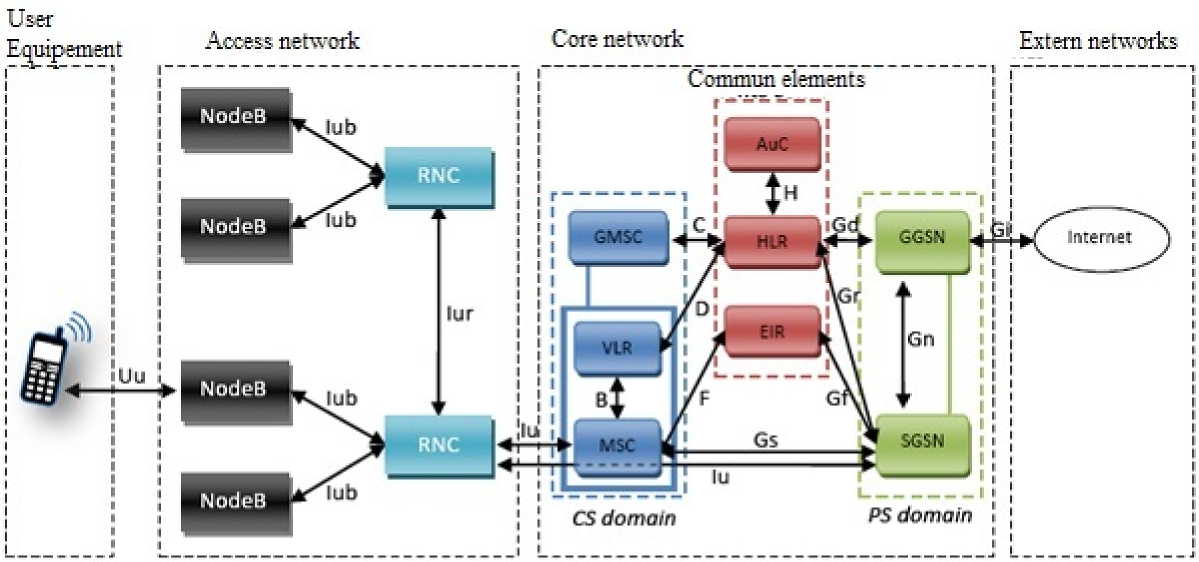

Figure 1.7: UMTS architecture

- Iu: it allows the communication between the RNC from the UTRAN and the SGSN from the core network.

- Iur: it allows the communication between two RNCs.

- Iub: it allows communication between Nodeb and the RNC.

The core network is composed of two parts, the circuit switched part that is responsible for the telephony and voice communications, and the packet switched part that allows the data communications.

With all these enhancement over the legacy GSM, the 3GPP's developers have make two other key enhancements over the UMTS called High Speed Downlink Packet Access HSDPA and High Speed Uplink Packet Access HSUPA, the combination of these key enhancements is called HSPA or 3G+. HSPA was deployed as a software upgrade to existing UMTS systems and by February 2010, it has been deployed by 303 operators in 130 countries around the world. To enhance the data rate, HSPA has defined a new downlink channel called the High-Speed Downlink Shared Channel HS-DSCH able to provide a theoretical data rate up to 14.4 Mbps. HSDPA uses 16 Walsh codes, 15 of them can be used for user traffic. A transmitter could use 5 to 10 codes to get higher throughputs, oppositely to the WCDMA that used time division multiplexing as the multiple access technique with a limited code division multiplexing. This new channel uses a frame length of 2 ms compared to frame lengths of 10, 20, 40, or 80 ms used by WCDMA channels. The practical deployments of HSDPA provided a typical user throughput in the rage of 500 kbps to 2 Mbps.

To achieve the high data rate, the HSPA has introduced some new techniques, this include:

- **Adaptive Modulation and Coding AMC:** It is an adaptive modulation and coding technique used in radio transmission and mobile telephony networks such HSDPA, WiMAX, LTE. AMC or link adaptation consists of varying the modulation and coding scheme per user and per frame basis depending on instantaneous downlink channel quality. For example, if the available modulations are QPSK and 16-QAM, if the channel is marked as good, the 16-QAM modulation is used, which provides better throughput but lower robustness. On the other hand, if the channel is marked as degraded, QPSK modulation will be used, allowing a lower throughput, but more robust and less sensitive to interference. The HSDPA uses the Channel Quality Indicator CQI to enable the base stations to select the best possible modulation and coding scheme.

- **Fast Dynamic Scheduling FDS:** A dynamic scheduler can ensure that the system is always operating at the highest possible rate; in fact, instead of scheduling users at fixed periods in time like in legacy systems, HSDPA uses a dynamic scheduler that attempts to exploit the diversity of channel conditions experienced by different users at different times. A dynamic scheduler could allocate all the cell capacity to a single user for a very short time when conditions are favourable. To use the Fast Dynamic Scheduling, HSDPA implements the scheduling functionality in the NodeB as opposed to the WCDMA where the scheduler is situated at the RNC.

- **Hybrid Automatic Repeat Request H-ARQ :** H-ARQ is a technology that allows the reliable transmission of data over a communication channel that can cause transmission errors. HARQ combines the principles of retransmission defined by the Automatic Repeat Request ARQ, and Forward Error Correction FEC. In the case of the simple ARQ, certain redundant bits are added to the data before being transmitted to enable the receiver to detect possible errors of transmission, for example by using cyclic redundancy control. If the receiver detects an error, it may request retransmission of the data. In the case of the HARQ, the data is encoded by means of a Forward Error Correction code, which also adds redundant bits to the original data and allows correcting a certain number of errors without retransmission. If the number of errors is too high to be corrected by the FEC, the receiver requests a retransmission.

Other enhancements have been applied to the HSPA network, namely HSPA+ or H+ or even 3G+, these enhancements attempt to improve the data rate. In June 2007, the HSPA release 7 is introduced, referred to as HSPA+, which contains a number of additional features that improve the system capacity, end-user throughput and latency. The enhancements consist of the introduction of new major features, these small improvements to existing structures which, when taken together as a package, represent a major increase in performance and capabilities. The key technical enhancements included in HSPA+ are:

- **Multiple Input Multiple Output MIMO :** It is one of the major features of the 3GPP's release 7, it allows increasing data rate through the use of multiple transmission streams. MIMO means the use of multiple antennas for transmitting and receiving data with supporting other enhancements such as the open closed loop transmission diversity and the spatial multiplexing. The MIMO schemes are designed to exploit certain properties in the radio propagation environment to attain high data rates by transmitting multiple data streams in parallel. However, to achieve these high data rates, a correspondingly high carrier-to-interference ratio at the receiver is required. The spatial multiplexing therefore is mainly applicable in smaller cells or close to the base station. Besides, the MIMO 2 * 2 allows a data rate up to 28 Mbps.

- **High order modulation** MIMO amplifies in peak data rate by exploiting the propagation conditions in the channel through multi-stream transmission. However this is not sufficient if the User Equipment or the base station are not equipped with multiple antennas. So, the high order modulation was introduced for the downlink channel up to

64QAM and 16QAM for the uplink. When using the 64QAM modulation and MIMO, the theoretical achieved data rate is 42 Mbps. In real situation it is much lower.

- **Discontinuous transmission and connectivity :** Packet data traffic is often highly bursty with occasional periods of transmission activity; the 3GPP's release 6 called HSPA forced the mobile terminals to transmit the physical control of the channel even in the absence of any data transmission which causes unnecessary battery drain. In the release 7, HSPA+ allows the downlink and the uplink transmissions for the UE to be completely discontinuous which means that the transmitter is tuned-of when there is no data to be transmitted. Discontinuous transmission and reception are very useful and power-saving techniques for bursty data applications and allows reducing the interference especially for the uplink direction which results in increasing of the network capacity. For example in the Web browsing service, it allows to save up to 50% of UE energy while for the VoIP calls, this could provide up to 50% increase in VoIP capacity compared to Release 6.

- **Dual-carrier downlink operation :** HSPA+ defines the use of dual-carrier operation in the downlink direction. It permits to double the peak data rate from 21 Mbps to 42 Mbps as well as doubling the average data rate and substantially increasing the overall cell capacity.

1.4.3 Long Term Evolution: LTE

All the improvements made over the original GSM network aimed to increase the capacity of the system, getting higher data rate and lesser latency. Further improvements have been achieved on the HSPA+ in the release 8 which defined the Long Term Evolution LTE.

The Long Term Evolution of 3G is a project aimed at drafting the technical standards of the future fourth generation mobile phone. It allows the transfer of data at a very high rate, with a greater range, a higher number of calls per cell and a lower latency. In theory, it permits to reach rates in the order of 50 Mbps in uplink and of 100 Mbps in downlink, to be shared between the mobile users of the same cell. For operators, the LTE involves modifying the core of the network and the radio transmitters; the subsequent network is called System Architecture Evolution SAE [23].

The LTE technology or 4G is based on all-IP transport network. The voice call is provided using the VoIP technology, unlike the 3G that carries voice in circuit mode and data using packet switching mode. LTE uses a bandwidth between 1.4 to 20 MHZ which allows LTE to achieve a peak data rate of 300 Mbps while the true 4G offers a peak downlink data rate of 1 Gbps. LTE offers several enhancements for the mobile users as well as to the service providers. It is designated as the successor networks 3G. It allows an efficient management of internet services emerging in the recent years. The Orthogonal Frequency Division Multiple Access OFDMA is the multiple access method used in the downlink direction. It combines Time Division Multiple Access TDMA and Frequency Division Multiple Access FDMA. It is derived from OFDM multiplexing, but it allows the multiple access of the radio resources shared among multiple users. For the uplink direction, Single Carrier-Frequency Division Multiple Access SC-FDMA method is used which is a variant of OFDMA. The two methods have the same performance (throughput, efficiency ... etc.), but SC-FDMA transmits sub bands sequentially to minimize the Peak -to- Average Power Ratio PAPR (OFDMA has a huge PAPR) which is the reason of choosing SC-FDMA in the uplink side to deal with the limited power budget (the use of battery by the UE) and to minimize the PAPR [24].

The general architecture of LTE, also known as the Evolved Packet System EPS, consists of two parts: the core network part, which is the Evolved Packet Core EPC and the radio network evolution part called the Evolved UTRAN E-UTRAN. The EPC can also be connected to other 3GPP and non-3GPP radio-access networks. It consists of some control-plane nodes, such as the Mobility Management Entity MME, control Home Subscriber Server HSS and two user-plane nodes, called Serving Gateway S-GW and Packet data Network Gateway P-GW. The LTE radio access network consists of the base stations, denoted as enhanced NodeB eNb, which are connected to each other through the X2 interface and connected to the EPC through the S1 interface. The mobile terminal is denoted as User Equipment UE. Figure 3.3. The elements constituting the LTE are described in the following

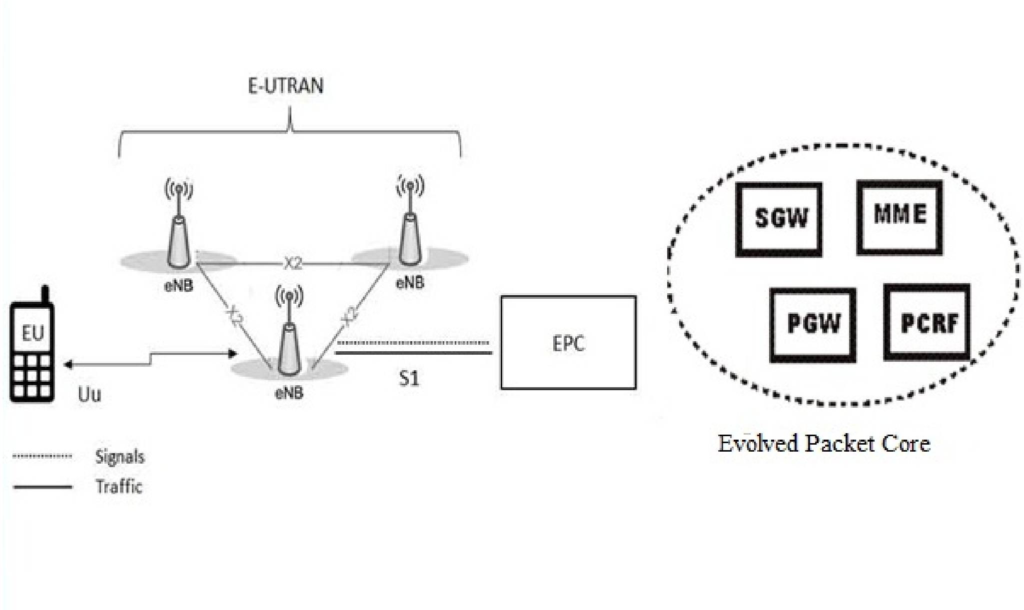

Figure 1.8: Architecture of LTE

- **Enhanced Nodeb eNb:** It represents the base station which is responsible of: radio resource management including the Radio Admission Control, Connection Mobility Control and Dynamic allocation of resources to UEs in both uplink and downlink; it is also in charge of IP header compression and encryption of user data stream, the selection of an MME at UE attachment, routing of user plane data towards SAE Gateway, measurement reporting configuration for mobility and scheduling. The eNb is in control of an important QoS task which is the efficient performance of radio resource allocation.

- **Serving Gateway S-GW:** It ensures the routing of packets data and it manages the inter eNb handover as well as the mobility between LTE network and other 3GPP networks. It also performs replication of the user traffic in case of lawful interception.

- **Packet-data Network Gateway P-GW:** It connects the UE to the external networks by being the point of exit and entry of traffic for the UE. It acts as the anchor for mobility between 3GPP and non-3GPP technologies such as WiMAX and 3GPP2 like CDMA 1X and EvDO.

- **Mobility Management Entity MME :** It has a crucial role in the LTE access network. It selects the gateway through which the messages will be exchanged; it determines level of resources for the UE in cases of attachment and handover; it also provides authentication / authorization in addition to location tracking using the HSS and intra-3GPP mobility.

In this section, we have traced the evolution of the 3GPP family of networks from first generation analogue voice systems to the development of LTE. Now, we will summarize the enhancements and performance improvements that have been achieved at each step of this evolution. The first and the second generations were voice-centric systems; the second generation was a milestone because of the transition from the analogue to the digital system as well as the use of the circuit switched technology that allows making vice conversations. The needs of higher data rate to manage the multimedia services especially the video service had forced the 3GPP developers to add the packet switching technology to exchange packets. The packet switching technology is considered as a huge innovation and enhancement that separates the 2G from the 3G; after that several releases have been proposed where each one provided enhancements in one or multiple aspects including: (a) radio performance improvements such as higher data rates, lower latency and increased voice capacity, (b) core network changes aimed at reducing its complexity and improving transport efficiency, and (c) support for new applications such as push-to-talk, multimedia broadcast, multicast services and IP Multimedia Services. The following Table summarizes the enhancements of the different 3GPP releases.

Table 1.2: 3GPP standard releases

3GPP releases	Year	Major enhancements
Phase 1	1992	Introduction of GSM.
Phase 2	1995	Enhanced full rate EFR codec.
Release 96	1996	GSM Features, 14.4 kbit/s User Data Rate.
Release 97	1997	Introduction of GPRS.
Release 98	1998	Introduction of EDGE.
Release 99	2000	Definition of the original version of WCDMA air-interface.
Release 4	2001	Taking a steps toward using IP transport in core network.
Release 5	2002	Introduction of the IP Multimedia Services IMS, the definition of the HSDPA.
Release 6	2004	Specified : HSUPA, Multimedia Broadcast/ Multicast Services MBMS. Added advanced receiver specifications, push-to-talk over cellular PoC and other IMS enhancements, WLAN interworking option, limited VoIP capability.
Release 7	2007	Definition of HSPA+ : higher order modulation 64QAM downlink and 16QAM uplink and MIMO. Also, reducing latency and improved QoS for VoIP.
Release 8	2008	Other enhancements on HSPA+: combination between MIMO and 64QAM and dual-carrier with 64QAM. Specification of a new

Table 1.2 – 3GPP standard releases

3GPP releases	Year	Major enhancements
		OFDMA-based LTE radio interface and a new all IP flat architecture with Evolved Packet Core EPC.
Release 9	2009	SAES Enhancements, WiMAX and LTE/UMTS Interoperability. Dual-Cell HSDPA with MIMO, Dual-Cell HSUPA. LTE HeNB.
Release 10	2011	LTE Advanced fulfilling IMT Advanced 4G requirements. Backwards compatible with release 8 (LTE). Multi-Cell HSDPA (4 carriers).
Release 11	2012	Advanced IP Interconnection of Services. national Service layer interconnection between operators/carriers as well as third party application providers. Heterogeneous networks HetNet improvements, Coordinated Multi-Point operation CoMP. In-device Co-existence IDC.
Release 12	2015	Enhanced Small Cells (higher order modulation, dual connectivity, cell discovery, self , configuration) Carrier Aggregation (2 uplink carriers, 3 downlink carriers, FDD/TDD carrier aggregation), MIMO (3D channel modelling, elevation beam-forming, massive MIMO), New and Enhanced Services (cost and range of MTC, D2D communication, eMBMS enhancements).
Release 13	March 2016	LTE in unlicensed, LTE enhancements for Machine-Type Communication. Elevation Beam-forming / Full-Dimension MIMO, Indoor positioning.
Release 14	June 2017	Energy Efficiency, Location Services LCS, Mission Critical Data over LTE, Mission Critical Video over LTE, Flexible Mobile Service Steering FMSS, Multimedia Broadcast Supplement for Public Warning System MBSP, enhancement for TV service, massive Internet of Things, Cell Broadcast Service CBS, Cell Broadcast Service CBS.
Release 15	Sept 2018	Support for 5G Vehicle-to-x service,IP Multimedia Core Network Subsystem (IMS), Future Railway Mobile Communication System.

1.5 QoS and types of traffic flow

In this section, we will discuss the Quality of Service QoS concept and the different traffic flows such as the Voice over IP VoIP, the video service especially the video conference and finally the best effort service like web navigation, mail service, etc.

1.5.1 Quality of Service QoS

QoS is the ability to uphold a given type of traffic in good conditions in terms of availability, throughput, transmission delays, jitter and packet loss rate. Quality of service is a concept that aims to optimize the resources of a network and to guarantee the good performance of the critical applications such as the real time application. The quality of service makes it possible to offer users rates and response times differentiated by applications (or activities) according to the protocols implemented at the level of the structure. Quality of service is particularly important for the transport of traffic with special requirements. In particular, developers have introduced technology to allow computer networks to become as useful as telephone networks for audio conversations, as well as supporting new applications with even rigours service requirements.

To quantify the performance of a communication network, many metrics can be used such as:

1 The bandwidth : Which measures the amount of data that can pass through a point in a network over time.

2 The latency : Which is a measure of the delay traffic experiences as it is forwarded in the network and the measure of the variation in that Delay called jitter.

3 Packet Loss Rate : The rate at which a packet is lost. This should also be as minimum as possible.

4 Packet Error Rate: This is the errors which are present in a packet due to corrupted bits. This should be as minimum as possible.

5 Reliability :The availability of a connection. (Links going up/down).

Each application has special requirements of the network which is running on. For example interactive audio i.e. VoIP requires low delay and jitter of its communication channel to support natural and conversational interaction, however it has relatively minimal bandwidth requirements. In contrast, video streaming requires high bandwidth and low jitter to provide a smooth viewing experience, though it has relatively minimal latency requirements; it's OK if the video takes a number of seconds to start, as long as it runs smoothly once it starts. In further contrast, interactive video i.e. video conferencing is the most demanding in which it requires high bandwidth, low latency, and low jitter, combing the challenges of the previous two examples. The defining characteristic of these QoS requirements are manifested in involving constraints on multiple performance metrics. The best effort services such as web navigation and mail service, have no QoS constraints as they tolerate delays and low bandwidth Figure 1.10.
The components in the QoS architecture are the following, we will detail the important ones which are: traffic specification, call admission control (CAC), resource reservation and packet scheduling (PS) [25]

1 **Traffic specification:** specifies source traffic characteristics and desired QoS. Traffic modelling plays an important role in QoS provisioning. It facilitates traffic specifications and accurate call admission control. Without a traffic model or characterization, measurement-based admission control needs to be employed with reduced accuracy and efficiency, compared to traffic-specification based admission control.

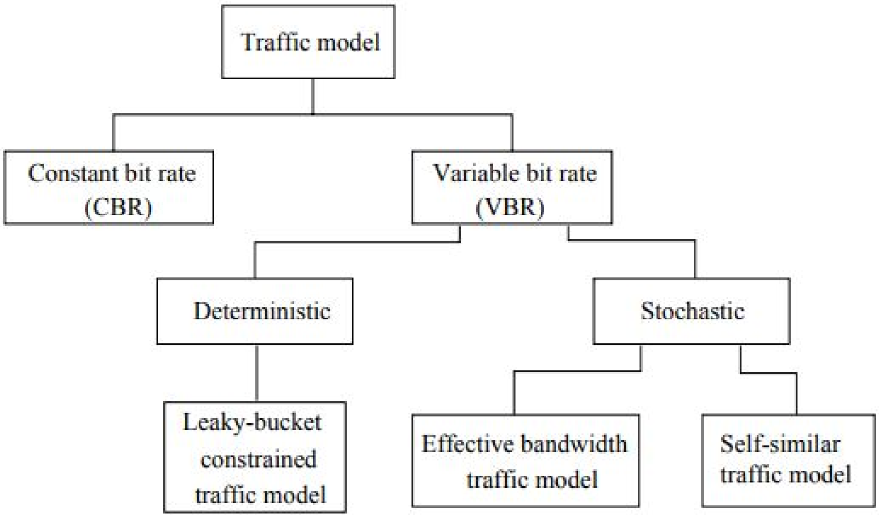

Figure 1.9: Traffic classification

2 **QoS routing:** provides route(s) between source and destination(s) that have sufficient resources to support the requested QoS.

3 **Call admission control:** decides whether a connection request should be accepted or rejected, based on the requested QoS and the network status. The objective of call admission control is to provide QoS guarantees for individual connections while efficiently utilizing network resources. Specifically, a CAC algorithm makes the following decision, given a call arriving to a network, can it be admitted by the network, with its requested QoS satisfied and without violating the QoS guarantees made to the existing connections.

The decision is made based on the availability of network resources as well as traffic specifications and QoS requirements of the users. If the decision is affirmative, necessary network resources need to be reserved to support the QoS. Hence, CAC is closely related to channel allocation, base station assignment, scheduling, power control, and bandwidth reservation. For example, whether the channel assignment is dynamic or fixed will result in different CAC algorithms.

The CAC problem can be formulated as an optimization problem, indeed, we must maximize the network efficiency/revenue subject to the QoS constraints of connections. The QoS constraints are signal-to-interference ratio (SIR), bit error rate (BER), call dropping probability, or connection-level QoS. In other words, a CAC problem is maximizing the number of users admitted or minimizing the blocking probability, subject to the BER violation probability not more than a required value ϵ.

4 **Resource reservation:** allots resources such as wireless channels, bandwidth, and buffers at the network elements, which are required to satisfy the QoS guarantees.

5 **Packet scheduling:** is to schedule packets to be transmitted according to the QoS requirements of the connections. Packet scheduling is an important QoS provisioning mechanism at a network node. Compared to the scheduler design for the wired networks, the design of scheduling for wireless networks with QoS guarantees, is particularly challenging. This is because wireless channels have low reliability, and time varying signal strength, which may cause severe QoS violations. Further, the capacity of a wireless channel is severely limited, making efficient bandwidth utilization a priority.

In wireless networks, the packet scheduling main function is to schedule such resources as time slots, powers, data rates, channels, or combination of them, when packets are transmitted. While the wired scheduler does not assign powers, data rates, and channels since packets are transmitted at a constant power, a constant data rate or link speed, and through one shared channel. Specifically, based on the source characteristics, QoS requirements, channel states, and perhaps the queue lengths, a wireless scheduler assigns time slots, powers, data rates, channels, or combinations of them, to the packets for transmission.

One of the features of wireless scheduling with QoS guarantees is its channel state dependency which means that the resources are scheduled depending on the channel state. This is necessary since without the knowledge about the channel state, it is impossible to guarantee QoS! A key difference between a wired scheduler and a wireless scheduler is that a wireless scheduler can utilize asynchronous channel variations or multi-user diversity while a wired scheduler cannot [26].

6 **Wireless channel characterization:** specifies the statistical QoS measure of a wireless channel, e.g., a data rate, delay bound.

1.5.2 VoIP: Voice over IP

VoIP is currently one of the most important evolutions in the field of Telecommunications. Before 1970, networks were dedicated to the telephony and the voice was transmitted analogically using an electromechanical technology. In the 1980s, the major enhancement was the transition to digital transmission TDM. The transmission of voice over IP packet-switched networks constitutes now a major development by comparison with the previous networks. Voice over IP is an emerging technology of vocal communication; it is a methodology for the delivery of voice communications and multimedia sessions over Internet Protocol IP. The Internet telephony specifically refers to the provisioning of communications services (voice,SMS, voice messaging) over the public Internet network, rather than via the public switched telephone network PSTN.

VoIP is not a protocol rather a generic name including a set of methods and protocols for encoding, transporting and routing audio calls. VoIP calls are similar to traditional digital telephony by involving signalling, digitization of the analogue voice signals and encoding. The difference resides in the transmission way instead of transmitting the digital signals over a circuit-switched network; however, they are gathered, and the transmission occurs as IP

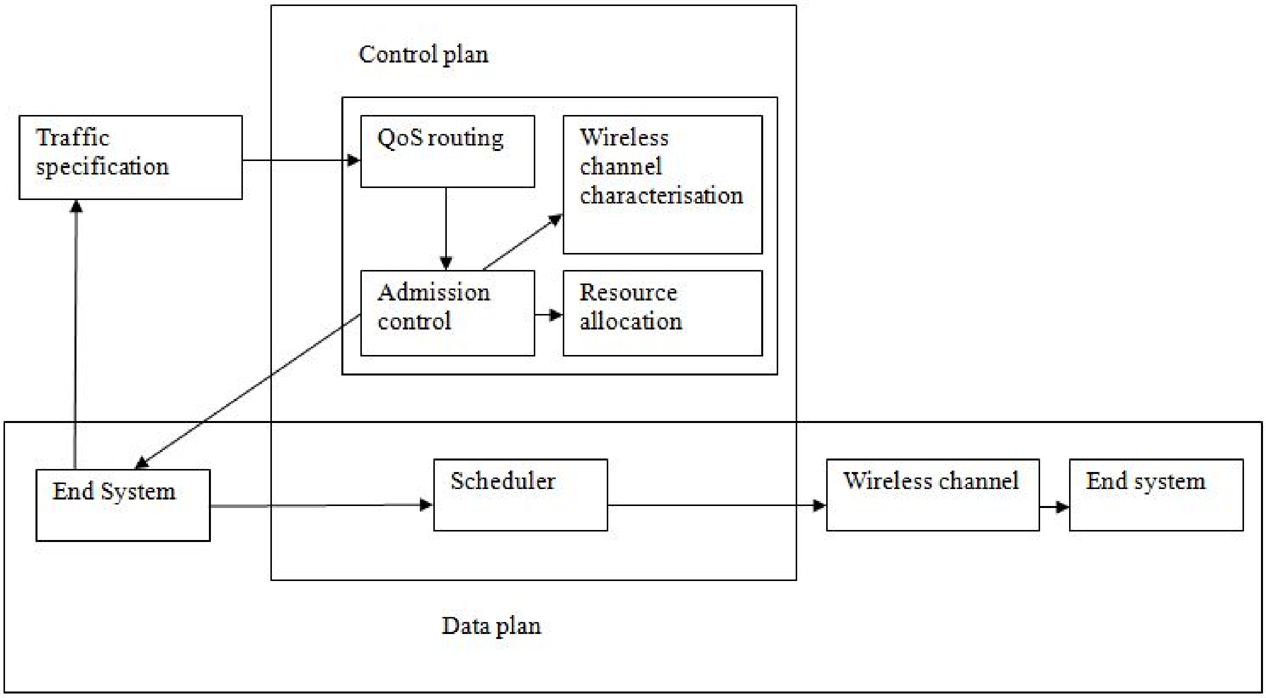

Figure 1.10: Traffic classification

packets over a packet-switched network. There are several protocols that support voice over IP such as the H.323, Session Initiation Protocol SIP and Media Gateway Control Protocol MGCP. The most common protocol currently used in VoIP solutions on the market is SIP. The transportation of audio streams is done using special media delivery protocols that encode audio and video with audio and video codecs. Various existing codecs optimize the media stream based on application requirements and network bandwidth; some implementations rely on narrowband and compressed speech, while others support high fidelity stereo codecs. Some popular codecs include G.711 and G.722, are popular and a widely used open source voice codec known as Internet Low Bit-rate Codec ILBC. This family of codec only uses 8 kbit/s each way called G.729. Now, VoIP is available on many smart-phones, personal computers, and on Internet access devices. Calls and SMS can be sent over 3G/4G or Wi-Fi networks [27].

The VoIP technology has many advantages such as the low cost and its ease to use just like the traditional telephonic technology. In addition to voice, other information such as video can be transmitted at the same time and at no extra cost, consequently, users get better quality in their international phone calls and videos, whether it's free or at the cheapest prices, especially with high quality providers. In the same time, it presents some drawbacks and the most known one is that the quality of the conversation is not always optimal even with the considerable progress made in this area. The quality often depends on the operator and the network radio condition. The architecture of VoIP technology is depicted in the following figure.

1.5.3 Video-conferencing service

The video-conferencing is a set of interactive telecommunication technologies that allow two or more different locations to interact simultaneously using video and audio transmissions.

In the late 90s, video-conferencing had made significant progresses in business, education, medicine and media due to the introduction of relatively low cost and high capacity broadband telecommunication services coupled with powerful computing processors and video compression techniques [28].

Recent advances in computer and telecommunications technologies are interested in compressed video systems. For instance, the vast amount of information in a normal TV transmission is squeezed into a fraction of its former size by a codec Coder-Decoder (video conferencing hardware that codes the outgoing video and audio signals and decodes the incoming signals). Some information is sacrificed in the process, which may result in diminished picture and sound quality. A video-conference system must have audio-visual equipment (monitor, camera, microphone and speaker) as well as the transmission hardware between sites. A broadband satellite connection which is a high-capacity communications path usually implies a speed greater than 1.54 Mbps with studio-quality equipment, it produces an excellent full-motion video connection, but the equipment and transmission expense is massive.

1.5.4 Best effort services

Best effort is a single service model in which an application sends data whenever it must, in any quantity, and without requesting permission or first informing the network. For best-effort service, the network delivers data if it can, without any assurance of reliability, delay bounds, or throughput. Best-effort service is suitable for a wide range of networked applications such as general file transfers or e-mail.

1.6 Mobility and Handover

Mobility

Mobility is the key defining feature of all wireless networks which differentiates them from the wired ones. Since their creation, the smooth execution of handovers has been an important factor in gauging the performance of wireless networks and this is due to the mobility of users .

Wireless networks are composed of multiple geographical units called cells. Each cell provides wireless coverage is administered by a single base station. Sometimes, mobile users might need to have ongoing connections transferred between various base stations to maintain active sessions a process labelled handover. Successful handover are possible when sufficient resources can be granted to an ongoing session by a new network access point. Or else, the session will be prematurely terminated/dropped, due to lack of resources, available with a new cell.

Handover process

Nowadays, the most important feature of all wireless network systems which differentiates them from the wired ones is the mobility. To achieve a seamless mobility, the wireless networks have to implement the handover mechanism. The Handover refers to the operations carried out to allow a mobile phone (called mobile station MS) to change the radio cells without interruption of the voice conversation or the data transfer. The handover process allows a mobile terminal to maintain the current communication during the movement, in which the mobile change the cell. Indeed, when the transmission signal between a mobile station and the base station is weakened, the handover mechanism searches for another reachable base station, which is capable of re-communicating in the best conditions [29].

The handover is necessary in three cases to ensure the continuity of the call:

- **Rescue Handover :**In this case, the mobile station leaves the area covered by the original cell for another. The need of the handover is determined by the transmission quality that is measured by the error rate, the received signal strength RSSI measurement, the interference level and the propagation delay.

- **Confinement Handover :** the mobile station would experience a lesser interference if it changed cells, where the interference is due to the other active mobile stations in the same cell. The mobile station is continuously listening to the other antennas to assess the quality of a connection. In addition, each mobile station is synchronized with several base stations to be ready in case of handover.

- **Traffic Handover :** Many mobile stations are connected in the same cell, and the neighboring cells can accommodate new ones. This decision requires knowing the load charge of the other base station.

In GSM, the handover is the responsibility of the Mobile Services Switching Center MSC, in the other hand, in 3G it is decided by the RNC whereas in 4G LTE networks which negotiates directly between the eNodeB that manages the radio cells concerned.

1.6.1 Types of Handover

We can classify the Handover process in accordance with the following characteristics. We have the horizontal handover that occurs when the mobile station changes the connected cell in the same network; we have also the vertical handover which happens when the mobile station swaps from a network to another. Moreover, we have the hard and soft handovers. In the following paragraphs, we will give more details about this procedure.

1.6.2 Horizontal handover

It is also called intra-system handover, it takes place when a mobile station changes the base station supporting the same network technology e.g., two geographically distant eNb of an LTE network. This kind of handover occurs all the time, due to the mobility of the users and generally it doesn't create any problem neither for the mobile user nor the network. This type of handover arises due to the following reasons: worse/loss of signal received, cell loaded and mobility of users. The following Figure 1.11 represents the intra-system handover.

1.6.3 Vertical handover

It is referred also to the inter-system handover; it happens when a mobile station changes the type of connectivity. It is used to access a supporting infrastructure. Vertical handover refers to the automatic switching from one technology to another in order to get a better communication. This latter has two major reasons; achieving better QoS and reducing the cost. This kind of handover is a complicated one because it implicates different networks which have various architectures. Recently, researchers have treated this problem and many papers have been published trying to make it seamless. The following Figure 1.12 describes the vertical handover.

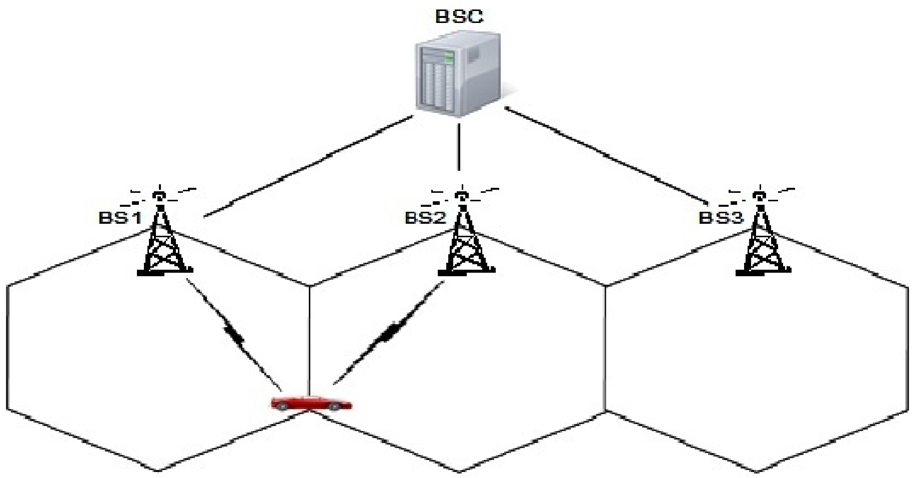

Figure 1.11: Horizontal handover

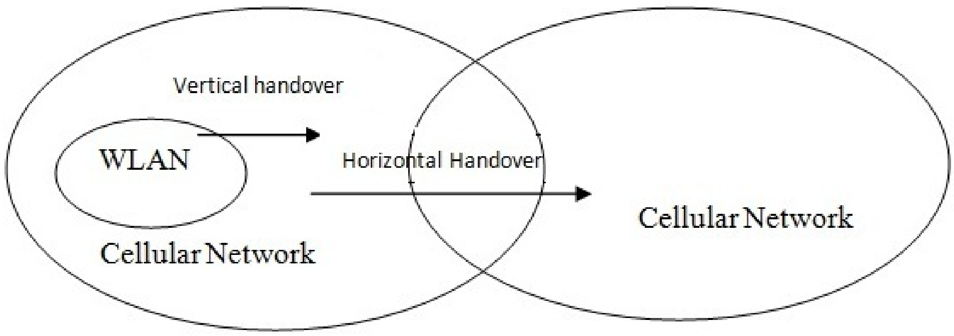

Figure 1.12: Vertical handover

1.6.4 Hard handover

Hard handover means that all the old radio links in the mobile station are removed before the new radio links are established. It can be seamless or not which means that the handover is not perceptible to the user. Practically, a handover that requires a change of the carrier frequency (inter-frequency handover) is always performed as hard handover.

1.6.5 Soft handover

Soft handover means that the radio links are added and removed in a way that the mobile station always keeps at least one radio link to the base station. It is also known as make before break because it maintains the previous channel from source until it gets the channel from source cell. Although soft handover increases the complexities, it has many advantages like the high hand over success rate, reduction of call drop probability and elimination of inference. Soft handover is performed by means of macro diversity, which refers to the condition that several antennas are used for transferring the same signal at the same time. Figure 1.13 illustrates the hard and the soft handover.

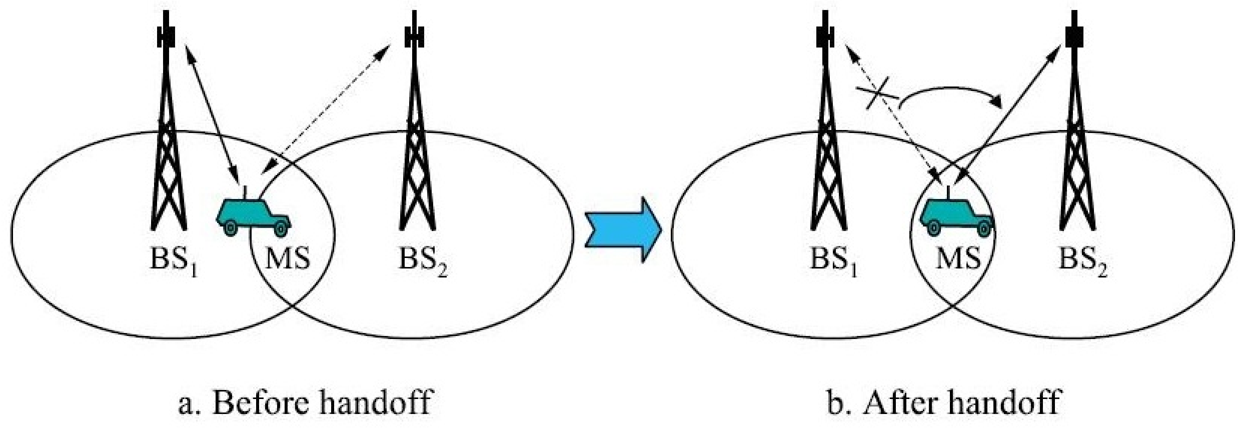

Figure 1.13: Vertical handover

1.6.6 Uncertainty in handover decision phase

The handover decision phase in heterogeneous networking is the most important step that directly affects a Mobile node [30]. It is a complex process which aims to answer three fundamental questions about vertical handovers:

- When? The answer to this question resulted in the area of handover prediction.
- Which? The answer to this question forms the area of network selection.
- How much? The answer to this question is sought through resource allocation and QoS management techniques.

Each of these fundamental questions forms a unique area of research attracting an immense amount of activity in both academia and industry.

The correct handover prediction in the vertical handover decision phase is very important. A wrong answer to the question "WHEN?" will disturb the knowledge of a network's availability can lead to an overall degradation in performance due to instability in other phases of the vertical handover, even resulting in connection loss. In fact the correctness and accuracy of decisions made in all the three handover stages depends on the accuracy of prediction of the availability of a network's coverage. The performance of all mechanisms linked directly or indirectly to vertical handovers can be greatly enhanced by obtaining an answer to the complex and important question "When" is the device expected to perform a vertical handover?" Therefore it is not an exaggeration to state that an accurate knowledge of the duration of availability of a network in relation to a mobile node motion within that network is crucial to the successful management of all aspects of functioning in 4G heterogeneous networks.

1.6.7 The need to avoid unnecessary vertical handovers

To have a successful vertical handovers in heterogeneous wireless networks, the unnecessary vertical handovers issue must be eliminated. The rule of a successful vertical handover is that the mobile node should stay connected to the new network for a duration equal to the handover recovery period. The handover recovery time is the period of time in which the data received on the new interface is equivalent to at least the amount that would have been received on the old interface in the duration equal to the total handover procedure. If the mobile node is forced to perform a vertical handover once again before the recovery duration period expires, the handover will be considered as unnecessary.

The main causes for unnecessary handovers are summarised in the following:

- The failure to recognise temporary coverage.
- Unavailability of required resources.
- Congestion in the new network.

Among these, the problem of predicting temporary coverage still remains largely unsolved. For instance, A mobile node roaming into a strong but temporary coverage of a WLAN may have access to the most optimal resources and the most favourable channel conditions. Despite all this, the fact that it will have to perform a vertical handover before successfully utilising these resources means that their availability is virtually useless unless it is harnessed in the correct manner. An unnecessary vertical handover actually results in an increased signalling overhead and delay.

1.7 Summary

In this chapter, we managed to present generalities concerning mobile wireless networks. We started by giving definitions, then we detailed the networks concerned in this thesis, notably, WLAN networks popularly known as Wi-Fi, we also discussed cellular networks focusing specifically on the 3GPP family with all the Generations from the analogue era passing through GSM to the LTE-A. Section 3 highlighted over the differences between the traffic flows while giving the QoS criteria required by this type of flow. The last section was devoted to the handover process, which is a crucial element in mobile wireless networks.

Chapter 2

A review on Network selection

Contents

The Heterogeneous Wireless Networks (HWNs) is a term referring to mixed networks made of different Radio Access Technologies (RATs) in the aim of achieving the Always Best Connected concept (ABC). In this case, users with multi mode terminals can connect to different wireless technologies such as 802.16, 802.11, UMTS, HSPA and LTE in the same time. The problem here is the choice of the suitable access network from the available RATs list; the decision process is called the Network Selection NS. The network selection process depends on several parameters such as QoS, mobility, cost, energy battery life . . . etc. Several methods and approaches have been proposed in this context, the objective was to satisfy the users with the best QoS and/or maximize the re-usability of the networks. This chapter represents a survey of the used methods in the field of network selection; in this constant we present the multiple attribute decision making methods (MADMs). Furthermore, we would illustrate the game theory concept and finally we talk about the use of the fuzzy logic theory in the network selection process.

2.1 Introduction

In the literature two famous of wireless networks families, the IEEE [1] and the 3GPP [2]. Two decades before, the IEEE family dominated the Internet market with the IEEE 802.11 standard; it was a big innovation for two reasons: (a) it was affordable in terms of cost, (b) it offered a high data throughput. This standard was and still a milestone. At the time, the 3GPP's cellular technology was the Global System of Telecommunication Mobile GSM, GSM is considered as a good technology for calling and sending Short Messages SMS, but it can't provide internet access. This shortcoming of GSM has pushed the 3GPP's developers to update GSM to the Global Packet Radio Service GPRS. Unfortunately, the internet access with higher data throughput was not possible either. At this moment the 3GPP's developers have realized that the problem resides in the internal process of GSM and GPRS. Hence, they change then the circuit-switched functionality to the packet-switched and called this release as Universal Mobile Telecommunications System UMTS or the third generation of networks 3G. In the circuit-switched technology, the user is billed depending to the antenna occupation time rather than the amount of data transmitted since the bandwidth is reserved for the user. However, the packet-switched technology allows each user transmissions to compete for the available bandwidth, and the user is billed for the amount of data transmitted, thus it utilizes the bandwidth much more efficiently. Moreover, the 3GPP has been upgraded with several enhancements called releases until the LTE-A.

Until the third generation networks, the network's radio access was mainly homogeneous; subsequently, the development of the networks technologies has led to an impressive growth of the Internet applications services as well as raising the mobile's users industry. Presently, people are equipped with Smart-phones and seek for the Always Best Connectivity ABC. It is obvious that no single RAT can offer the ABC. Therefore, it was necessary to change the idea, i.e., switching from the homogeneous systems to the heterogeneous ones. The aim of the fourth generation networks is to achieve the ABC concept, by offering mobile users the ability to take advantages from the networks having different architecture and performance.

So, nowadays, we have a variety of Radio Access Technologies RAT, the WLAN basically IEEE802.11, UMTS, HSPA and the LTE. This variety constitutes elements of the heterogeneous environment, (Figure 2.1). The heterogeneous system allows mobile users to choose

[1] http://standards.ieee.org/
[2] http://www.3gpp.org/

among multiple RATs based on several criteria, this choice is called the Network Selection (NS) which is the scope of this chapter. The network selection procedure consists of selecting the best network among available ones, Figure 2.1. However, the large number of parameters involved makes the decision very hard and complex, cost, QoS, energy consumed . . . etc. The basic definition of the network selection problem is the dynamic and the automatic choice of the best wireless access network. In the classical cellular systems, the network selection was directed by physical layer parameters, and the mobile terminal is often associated to the best-received base station, which is why such selection policy is obviously not suited for the heterogeneous wireless access technologies. For instance, a user may favour to connect to a less loaded RAT at larger distance with a lesser Received Signal Strength (RSS), rather than a more loaded cell with high RSS.

The network selection in a heterogeneous environment can be reported to the multiple attribute decision making MADM problem because of the number of parameters involved and the criteria. This mathematical approach has been widely used to solve the NS problem [4] [5] [1] [2]. Other methods like fuzzy logic and game theory have been proposed to solve the network selection problem in [31] [32]. In this chapter we try to make a survey of the methods and approaches used to solve the NS problem.

This chapter is organized as follows, an introduction that contains history and backgrounds, section two is devoted to the network selection procedure, while in section three we will focus on the approaches and methods that are proposed to solve the NS problem in the literature along with our critics. Concerning section four, we make a summary of the discussed methods and approaches with a recap table. Finally, we finish with a conclusion.

2.2 Network selection process

As we mentioned above, the NS process consists on switching between RATs to be always best served. So, when a multi-mode user discovers the existence of various RATs within it's area, it should be able to select the best network among them to operate-in and get the desired service. The different RATs provide different characteristics in terms of delay, jitter, throughput and packet loss rate. For this reason, in the context of 4G systems, selecting a RAT and being connected-to as long as possible is very important to minimize energy consumption, cost while maximizing the system efficiency, (Figure 2.2). The NS procedure

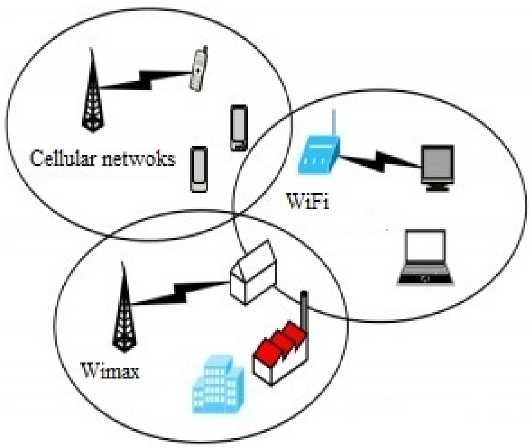

Figure 2.1: Heterogeneous wireless environment

is the general case of the handover process HO which can be either centralized, i.e., network-centric or decentralized, i.e., user-centric.

For the network-centric approach, the operator controls the whole process and makes decisions; the users obey these decisions and execute them. This can be seen as a good strategy to avoid problems like the selfish behaviour of users who tries to get the best RAT in the same time resulting in congestion situation. On the other hand, this approach assumes a situation of a single operator with multiple RATs and this can pose a problem in the case of multiple operators.

For the user-centric approach, users make decisions by themselves; this approach is decentralized and can easily generate a congestion situation because of the selfish nature of users. Nowadays, almost all operators offer 3G and 4G radio access in addition to the Wi-Fi connections, so, it is better to use the first approach.

Many parameters influence the decision process of the best RAT, the battery status level, energy required to get the requested services, the RSS received, the cost to pay, the bandwidth acquired, the user preferences: the required quality ... etc. These parameters can be categorized in various sets, Table 3.2.

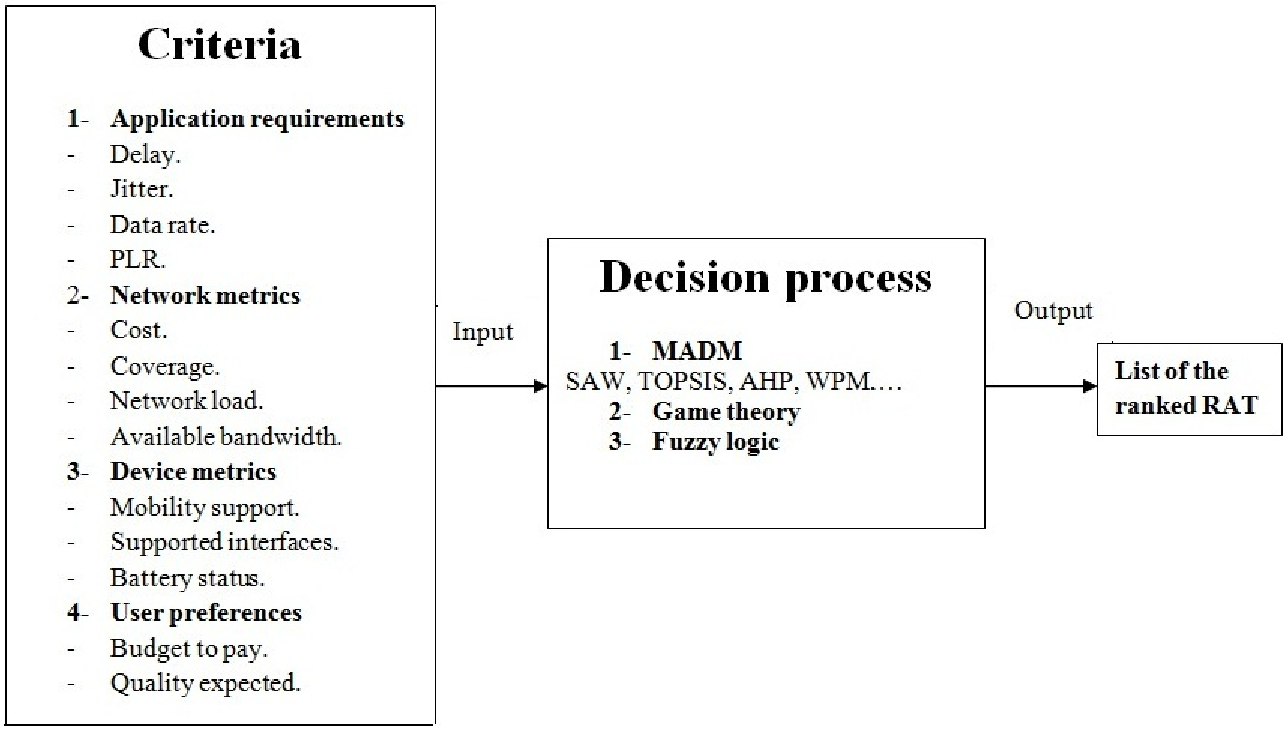

Figure 2.2: Network selection process

- Network conditions parameters: it represents information about the network conditions, among others such as network load, coverage area, network connection time, available bandwidth, etc.
- Application requirements parameters: it corresponds to the information about the threshold needed by the service application to be in the normal state, amongst others

we can mention: the required throughput, delay, jitter, the packet loss rate and the energy needed for the application.

- User preference parameters: it represents information related to the end users, it includes among others, the budget that user is ready to pay, the preference between cost and service quality.

- Mobile equipment parameters: it regroups information about the user's device, among others: the battery level status and mobility.

The parameters have also another classification; they can be dynamic or static, to be maximized or to be minimized. For example, the delay parameter related to the network conditions, the network load and the user's received signal strength. These parameters are a dynamic and must be minimized for the best of the user. Table 3.2.
Considering all the implicated parameters and their classifications, we can say that the NS problem is very complex task belonging to the NP-hard class. Several works in the literature have treated the NS problem. These studies focus on the optimization of the network selection decision for users in order to support many services with best QoS to maintain the users stay connected with the current access network as long as possible, this will be detailed in the next section.
The network selection NS process consists mainly on these actions :

- Monitoring step: it consists of discovering the available RAT, collecting the network radio conditions and other characteristics of the RAT. In this stage, some of the parameters are estimated and others are calculated.

- Decision step: the network selection decision is initiated. The choice of best network is based-on the monitoring process and other parameters provided by the mobile device such as the network and the user preferences. In this stage, the decision process is applied to rank the RATs.

- Execution step: it consists of connecting the user to the target RAT.

The network selection procedure is started in the following instances. When a new service is requested, a video/VoIP call or data transfer service, also when the received RSS becomes lesser than a given threshold and once the user's radio condition becomes worse in the mobility case for instance. Moreover, it can be triggered in the case of the weakening of user's battery status.

For the application requirements, called also the QoS parameters, it depends on the type of service desired. For the VoIP service the important parameters are delay and packet loss rate, for the video streaming the important parameters are bandwidth and delay, while for the best effort service, the important thing is the bandwidth acquired. The decision process is represented by a utility function. When using the MADMs methods, this means that we use the predefined function utilities of these methods. The objective functions are monotonic or non-monotonic; concerning the monotonic utility, it is a function that has an increasing or decreasing behaviour. It has a constant sense of change. In a given interval, the graphical representation of a monotonic function is a curve that rises or falls continuously. However, the graphical aspect is not the only form in which the property of monotony is revealed: a monotonic function is a function that has the same effect on the relation order. For an increasing function, the order between two variables x and y is reflected in the order of their images f(x) and f(y), for a decreasing function, the images order is reversed compared to the original order of the variables. Thus, one can generalize the concept of any monotonic function depending on an ordered set. If the function does not respect the order property, it is said non-monotonic function. Normally, the monotonic utility functions are used in the context of network selection problem.

The objective of this section was to give the readers a detailed description of the network selection procedure. Particularly, we describe the parameters that influence the decision process, and also we define the actions of the network selection algorithm, we also discussed the objective functions. In the next section, we are going to give a state of the art concerning the approaches and the used methods.

Set of parameter	Parameters	Static/ Dynamic	Maximized/ Minimized
Network conditions.	Network load. Network coverage. Network connection time. Available band-width.	Dynamic. Static. Dynamic Dynamic.	Minimized. Fixed. Minimized. Maximized.
Application require-ments	Throughput. Delay. Jitter. PLR. Energy consump-tion.	Dynamic Dynamic. Dynamic. Dynamic. Dynamic.	Maximized. Minimized. Minimized. Minimized. Minimized.
User preferences.	Budget. Cost.	Static. Static.	Fixed. Fixed.
Mobile equipment	Battery level. Mobility.	Dynamic. Dynamic.	Fixed. Fixed.

Table 2.1: Network selection's parameters classification

As we mentioned before, several parameters and attributes must be taken in consideration in the NS process. So, many authors have modelled the solution in order to find the most appropriate one leading to different solutions. In this section, we present a survey of the methods used to solve the NS problem.

2.3 Multiple Attribute Decision Making and network selection

Decision making is the study of identifying and choosing alternatives based on the values and preferences of the decision maker. Making a decision implies that there are alternative choices to be considered, and in such a case we want not only to identify as many of these alternatives as possible but to choose the one that best fits with our goals, objectives, desires, values, and so on. A standard feature of multi-attribute decision making methodology is the decision table. In the decision table each row belongs to a criterion and each column describes the performance of an alternative. The score a_{ij} describes the performance of alternative A_j against criterion C_i. In the decision table, weights w_1,...,w_m are assigned to the criteria. Weight w_i reflects the relative importance of criteria C_i to the decision, and is assumed to be positive. The weights of the criteria are usually determined on subjective basis. They represent the opinion of a single decision maker or synthesize the opinions of a group of experts using a group decision technique, as well.

Multiple Attribute Decision Making MADM is a famous mathematical approach in the context of preferential decisions; it treats multi criteria decision problems. This branch of decision making is widely used in various fields such as the economy sector [33] [34]. Simple people use this approach in their daily life , for example, buy a car or a house having different characteristics. This approach is very adapted to the network selection because of the multi criteria nature of the NS problem [35]. In fact, the basics of this approach MADM are as follows:

- Alternatives: the set of the actors who will be ranked. In the NS scenario, the alternatives are the lists of RATs.
- Set of attributes: it represents the parameters or the criteria used in the decision making process. For the NS scenario, parameters are throughput, jitter and delay . . . .
- Weights: this translate the importance of the parameter or the criteria in the decision process.

Finally, we get a decision matrix representing the system, where the columns are the criteria and the lines are the alternatives.

Several methods have been proposed in this context, such as: Simple Additive Weight (SAW), Technique for Order Preference by Similarity to Ideal Solution TOPSIS, Weighted Product Model WPM, Analytical Hierarchy Process AHP . . . etc [34]. SAW, TOPSIS and WPM are qualified as ranking methods that need other methods to weight criteria. Alternatively, the AHP method is a ranking/weighting method, it contains a process that generates the weights for the criteria. It is very important to note that these methods are applicable only when all the data of the input matrix are expressed in the same unit; this means that it is necessary to normalize data, Table 3.3. If this is not the case, then the final result is equivalent to "compare olives and banana". So, data must be normalized and it is an important step in the network selection procedure, Table 3.3 represents a non exhaustive list of normalization methods.

Another important point is the weight of the parameter for the user. In fact, users do not consider that parameters are equally important. Users can be QoS based ones, others are cost based and/or energy based users. Thus, weights are related to the user's profile. The weights can be subjective or objective. The subjective weights are empirical values based on experience, for example, in the case of QoS based users, when a user initiates a VoIP session, the most important parameters are delay and packet loss ratio which have 60 to 70 % of importance; conversely, the bandwidth it is not as important. In the case of the video session, the bandwidth is more important than other parameters, 50 % of importance is given to the bandwidth. The objective weights are given using mathematical methods, table 3.5. In the same time, the AHP method contains an auto-creative system to generate the weight vector using the eigenvectors and eigenvalues of the input matrix.

Method	Description
Max-Min	$e_{ij} = \frac{x_{ij} - min(x_{ij})}{max(x_{ij}) - min(x_{ij})}$
Sum	$e_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$
Square root	$e_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}}}$

Table 2.2: Normalization methods

Method	Description
Entropy	$w_j = 1 - \frac{1}{N} \times \sum_{i=1}^{n}[x_{ij} \times ln(x_{ij})]$
Variance	$w_j = \frac{1}{N} \times \sum_{i=1}^{n} x_{ij}$
Eigenvector	$(B - \lambda \times I) \times w = 0$

Table 2.3: Weighting methods

2.3.1 Simple Additive Weight SAW

SAW also known as weighted linear combination or scoring methods is a simple and one of most used multi attribute decision technique, The method is based on the weighted average score. An evaluation score is calculated for each alternative by multiplying the scaled value given to the alternative of each attribute with the weight of relative importance directly assigned by decision maker followed by summing of the products for all criteria. This method assumes that the treated data has the same unit, so to get a comparable scale among parameters, it is mandatory to normalize the data for each parameter [34] [36] [37]. Finally the candidate having the highest or the lowest value is selected.

The mathematical formulation of SAW is:

$$\mathrm{R}_{SAW} = \sum_{i=1}^{n} (\mathrm{w}_j \times \mathrm{r_{ij}}) \tag{1}$$

R_{SAW}: is the value of each alternative. w_j is the weight value of the criteria j and r_{ij} is the normalized value of criteria j and alternative i. Then we choose the maximum/minimum value, depending on the utility function tendency (to be maximized or to be minimized).

Since 1999, SAW method has been widely used in the context of network selection, [34] [37] [38]. In [37] and [38], the authors have used the Simple Additive Method SAW to get a ranked list of networks, while in [34] the authors make a mix between game theory and SAW method. The main benefits of SAW is its simplicity and the low complexity. However, it has two major drawbacks, first: a parameter can be outweighed by another one, second: the rank reversal phenomenon that represents a problem of the entire MADM approach.

2.3.2 Technique for Order Preference by Similarity to Ideal Solution TOPSIS

The Technique for Order of Preference by Similarity to Ideal Solution TOPSIS is a multi-criteria decision analysis method, which was originally developed by Hwang and Yoon in 1981 with further developments by Yoon in 1987, and Hwang, Lai and Liu in 1993. TOPSIS is an aggregating compensatory method based on the concept in which the chosen solution should have the shortest geometric distance from the positive ideal solution [39] and the longest geometric distance from the negative ideal solution. The normalized data for each parameter is weighted and therefore the geometric distance between each alternative and the ideal alternative is computed.

An assumption of TOPSIS is that the criteria are monotonically increasing or decreasing. Normalisation is usually required as the parameters or criteria are often of incongruous dimensions in multi-criteria problems. Compensatory methods such as TOPSIS allow trade-off between criteria, where a poor result in one criterion can be negated by a good result in another criterion. This provides a more realistic form of modelling than non-compensatory methods, which include or exclude alternative solutions based on hard cut-off.
The TOPSIS process is carried out as follows:

- Create an evaluation matrix consisting of m alternatives and n criteria, with the intersection of each alternative and criteria given as, x_{ij} therefore we have a matrix $(x_{ij})_{m*n}$.

- The matrix $(x_{ij})_{m*n}$ is normalized to get $(r_{ij})_{m*n}$ using one of the above mentioned methods Table 3.3.

- Calculate the weighted normalized decision matrix, where
  $t_{ij} = w_j \times (r_{ij})_{m*n}$
  $w_j = \frac{W_j}{\sum_{i=1}^{n}(W_j)}$
  $\sum_{i=1}^{n}(W_j) = 1$ and j=1...m; i=1...n

- Determine the best and worst alternative A_b and A_w respectively.
  $A_b = \{b_j^P = max t_{ij} / j \in J_+\}$
  $A_{wo} = \{wo_j^P = min t_{ij} / j \in J_-\}$
  where J_+ and J_- contain the criteria with positive and negative impact respectively.

- Calculate the separation measure for each alternative
  $D_P := \sqrt{\sum_{i=1}^{n} \left(w_j^2 \times (r_{ij} - b_j^P)^2\right)}$ and
  $D_N := \sqrt{\sum_{i=1}^{n} \left(w_j^2 \times (r_{ij} - a_j^N)^2\right)}$

- Calculate the relative closeness to the ideal solution

$$\mathrm{R}_{TOPSIS} = \frac{\mathrm{D}_P}{\mathrm{D}_P + \mathrm{D}_N} \tag{2}$$

TOPSIS has been applied in the network selection problem in [37] [35] [40]. In [37], authors compare the performance of vertical handover HO using SAW and TOPSIS. The authors concluded that TOPSIS outperforms the SAW method. The authors in [40] propose a network selection algorithm based on TOPSIS method. The networks are ranked based on the closeness to the ideal solution using TOPSIS method. Numerical examples have been used to evaluate the proposed solution. The parameters considered in the decision matrix are: available bandwidth, QoS level, security level, and cost. The results show that TOPSIS is sensitive to user preference and the parameter values.

In general, we can say that compensatory methods namely TOPSIS managed to avoid the problem that a parameter can be overshadowed by another one by allowing a trade-off between criteria. This means that a poor value in one criterion can be neglected by a good value in another criterion. This provides a better performance and gets more sense than non-compensatory methods, which use threshold system.

2.3.3 Weighted Product Model WPM

WPM called also Multiplicative Exponential Weighting MEW is similar to SAW [41]. The difference is the replacement of the addition operation in the SAW method with the multiplication operation in WPM method; each decision alternative is compared with the other ones by multiplying a number of ratios, one for each decision criterion. Each ratio is raised to the power equivalent to the relative weight of the corresponding criterion. The mathematical description of this method is as follows:

$$\mathrm{R}_{WPM} = \prod_{i=1}^{n} (r_{ij})_j^w \tag{3}$$

WPM has two variants, the difference resides in r_{ij}:
$r_{ij} = x_{ij}$ or $r_{ij} = \frac{x_{ik}}{x_{jk}}$

For the sake of simplicity, we use the first representation in the following when we mention the WPM method.
Authors in [42] have made a comparison between the SAW and WPM methods in the context of vertical handover, they use the relative standard deviation as a metric of comparison and they obtain a conclusion that WPM is better than SAW method. In [43], authors use the WPM method in the context of heterogeneous systems. The aforementioned work conclusion was that the WPM method is a more robust approach for the dynamic decision making and it penalizes the attributes with poor quality.

2.3.4 Analytical Hierarchy Process AHP and Grey Relational Analysis GRA

AHP considers decomposition of one complicated problem to a multiple hierarchy simple sub-problems. The AHP steps are as follows:

- Decompose the problem to a hierarchy sub problems, where the top node is the final goal and for each criterion we list the alternatives.
- Pair-wise comparison of attributes and translating them into numerical values from 1 to 9.
- Calculates the weights of each level of hierarchy.
- Synthesizing weights and getting overall weights.

Regarding GRA method, it is used to rank the candidate networks and it regroups the following steps:

- Normalization of data is performed considering three situations: the higher is better, the lower is worse, and nominal is the desired.
- Definition of the ideal sequence in the three considered situations: the ideal sequence contains the higher bound, lower bound and moderate bound.
- Computing the grey relational coefficient (GRC): the more favourable sequence is when the GRC is larger.

The AHP method usually is coupled with Grey relational Analysis GRA method. The AHP for weighting and the GRA is used to rank alternatives. Authors in [44] have used a modified AHP and compared it with the normal AHP using a Quality of Experience (QoE) criterion; finally their Numerical results show that the proposed scheme outperforms the conventional AHP scheme resulting in a good load balance. In [45], authors use AHP to rank various criteria to compare the desirability of different Internet advertising networks; the proposed model provides an objective and effective decision model to be used by advertisers in selecting an Internet advertising network.

As a summary, MADM are widely used in order to solve the network selection problem, this is due to the fact that the network selection has the same scheme and characteristics of the problems solved by the MADM. Moreover, these methods are known by their facility of use, clarity and low complexity of computation. The disadvantages of these methods can be reported to the following reasons:
First, these methods do not have the same performance toward the different services (VoIP, Video Calls and web browsing). For instance, a method can produce a good performance with VoIP service and a bad performance for the best effort services which is not ideal Table 3.6. Secondly, these methods suffer from the problem of ranking abnormality, i.e. it is a phenomenon that occurs in the MDAM methods when an exact replica or a copy of an alternative had introduced or eliminated. Authors in [46] have shown that the rank reversal problem occurs in the most of the well known MADM methods. This problem has been addressed in other works [47][48] by modifying methods, but the original versions of MADM methods suffer from the rank reversal phenomena. Additionally, AHP method is very complicated and has a high complexity computation, when calculating the vector of weights. For all these reasons, we can say that MADM are a good solution, but the lack of a general method that serves all kinds of services is a problem.

Methods	Advantages	Inconvenient
MADM	Easy to understand Easy to implement Good results in some cases	Rank reversal phenomena High complexity such as AHP and ELECREE Good performance with some services and bad results for others

Table 2.4: MADM method advantages and inconvenient

2.4 Game theory

Game theory is a discipline aimed to model competitive situations, which implies the interaction of rational decision makers with mutual or conflicting interests [37] [35] [38]. It provides analytical tools to predict the outcome of complicated interactions between rational interacting entities. Its first application was in economics field where it has been immensely successful. Consequently, game theory has been widely applied to other fields. In telecommunication, it was largely used in routing networks, energy consumption and resource allocation in competitive systems. In wireless networks, mobile devices are rational, independent entities that compete to increase their gain. For this reason, game theory is an adapted tool to the model conflicting situations of wireless networks. In this chapter, we focus on the use of game theory for modelling the network selection problem [3].

In 1838, Antoine Cournot has used a formal game-theoretic analysis to study the duopoly problem. It was also used by the mathematician Emile Borel in 1921 in his researches. The Game theory was considered as a field in its own right in 1944 due to the monumental book named "Theory of Games and Economic Behavior" published by Von Neumann and Oskar Morgenstern. This book provided the basic terminology and problem set-up that is still in use today. Game theory received special and more attention with the awarding of the Nobel prize in economics to Nash, John Harsanyi, and Reinhard Selten. So, In 1950, John Nash demonstrated that finite games have always an equilibrium point, at which all players choose actions which are best for all of them given their rival choices, this concept is called then Nash equilibrium. This central concept of non-cooperative game theory has been a focal point of analysis since then. In the 1950s and 1960s, game theory was expanded theoretically and applied to problems of war and politics. Since the 1970s, it has driven a revolution in economic theory. Additionally, it has found applications in sociology and psychology, and established links with evolution and biology.

At the end of the 1990s, a high-profile application of game theory has been the design of auctions. Prominent game theorists have been involved in the design of auctions for allocating rights to the use of bands of the electromagnetic spectrum to the mobile telecommunications industry. Most of these auctions were designed with the goal of allocating these resources more efficiently than traditional allocation schemes.

So, in the telecommunication, it was largely used in routing networks, energy consumption, and resource allocation in competitive systems. In wireless networks, mobile devices are rational, independent entities that's compete to get most possibly interest, for this reason game theory is an adapted tool to model conflicting situations of wireless networks. An example of conflicting scenario in which a game theory solves the problem is the Samsung profits does not depend only on its choice in terms of price but it depends also on the prices of challengers such Apple. As a rule, the fundamental characteristic of a game is that the gain of a player depends on his choices and also on the choices made by other players of the game which is called a strategic game.

A game is represented mainly by three sets: a) set of players that contains the rational actors competing to get more payoff, b) a set of actions or strategies which depends on the available information in the system. Obviously, each player seeks the action which maximizes its revenue. The payoff is a utility function representing the player revenue when choosing a given strategy, c) the pay-off for each player can be represented as the actual or expected utility which a player receives by playing the current strategy. The game is played whilst the player get more gains. When the payoff cannot be further enhanced with any other strategy combination, then, the equilibrium known as Nash Equilibrium is reached. Basically, a Nash Equilibrium occurs when each player cannot get more gains by changing only his strategy while keeping the other players strategies' unchanged, it is the combination of the best strategies for all players. A detailed representation of a strategic game is written as:

- $Game = \{P, A, G\}$.
- $P = \{1, n\}$ the set of players..
- $A = \{a_1, a_n\}$ the set of actions, denoting the set of available strategies of a player i, $1 \leq i \leq n$.
- $G : \{payoff\}$, it represents a reward achieved when choosing a strategy. Here, to simplify things, the payoff function is the linear summation of local gains using the weights of each parameter.

A game can be either in strategic or in the extensive form.
The strategic form called also the normal form is the basic type of games studied in the non-cooperative game theory. A game in strategic form lists each players strategies, and the outcomes that result from each possible combination of choices. An outcome is represented by a separate pay-off for each player, which is a number (also called utility) that measures how much the player likes the outcome, Figure 3.1 .

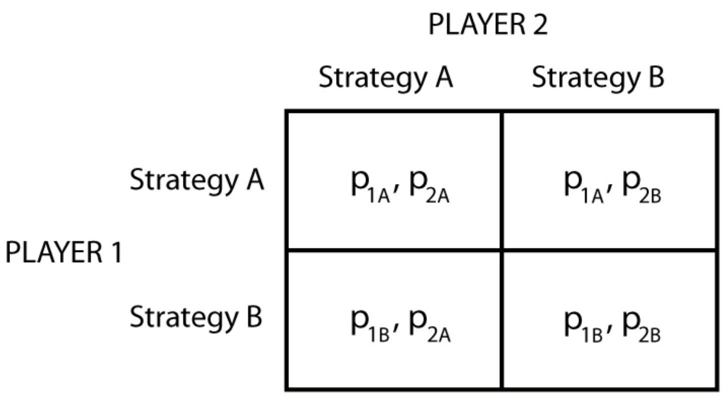

Figure 2.3: Strategic form of a game

The extensive form, also called a game tree, is more detailed than the strategic form of a game. It is a complete description of how the game is played over time. This includes the order in which players take actions, the information that players have at the time they must take those actions, and the times at which any uncertainty in the situation is resolved. A game in extensive form may be analysed directly, or can be converted into an equivalent strategic form, Figure 3.4.

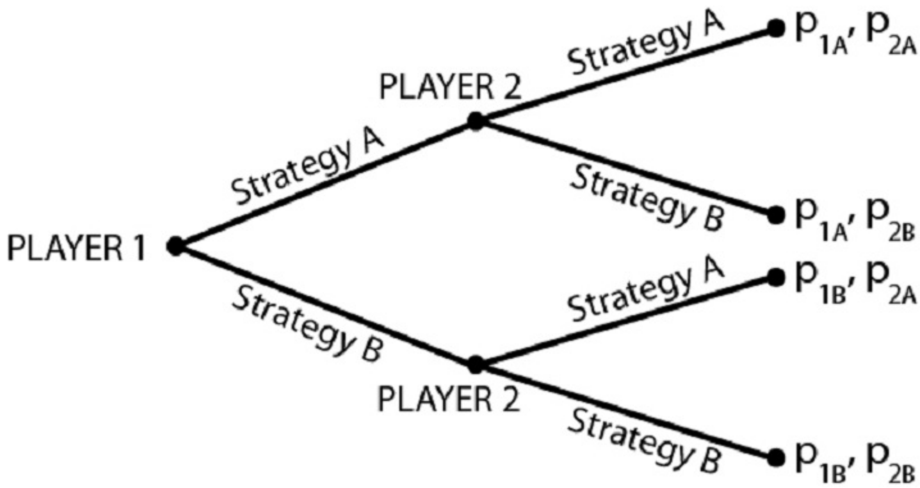

Figure 2.4: Extensive form of a game

As we mentioned above, a game is composed of three sets, players, strategies and pay-off. each player seeks to maximize his profit by choosing strategies depending on the information available in system at a certain moment. When each player cannot rising his pay-off function any more by changing his strategy unilaterally by keeping the other players strategies unchanged, then we say that the solution of the game represents a Nash Equilibrium. In contrast, if there is no other outcome that makes every player at least as well off and at least one player strictly better off, the pay-off is said Pareto optimal. So, a Pareto Optimal outcome cannot be improved upon without hurting at least one player.

In game theory, peoples might confuse between these concepts, but in the reality, both of them answer two totally different questions regarding a particular game. A Nash equilibrium looks for the feasible strategies in the non-cooperative games while the Pareto optimal concept is about the efficiency of the solutions. A Nash equilibrium may not be Pareto optimal.

Taking the example of Prisoners Dilemma. If both prisoners stay quiet they each get 1 year, if both confess against the other they both get 2 years. In case only one confesses against the other, he gets free while the other guy gets 3 years. In this case the Nash Equilibrium is both confess while the Pareto optimal equilibrium will be the one where both don't confess as in this case both get a 1 year sentence while in the Nash equilibrium they get a 2 year sentence each.

Different types of games are used to model various cooperative or competitive situations between rational decision makers. Some of the most widely used game theoretic models are outlined below.

2.4.1 Cooperative and non-cooperative games

Cooperative game theory called also coalition games is a description of games in which we specify only the pay-off that can obtain by the cooperation of its members, or coalition.

In the same time, the process by which the coalition forms is not explicit. John Nash has proposed a solution for the division of gains from agreement in a bargaining problem which depends solely on the relative strengths of the two bargaining parties. The model of Nash fits within the cooperative framework in that it does not specify a time-line of offers and counteroffers, but rather focuses solely on the outcome of the bargaining process.

On the other hand, the non cooperative games called also a competitive games; specify all the strategic options available for the players, while the contracts underlying the coalitions in a cooperative game are not described. Each player tries to get his goal without any regards to the other players; here, the players are called rational [49]. Generally the non-cooperative games admit a solution called the Nash Equilibrium, while for the cooperative games the solution is the Shapely value that means that for each cooperative game, it assigns a unique distribution among the players of a total surplus generated by the coalition of all players. It proposes a fair distribution of the gains for the coalitionist players.

2.4.2 Games with complete/ perfect or incomplete/imperfect information

Complete information is a term used in game theory to describe games in which knowledge about other players is available to all participants. Every player knows the pay-off and strategies available to the other players. Complete and perfect information are significantly different. In a game of complete information, the structure of the game and the pay-off functions of the players are commonly known, but players do not see all of the moves made by other players. For the perfect information games, each player is perfectly informed of all the events that have previously occurred, but may lack some information about pay-off of other players, or on the structure of the game.

Inversely, in games with incomplete information, some players do not know the information of other participants, like other player's payoffs. In a game with imperfect information, players are simply unaware of the actions chosen by other players. However, each player know who the other players are, their possible strategies and their preferences. Hence, data about the other players is imperfect, but complete [3][49].

2.4.3 Repeated games

In the strategic form of games, the players make their decisions simultaneously at the beginning of the game. On the contrary, the extensive game model defines the possible orders of the events. In this case, the players can make decisions during the game and they can react to other player's decisions. Some of the extensive games which be finite or infinite. A type of extensive games are repeated games, in which a game is played numerous times and the players can observe the outcome of the previous game before attending the next repetition [50][49].

The stage game is an extensive form of games which consist in some number of repetitions of some base game. one of the well-studied 2-person games is the stage game. It captures the idea that a player will have to take into account the impact of his current action on the future actions of other players; this is sometimes called his reputation. The presence of different equilibrium properties present the threat of retaliation, since one will play the game again with the same person. It can be proved that every strategy that has a payoff greater than the min/max pay-off can be a Nash Equilibrium, which is a very large set of strategies. Single stage game or single shot game are names for non-repeated games.

2.4.4 Zero-sum games

A zero-sum game is a situation in which one person's gain is equivalent to another's loss, so the net change in wealth or benefit is zero. A zero-sum game may have two players, or millions of participants and they are usually called the strictly competitive games in the game theory discipline, but are less widespread than non-zero sum games. In contrast, non-zero-sum game describes a situation in which the interacting parties aggregate gains and losses can be less than or more than zero. A non-zero-sum games can be either competitive or non-competitive. Zero-sum games are most often solved with the minimax theorem which is closely related to linear programming duality and with Nash equilibrium.

Poker and gambling are popular examples of zero-sum games since the sum of the amount won by some players equals the combined losses of the others, [50].

2.4.5 Game theory on network selection

In the wireless environment, game theory has been used in order to solve many distributed power control [51], resource management and allocation, and dynamic pricing [52] related problems. A more comprehensive survey on general game theory application in wireless networks is offered by Charilas et al. in [50]. The book [53] presents a collection of fundamental issues and solutions in applying game theory in different wireless communications and networking domains (e.g., wireless sensor networks, vehicular networks, power control games, economic approaches, and radio resource management).

Game theory is adapted to solve the network selection problem, a mapping of network selection components on game theory s provided in the Table 3.8. The players of the game can be the mobile users, the networks or a mixture between them. The players seek to maximize their pay-off by choosing different strategies, such as: available bandwidth, available APs ...etc. The pay-off are estimated using utility functions based on various decision criteria: monetary cost, energy conservation, network load, availability, etc.

Game theory components	Network selection correspondent components
Players	The agents who play the game, they can be users or networks.
Strategies	The action taken by the players.
Pay-off	The profit estimated by using a utility function.

Table 2.5: MADM method advantages and inconvenient

In this subsection we will discuss the works that use game theory in the network selection problem. As we mentioned above, a game is defined by three sets, players, actions and pay-off. Players can be: users, networks, or both. In the following we will discuss all these cases separately.

Game between users

We started by making the difference between the cooperative and the non cooperative games, so, in each case the game can be either cooperative or not, in the case of the non cooperative games, the users compete against each other while seeking to maximize their own utility.

In [31], authors modelled the competition between users for one access point as an evolutionary game. The users represent players that compete to maximize their transmission rate where this latter represents the strategy, hence the pay-off is modelled by a utility function. This function takes the delay and packet loss rate and determines the mean opinion score MOS, which is a subjective measure for voice quality. In this work, the notion of free users is used and the VoIP service is the only service assessed and the reached equilibrium is optimal.

In [54], authors emphasize the problem of the least congested access point selection; here players are users, the set of strategies are the different access points available, the pay-off is a trade of between bandwidth gained and the effort generated when moving to the new access point.The authors demonstrate that the result of this game is the distribution of users on access points.

In [8], authors consider the scenario of a single network Wi-Fi and multiple access points. In this case study, users represent players that can choose one access point to connect to, the pay-off function depends on the congestion level of the access point and the monetary cost so, users achieve a Nash equilibrium.

In order to model the network selection problem, a Bayesian game is proposed in [7]. The players are the users, and their action set is represented by the selection of an available access network. Each user has partial information about the preferences of other users. The authors show that a Bayesian Nash equilibrium can be reached in an environment with incomplete information.

The authors in [55] have proposed an auctioning game to model the resource allocation problem between the wireless users. They consider the following scenario, multiple wireless users are located in the coverage area of a number of base stations BSs, each user is interested in buying a certain amount of bandwidth from the BS. Each user has a maximum budget allowed to spend, and from which he bids for a BS allocation. Based on the users bids, every BS will allocate its available bandwidth among the wireless users in a proportionally fair manner. The authors claim the existence of Nash Equilibrium for the case where each user can access all BSs. However, in the case where users that can access only a subset of all BSs the existence of Nash Equilibrium is not guaranteed.

A few work have addressed the NS problem using a cooperative game between users. In [56], the authors consider the scenario of sharing bandwidth of a single cell of one BS, they opt to use two approaches, a cooperative bargaining game and a bankruptcy game. In the first approach, they claim that the Nash Equilibria is attended. For the second approach, the results show that the maximization of total capacity is reached by using Constrained Equal Awards CEA, but in terms of maximum fairness the Random Arrival RA rule act better.

The remarks that can be deducted from the above-mentioned works showcase that a game between users means that players in general represent the non cooperative behaviour and because of the selfish nature of users, these games lead to system congestion and also to the monopolization of resource by certain users. Additionally, if each user makes his calculation, this will lead to substantial battery consumption of the mobile users. Hence, this aspect must be considered when modelling a game between users. In the case of game between users, we propose to enforce the cooperative behaviour instead of using the competition to limit the selfish side of the users.

Game between networks

In the case when players represent the networks, the latter compete to maximize the user's number and get more revenue. In [6] authors present a non cooperative game mechanism between networks, the players compete for service requests and try to admit the maximum number of customers to gain an amount of point representing the user's satisfaction. Still this scheme has a problem as preferences toward players are the same.

In [57] authors used a game theory model where they introduce the strategy space and quality point concept, players are the networks and payoff function determines which access network will provide the service requested by the user; this corresponds to a distribution of service requests amongst the networks.

In [58], authors investigate the admission control problem by modelling a multi-round game between two WI-Fi networks. Here the players in the game are the two networks and the strategy set is the user service requests. The pay-off of the game is the distribution of the service requests between the networks.

The network selection problem is modelled as a non-cooperative game between the networks in [6] in which, the networks compete against each other in order to maximize their own profit. The pay-off are defined based on a utility function which models the user preferences. The utility function follows a three zone-based structure, which was previously proposed in [59], that defines the users level of satisfaction relative to delay: satisfied, tolerant, and frustrated. The authors argue the existence of Nash Equilibrium and observe that under Nash equilibrium the networks pay-off are maximized when the users with higher preferences for the specific network are selected.

A cooperative game was proposed in [60] which studies the interaction between networks. The authors consider a scenario where a wireless multi-mode terminal can be served simultaneously by three different access networks owned by different cooperating network operators. In this context, the bandwidth allocation and admission control problems are modelled using a bankruptcy game. In this game the mobile user who initiates a connection request is seen as the bankrupt company, the bandwidth requirement is the money that has to be distributed among different networks. The access networks involved cooperate in order to provide the required bandwidth to the mobile user by using a coalition form and a characteristic function which is used to express the pay-off of the coalition. The solution of the bandwidth allocation problem is computed by using the Shapley Value and the core concept is used in order to analyse the stability of the allocation.

From all these works, formulating the games between networks has a well known strategy which is seeking for the maximum revenue, i.e. maximizing the number of users connected to this network. This approach guides users to think about their corresponding schemes for network selection under a competitive environment.

Game between users and networks

In this case players can be mobile users and/or networks. Users compete against networks, each seeking to maximize their own utility. On one side, the users try to maximize their benefits from the service for the price they pay. On the other side, the networks try to maximize the profit for the provided services.

In [61], authors propose a reputation-based network selection mechanism by modelling the interaction between user and network using repeated Prisoner's Dilemma game. To reinforce the cooperation between users and networks they combine the reputation-based systems and game theory. The network reputation factor represents network past behaviour in the network selection decision. They show that using reputation is essential in the case of repeated interaction to maintain the cooperation.

In [62], the author's addressed the network selection problem as a non-cooperative auction game in which the buyers represent the users, sellers/bidders are the available network operators and the auction thing is the requested bandwidth with associated attributes. The auction that maximizes the user's utility is the winning bid.

In [63], authors present a non-cooperative resource allocation based on the Cournot game between a provider and his customers where users are classified into three classes: premium, gold and silver. The strategies for the provider and the customer are: the provider seeks for customers who bring high revenue, while for the customers; if they are unsatisfied from the received QoS, they can decide to leave the network. Users are accepted into the network if the new utility computed when a new customer arrives is less than the providers' utility value. Finally, the authors find the equilibrium for resource distribution.

A network selection situation modelled as a user-network cooperative repeated game is presented in [64], The approach separates propose four strategies for users and two for networks.
Each user has four strategies: Grim strategy stipulating that the user participates in the relationship but if dissatisfied he will leave the relationship forever. Cheat-and-Leave strategy gives the user the option to cheat and then leave the network after cheating. Leave-and-Return strategy dictates that in case the network cheats the user leaves for only one period and returns in the subsequent interaction, and Adaptive Return strategy in which the user returning is dictated by the normalized weight of network's past degradation behaviour.
The network can choose between two strategies: Tit-for-Tat strategy which mimics the action of the user, and Cheat-and-Return strategy which gives the option to the network to cheat and return accepting the user's punishment.
The authors show that employing the proposed Adaptive Return strategy can motivate cooperation, resulting in higher pay-off for both players.

The analysis of this kind of game can be summarized in which users compete against networks while each seeking to maximize their own utility. On one side, the users try to maximize their benefits; it can be cost-benefit and/or QoS benefits. On the other side, the networks try to maximize the profit for the provided services.

2.5 Fuzzy logic

The idea of fuzzy logic was first advanced by Dr. Lotfi Zadeh [65] of the University of California at Berkeley in the 1960s. Dr. Zadeh was working on the problem of computer understanding of natural language. Natural language like most other activities in life, indeed, the universe is not easily translated into the absolute terms of 0 and 1. Whether everything is ultimately describable in binary terms is a philosophical question worth pursuing, but in practice much data we might want to feed a computer is in some state in between and so, frequently, are the results of computing. It may help to see fuzzy logic as the way reasoning really works and binary or Boolean logic is simply a special case of it.

In fuzzy logic, there are degrees of satisfaction of a condition [65]. Instead of Boolean algebra that considers a proposal is true or false often called "crisp" values, fuzzy logic adds a degree of truth to choose for example in [0,1]. Formalized by Lotfi Zadeh [3] in 1965 , it is a tool of artificial intelligence used in various fields, [66]. It is based on the mathematical theory of fuzzy sets Lotfi Zadeh who has an extension of the theory of classical. Fuzzy logic can handle the concept of partial truth, where the truth value may balance from completely true to completely false.

Fuzzy logic seems closer to the way our brains work. We aggregate data and form a number of partial truths which we aggregate further into higher truths which in turn, when certain thresholds are exceeded, cause certain further results such as motor reaction. A similar kind of process is used in neural networks, expert systems and other artificial intelligence applications. Fuzzy logic is essential to the development of human-like capabilities for AI, sometimes referred to as artificial general intelligence: the representation of generalized human cognitive abilities in software so that, faced with an unfamiliar task, the artificial intelligence system could find a solution.

2.5.1 Fuzzy logic concepts

Suppose that A is part of a set E, usually associated with its characteristic function, this function is applied to the elements x of the set E. It takes the value 0 if x does not belong to A and 1 if x belongs to A. We define a fuzzy part A that's assigned to each element x of E a membership degree. This value will be equal 0 if x is not element of A, it will be worth 1 if x belongs to A, and it takes a value between 0 and 1 if we are not sure of the membership of x in A. It is therefore necessary to define a fuzzy part as follows: A fuzzy part of a set E is a mapping of E into [0,1]. More generally, if L is a complete lattice, distributive and complemented it defines part L-fuzzy as an application from E to L. If L = [0,1], we find the above definition of blurred party, and if L = 0,1, we find the usual concept of E.

A fuzzy logic system FLS can be defined as the non-linear mapping of an input data set to a scalar output data [67]. A FLS consists of four main parts: fuzzifier, rules, inference engine, and defuzzifier. These components and the general architecture of a FLS is shown in Figure3.4 The process of fuzzy logic consists on firstly, the conversion of a crisp set of input data to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as fuzzification. Afterwards, an inference is made based on a set of rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step.

2.5.2 Fuzzy logic in network selection

Few studies have addressed the network selection problem using fuzzy logic as a core of the ranking scheme; basically, the authors use fuzzy logic in network selection in two ways: first

[3]http://wi-consortium.org/wicweb/pdf/Zadeh.pdf

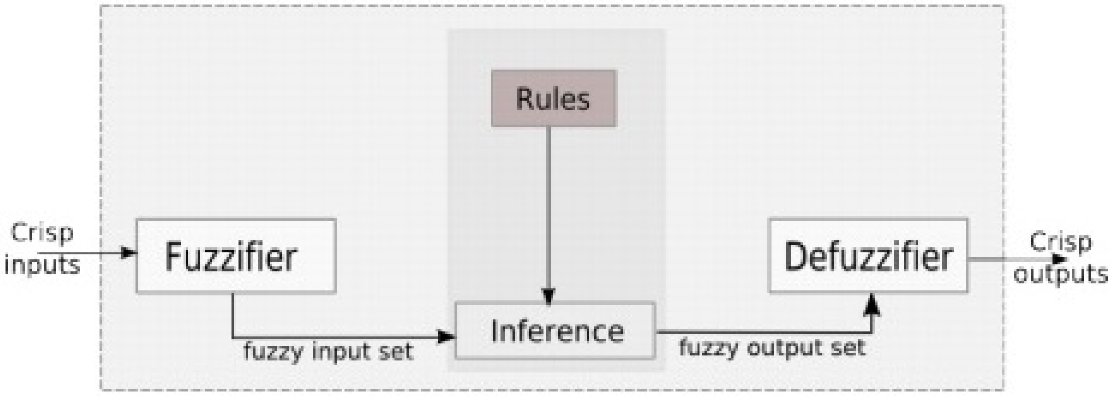

Figure 2.5: Fuzzy Logic System FLS

a combination of MADM method with fuzzy logic; secondly, fuzzy logic is used as a scheme of selection.

In [32], authors propose a general scheme to solve the multi-criteria network selection problem. In the proposed scheme model, the multi-criteria network selection solution is done by considering the user requirements, the QoS, and the operator points of view. The proposed scheme is scalable and is able to handle any number of RATs with a large set of criteria. The simulation results show that the proposed solution has a better and more robust performance over the reference solutions.

In [68], authors describe two novel ranking schemes based on fuzzy logic. These schemes enable users to evaluate the correctness of combination different P2P-based grid networks. The authors used a fixed set of communally used attributes such as cost, capacity, and reliability. The proposed ranking algorithm is based on intuitive design of rules optimization by applying the boolean logic to capture the combinations of inputs that do not apply.

In [69], authors propose a fusion method based fuzzy logic for the multiple network schemes; the main advantage is the consideration of the relative importance of the different networks; the authors show that the proposed scheme improves the generalization capability significantly.

To summarize, the fuzzy logic theory has been studied since the 1920s, it was applied to many fields, from control theory to artificial intelligence. Most of the recent works using the fuzzy logic are combined with MADM methods [70] [71]; generally, in the network selection area, the use of fuzzy logic as a core of the ranking scheme is not widely adopted; instead of that, fuzzy logic has always been combined to the MADM.

2.6 Analysis and discussion

In this section, we begin with a classification scheme in Figure 3.1 which regroups the methods used to solve the network selection problem, after that, we discuss their advantages and shortcomings.
In our study, we focus on the well known approaches used to solve the network selection; in this chapter, we only study the famous families and approaches used extensively in the literature. Other methods may exist, but they are not very well known and hence, rarely

used.
Figure 3.1. Regroups the approaches described in this study. The summit of the flow

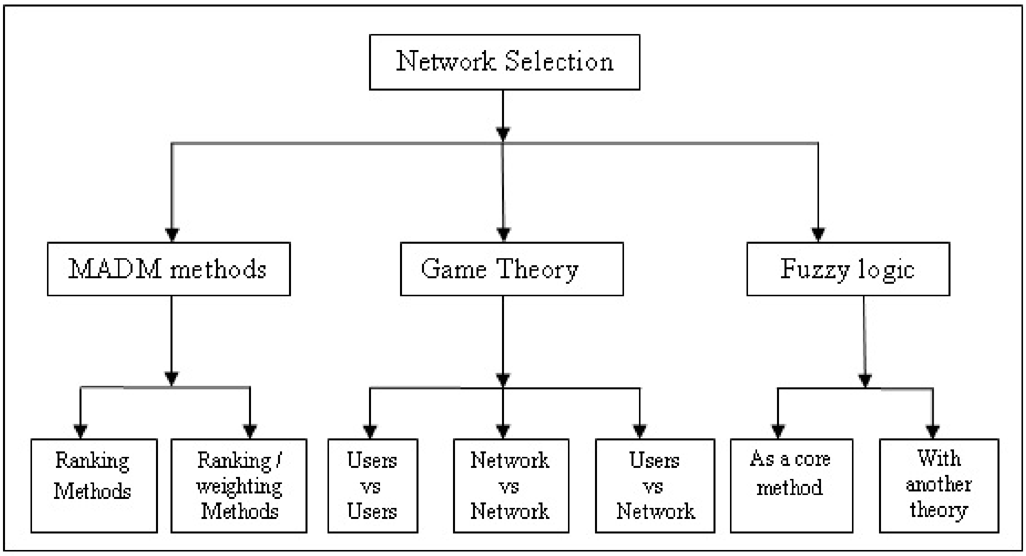

Figure 2.6: Recap of the used methods in NS problem

chart is the NS problem. In the second stage we have three approaches and families, namely MADM, game theory and fuzzy logic, each family has its own specific methods that are regrouped in the third stage.

- Ranking methods

It regroups TOPSIS, SAW, WPM and others. These methods require another method to get the weight vector. Such methods suffer from a lot of problems like the rank reversal phenomenon, and the influence of a bad value for a criterion on a good value of another criterion.

- Ranking/Weighting methods

A method such as AHP contains a process of getting the weight vector. For AHP, it uses the eigenvector method. These methods suffer from the high complexity and the rank reversal which can occur here as well. Because, it is a problem inherited from MADM approach.

- Game theory

Game theory has been used also in this area of research. There are three possible scenarios of the game, each of them can be either cooperative or non cooperative.

- The game can be played between users: here we risk the problem of congestion due to the selfish behaviour of users.
- The game between networks and users: in this scenario, the cooperative behaviour is preferable to avoid competition between the different players.

- The game between networks needs an authority which manages the game because the objective of the networks is the maximization of the network utilization, nowadays, all operators propose a variety of radio technologies so this approach can be seen as the most suitable one.

In general, we can say that the game theory is a good tool to model the network selection problem: the main issue can be the computation time due to the relatively high complexity and this issue can be avoided in the case of game between networks.

- Fuzzy logic

This discipline has been adapted to solve the network selection problem but, generally, it is not used alone. It needs to be complemented by the MADM methods or the genetic algorithms.
Table 3.7 contains a recap of the discussed methods in this study, it gives both the advantages and drawbacks of each approach.

As a summary, it is well known that wireless LANs provide a high speed internet access, while the cellular technologies generally have more ubiquitous coverage. Consequently, a suitably equipped mobile user might be able to use both wireless LAN and a cellular technology for Internet access. Thus, this mobile user might want to use a wireless LAN connection whenever one is available plus in the case of low mobility, i.e. in the pedestrian case, and to use a cellular connection when the wireless LAN is unavailable and/or in the case of moderate/high mobility.

Approaches	Kind	Advantages	Inconvenient
MADM	-Ranking methods.	-Easy to understand and use. -Relatively good results. -Low complexity.	-Rank reversal phenomenon. -Lack of a weighting process. -Different behaviour with different applications.
	-Weighting/Ranking methods.	-Weighting process.	-Rank reversal phenomenon. -High complexity.
Game Theory	-Users vs Users	-No rank reversal phenomenon.	-Selfish behaviour of users. lead to the congestion.
	-Users vs Networks		
	-Networks vs Networks		
Fuzzy logic	-As a core method	-Overcome the MADM's drawbacks.	-Few works in the literature. -Time consuming due to coupling other approaches.
	-With another approach		

Table 2.6: Summary of the discussed methods

2.7 Summary

In order to achieve the ABC concept, users with multi mode mobile equipment must be able to select a target network that meets their needs. This assumption implies a multi-technology environment. This chapter presents a review of the popular used method in this field; we first talked about the network selection problem and the utility functions, after that, we presented the most popular theories used basically the MADM methods, game theory and fuzzy logic. In this study, we managed to provide some critics to the discussed methods; regarding the MADM, we highlighted the well known problem of rank reversal and the influence of one parameter over another in some cases. The major problems of the game theory can be the congestion situation in the case of the game between users and the relatively high computation complexity. Furthermore, fuzzy logic needs another method to make decisions; it is generally coupled with the MADM or the genetic algorithms. Obviously, other approaches might exist in the literature, but in reality, researchers focus on the exposed methods, which represent the core of the treatment. Thus, the other approaches are used only in combination with the exposed methods. In this chapter, we give our critics of all mentioned methods seeking to make these solutions more robust and flexible.

Chapter 3

First contribution: Network selection

Contents

In this chapter, we propose a modified-SAW function to deal with the drawbacks of the existing solutions. Indeed, the existing MADMs methods suffers mainly from the famous problem of rank reversal once an alternative is added or removed, other problems occur in the legacy MADMs such as the sensitiveness to the user preference in TOPSIS and the penalization of the alternatives with poor attribute values in WPM. Our simulations brought out the lakes of the traditional MADMs approach in the context of network selection. We modify the SAW method intelligently and we use it to solve the network selection problem. Finally, we compare the performance of our solution with the previous works in different scenarios. The simulations show that our proposal outperforms the other existing methods. Radio access Technology, Always Best Connected , network selection, Multiple Attribute Decision Making, weighted sum.

3.1 Introduction

Up to the third generation network, the radio access network was mainly homogeneous and the user has to connect to only one RAT. After that, the development of the network technologies has led to an impressive growth of the internet applications and services, and an increase in the mobile user's industry. People now are equipped with Smart-phones and are seeking for the Always Best Connected (ABC) concept. It is obvious that no single network technology can sustain that, therefore, it was necessary to change the whole design, i.e. switching from the homogeneous systems to heterogeneous ones. The aim of the fourth generation network is to achieve the ABC concept by offering the mobile users the ability to take advantage of networks having different architectures and performances.

The network selection in a heterogeneous environment can be mapped to the Multiple Attribute Decision Making MADM problems, since it involves a number of criteria. The MADM approach has been widely used to solve the NS problem [5], [1], [2]. Other methods like fuzzy logic and game theory [6] [7] [8] have also been used to solve the network selection. Many studies in the literature propose algorithms that select a network among a set of candidate networks. These algorithms try to find and connect to the best networks and this is not always good solution because, several users may choose the same network simultaneously. Thus the network becomes loaded and potentially unreachable. So, it would be interesting to have the optimal rank of networks and then select the best network based on load. This is the motivation of this work.

Our study seeks to find the best network in addition, we give the users the other alternatives in the case when the best network is unreachable, i.e. we give the list of the ranked networks to select one of them according to the network load. In this chapter, we try to bring out our solution based on the modified-SAW and we make a comparison with the legacy MADM methods.

The rest of the chapter is organized as follows Section 2 presents the problem formulation and the related works and in Section 3 we give the mathematical description of the proposed method. In Section 4, the simulations are presented and the performance analysis is provided, finally a conclusion and future works are given in the section 5.

3.2 Network selection

In this section, we present the context of network selection; we discuss the involved criteria in the process and the steps of the NS procedure. After that, we give an overview of the previous research in the field of NS. Subsequently, a detailed description of the methods used in this chapter can also be found in this section.

3.2.1 Problem context

In the Next Generation Network, the heterogeneous wireless access is a promising feature in which the users are sufficiently flexible to select the most appropriate network according to their needs. In these circumstances, the network selection has an important task for the smooth functioning of the whole communication system. Indeed, the NS process consists of switching between RATs to serve the user with the best network. So, when a user with a multi mode terminal discovers the existence of various RATs within the same area (Fig.1), he should be able to select the best network and operate-in to get the desired service.

The different RATs provide different characteristics in terms of delay, jitter, throughput and packet loss rate. For this reason, in the context of heterogeneous systems, selecting the best RAT is a hard task. Many parameters influence the decision process of the best RAT [3] such as the battery level status, the energy required for the requested services, the Signal to Interference plus Noise ratio received (SINR), the cost to pay, the bandwidth required, the user preferences (QoS) . . . etc. The NS procedure is the decisive part of the vertical handover

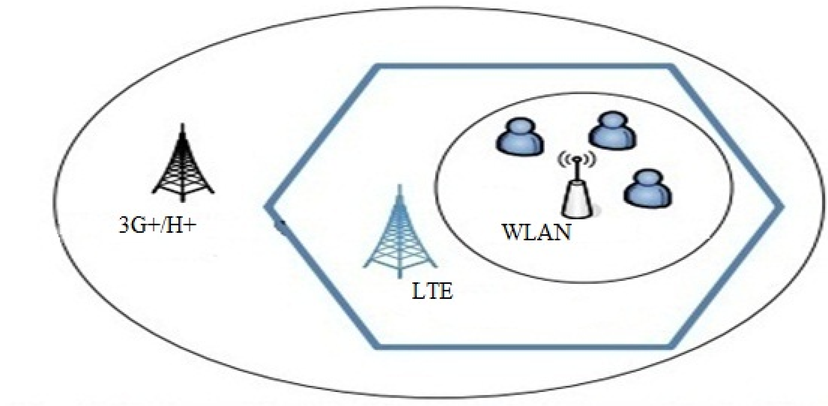

Figure 3.1: Heterogeneous wireless system

process VHO, it can be either centralized (network-centric) or decentralized (user-centric).

For the centralised approach, the operator controls the whole process and makes the decisions, the users obey these decisions and executes them, this can be a good strategy to avoid problems like the selfish behaviour of users who always try to get the best RAT resulting in congestion. On the other hand, this approach assumes a situation of a single operator. With multiple networks RATs, this strategy can not be used in case of multiple operators.

The user-centric approach, users make decisions by themselves; this approach is decentralized and can easily generate a congestion situation because of the selfish nature of users. Nowadays, almost all operators offer 3G and 4G radio access and even the Wi-Fi connections, so, the centralized approach is suitable to use.

The network selection procedure consists of the following parts:

- **Monitoring step:** the objective of this step is to discover the available RAT, collect the network radio conditions, and the other characteristics of the RATs. In this stage, some of the parameters are estimated and others are calculated.
- **Decision step:** the network selection decision is started. The choice of the best network is based-on the monitoring process and other parameters provided by the mobile device, the network and the user preferences. At this stage, a decision algorithm is used to rank the different RATs.
- **Execution step:** consists on connecting to the target RAT.

3.2.2 Related work

Several works in the literature treat the NS problem, we have already detailed that in the previous chapter. Meanwhile, we just make a little reminder of some works which they were used the MADM to solve the network selection problematic, these works focus on the optimization of the network selection decision for the users in order to support many services with the best QoS in a smooth manner. In the following we present some of these related work on field of network selection.

In [37] and [38] the authors have used the Simple Additive Method SAW to get a ranked list of networks, while in [34] the authors make a mix between game theory and SAW method. The main benefits of SAW are the simplicity and the low complexity, meanwhile, it has two major drawbacks, first: a parameter can be outweighed by another one, second: the rank reversal phenomenon that represents a problem of the entire MADM approach.

A comparison study is performed in [37], indeed, the authors compare the performance of the vertical handover VHO using SAW and TOPSIS. The authors conclude that the TOPSIS method outperforms the SAW method. In general, we can say that the compensatory methods such as TOPSIS avoids the problem where a parameter can be outperformed by another one, this is allowed using some trade-off between the criteria. This means that a poor value in one criterion is neglected by a good value in the other criteria, this concept provides more credibility for whole NS process comparing to the non-compensatory methods, which use the thresholds and do not allow any kind of trade-off between criteria.

A comparative study between SAW and WPM in the context of vertical handover is done in [42]. The authors use the relative standard deviation as a metric of comparison and they obtain a conclusion that WPM is better than SAW. In [43], the authors use the WPM method in the context of heterogeneous systems, their conclusion is that the WPM method is a more robust approach for dynamic decision making while it penalizes the attributes with poor quality to a greater extent.

The AHP method is applied in [45] to rank the importance of various criteria used and to compare the desirability of different Internet advertising networks. The proposed model provides an objective and effective decision model for advertisers to be selected. Authors in [44] have compared the original AHP with a modified version called Fuzzy AHP, the important criterion used is the Quality of Experience (QoE); indeed, the authors used the fuzzy complementary matrix and the fuzzy consistency matrix to relax the consistency requirement in the conventional AHP. The Numerical results show that the proposed Fuzzy AHP scheme outperforms the conventional AHP scheme.

Other approaches were used in this context, like game theory and fuzzy logic. In [6] authors present a non cooperative game theory approach for 4G environment; the authors designed a model of game between networks; they state that in the case of selfish behaviour of users, they must enhance the utility function with other parameters. A Bayesian game with incomplete information is used in [7] where users represent the players. The information is incomplete, which means that users have a portion of information about other users. The authors show that a Nash equilibrium is reached. A vertical handover solution in a homogeneous environment is proposed in [8]. The aim was to choose the best Access Point.

In [72], Authors claim that QoS criteria are highly dynamic; hence, they affirm that the existing static optimization based solutions might not be as effective. Therefore, they proposed a network selection mechanism based on sequential Bayesian estimation, which takes into account dynamic QoS factors. The drawback of the solution is the only consideration of the QoS factors while criteria like mobility, cost and energy were not considered.

3.2.3 MADM methods: Mathematical description

In this subsection, we describe briefly the well-known methods used to solve the network selection, these methods are used in Section 4 to make the comparison with our proposal.

MADM approach is a famous mathematical approach in the context of preferential decisions; it treats problems involving many decision criteria and many alternatives. This branch of decision making is widely used in various fields such as, the economy sector [33], [34], [37]. Indeed, ordinary people in their daily life use this approach, for example, to buy a car or a house having different characteristics . . . etc.
This approach MADM is very suitable to the network selection problem because of the multi criteria nature of the NS problem [35]. In [73], the authors present the basics of this approach:

- **Alternatives:** represent the different choices of actions available to the decision maker. The set of alternatives is assumed to be finite. In the NS scenario, the alternatives are the different RATs.
- **Set of criteria:** called also goals, it represents the attributes used in the decision making process. The attributes might conflict with each other because they represent the different dimensions from which the alternatives can be viewed; for instance cost may conflict with profit, etc. Here, an important aspect must be clarified, the MADMs are hard to solve because of the possible different units of the attributes. So, the attributes are heterogeneous and do not have the same unit. The solution is the use of the normalization methods to eliminate the unit. For the NS scenario, criteria are throughput, jitter, delay, cost ... etc
- **Weights:** it means the importance of the criterion in the decision process, whereas the sum of the weights is equal to one. Different methods are used to determine the weight values.

Finally, we get a decision matrix representing the system, where the columns are the criteria and the lines are the alternatives. Several methods use the MADM philosophy have been proposed in this context, such as: Simple Additive Weighted (SAW), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), Weighted Product Model (WPM), Analytic Hierarchy Process (AHP), Grey relational analysis (GRA). . . etc [34]. In the following we give a general description of these algorithms.

- **Simple Additive Weight SAW**

SAW is a well-known method for the case of multiple criteria systems. It assumes that data treated have the same units. So to get a comparable scale among parameters, it is mandatory to normalize the data for each parameter [34], [37], [36]. Finally, the alternative with the highest value is selected. The mathematical formulation of SAW is:

$$\mathrm{R}_{SAW} = \sum_{i=1}^{n} (\mathrm{w}_j \times \mathrm{r_{ij}}) \tag{1}$$

R_{SAW}:is the value of each alternative. w_j is the weight value of the criterion j.
n is the number of the decision criteria.
$r_{ij} = \frac{m_{ij}}{max(m_{ij})}$ is the normalized value of the criterion j and the alternative i. Then, the chosen alternative is the one with the maximum value.

- **Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)**

TOPSIS is an aggregating compensatory method based on the concept that the chosen solution should have the shortest geometric distance from the positive ideal solution [33], [39] and the longest geometric distance from the negative ideal solution. The TOPSIS method consists mainly of these actions: normalizing the data, computing the geometric distance between each alternative and the ideal alternative. The TOPSIS process is as follows:

$$\mathrm{R}_{TOPSIS} = \frac{\mathrm{D}_P}{\mathrm{D}_P + \mathrm{D}_N} \quad (2)$$

where
$D_P = \sqrt{\sum_{i=1}^{n} \left(w_j^2 \times (r_{ij} - b_j^P)^2\right)}$ and
$D_N = \sqrt{\sum_{i=1}^{n} \left(w_j^2 \times (r_{ij} - a_j^N)^2\right)}$
The w_j represents the weight value, r_{ij} is the normalized value of the parameter j and the network i. b^P represents the best alternative, the a^N is the worst alternative and n is the number of the decision criteria.

- **Weighted Product Model (WPM)**

WPM called also Multiplicative Exponential Weighting (MEW) is similar to SAW, [33], [41]. The difference is the replacement of the addition operation in SAW with the multiplication operation in WPM; each decision alternative is compared with the others ones by multiplying a number of ratios, one for each decision criterion. Each ratio is raised to the power equivalent to the relative weight of the corresponding criterion. The mathematical description is as follows:

$$\mathrm{R}_{WPM} = \prod_{i=1}^{n} (\mathrm{w}_j \times \mathrm{r_{ij}}) \quad (3)$$

- **Analytical Hierarchy Process AHP**

AHP considers the decomposition of a complicated problem into multiple hierarchical simple sub-problems [33]. The AHP steps are as follows:

- First, decompose the problem to set hierarchical sub-problems, where the top node is the final goal and in the lower nodes are the alternative solutions to the problem.

- Second, determination of the relative importance of the criteria with respect to the objective. In each level the decision factors are compared pairwise according to their levels of influence with respect to the scale shown in Table 3.1.

- Third, calculating the weights in all the hierarchy levels.

- Fourth, the AHP method in general is coupled with Grey Relational Analysis (GRA) to rank alternatives.

Definition	Importance
Equal importance	1
Moderate Importance	3
Strong importance	5
Very strong importance	7
Extreme importance	9
Intermediate values	2,4,6,8

Table 3.1: AHP scale of importance

In this sub-section, we presented the famous methods of the MADM approach. These methods are widespread to solve the network selection problem, however, these methods present certain shortcomings. In the next section we proposed a solution for network selection, this solution solves the shortcomings of the legacy MADM methods.

3.3 NS2 and NS3 simulations

In this section, we present an important part of the work which consists on extracting the data values used latter in our simulations. The targeted values are, data rate, delay, packet loss rate. In this section we will show how we have generated these values which will be used in the entire document.

To be able to make simulations on the different RATs used in this thesis which are (Wi-Fi, HSPA, HSPA+ and LTE) and apply our proposed algorithms using the extracted values (data rate, delay and plr), we were forced to make simulation on these RATs. The major problem was the lack of a simulator that allows to make simulations overs the overall RATs used in this thesis so, we have decided to use different simulators and regroup the results while ensuring the same condition of the simulations (same number of simulation, same mobility model which is by the way "random mobility", same number of mobile users etc).

Each simulation on each simulator contains the different traffics targeted in this thesis namely: VOIP, video and best effort service. The users on each simulation are divided in this way, 40 % VOIP users, 40 % video users and 20% best effort users. The reason of a such division of users is because nowadays, most traffic in internet is dedicated to the voice and video communications. We mention that this division doesn't affect the performance of the simulation, it is just a way to have a realistic scenario.

In this thesis, we have used three simulators namely Network Simulator 3 (NS3) [74] to get results for Wi-Fi networks and LTE, LTE simulator (LTESim) [75] to validate the results of LTE network generated by NS3 and Network simulator 2 (NS2) with eurane module [76] for 3G and HSPA networks.

Based on the scenario presented described, we get a margin values for The throughput, end to end delay and packet loss ratio 3.2. The values presented in table 3.2 are margin vales which means that the values used latter in our work are a random values in between the intervals.

3.3.1 NS3 simulator

NS3 is a discrete-event network simulator, targeted primarily for research and educational use. Free software, it is licensed under the GNU GPLv2 license, and is publicly available for research and development.

The goal of the NS3 project is to develop a preferred, open simulation environment for networking research. It should be aligned with the simulation needs of modern networking research and should encourage community contribution, peer review, and validation of the software.

The NS3 project is committed to building a solid simulation core that is well documented, easy to use and debug, and that caters to the needs of the entire simulation work flow, from simulation configuration to trace collection and analysis. Furthermore, the NS3 software infrastructure encourages the development of simulation models which are sufficiently realistic to allow NS3 to be used as a real-time network emulator, interconnected with the real world and which allows many existing real-world protocol implementations to be reused within NS3.

The NS3 simulation core supports research on both IP and non-IP based networks. However, the large majority of its users focuses on wireless/IP simulations which involve models for Wi-Fi, WiMAX, or LTE for layers 1 and 2 and a variety of static or dynamic routing protocols such as OLSR and AODV for IP-based applications.

NS3 also supports a real-time scheduler that facilitates a number of "simulation-in-the-loop" use cases for interacting with real systems. For instance, users can emit and receive NS3-generated packets on real network devices, and NS3 can serve as an interconnection framework to add link effects between virtual machines. Another emphasis of the simulator is on the reuse of real application and kernel code. Frameworks for running unmodified applications or the entire Linux kernel networking stack within NS3 are presently being tested and evaluated.

3.3.2 LTE Simulator

LTE-Sim is an open source simulator software designed to simulate uplink (UL) and downlink (DL) scheduling strategies in multi-cell/multi-user environments of LTE networks that take into account user mobility, Radio resource optimization, frequency re-use, adaptive modulation (AMC), and other significant aspects for industry and the scientific community. It was mainly developed by G. Piro and F.Capozzi at "Politecnico di Bari". It allows to simulate a telecommunication network according to the scenarios that the user defines to him, for example the simulation of a transmission in an urban macro-cells taking into account the multi-user, multi-service, mobility of users.

It encompasses several aspects of LTE networks, including both the Evolved Universal Terrestrial Radio Access (E-UTRAN) and the Evolved Packet System (EPS). It supports single and heterogeneous multi-cell environments, QoS management, multi users environment, user mobility, handover procedures, and frequency reuse techniques. Four kinds of network nodes are modeled: user equipment (UE), evolved Node B (eNB), Home eNB (HeNB), and Mobility Management Entity/Gateway (MME/GW). Four different traffic generators at the application layer have been implemented and the management of data radio bearer is supported. Finally, well-known scheduling strategies (such as Proportional Fair, Modified Largest Weighted Delay First, and Exponential Proportional Fair, Log and Exp rules), AMC scheme, Channel Quality Indicator feedback, frequency reuse techniques, and models for physical layer have been developed.

3.3.3 NS2 simulator

NS2 is an open-source simulation tool that runs on Linux. It is a discreet event simulator targeted at networking research and provides substantial support for simulation of routing, multicast protocols and IP protocols, such as UDP, TCP, RTP and SRM over wired and wireless (local and satellite) networks. It has many advantages that make it a useful tool, such as support for multiple protocols and the capability of graphically detailing network traffic. Additionally, NS2 supports several algorithms in routing and queuing. LAN routing and broadcasts are part of routing algorithms. Queuing algorithms include fair queuing, deficit round-robin and FIFO.

NS2 started as a variant of the REAL network simulator in 1989 (see Resources). REAL is a network simulator originally intended for studying the dynamic behaviour of flow and congestion control schemes in packet-switched data networks.

Currently NS2 development by VINT group is supported through Defense Advanced Research Projects Agency (DARPA) with SAMAN and through NSF with CONSER, both in collaboration with other researchers including ACIRI (see Resources). NS2 is available on several platforms such as FreeBSD, Linux, SunOS and Solaris. NS2 also builds and runs under Windows.

Simple scenarios should run on any reasonable machine; however, very large scenarios benefit from large amounts of memory. Additionally, NS2 requires the following packages to run: Tcl release 8.3.2, Tk release 8.3.2, OTcl release 1.0a7 and TclCL release 1.0b11.

3.4 The proposed algorithm

3.4.1 Critics for MADM

The MADM methods are widely used to solve the network selection problem, this is due to the multiple criteria nature of this problem; in addition, these methods are easy to use and understand and to implement, some of the MADM methods are also known with the low computational complexity. Meanwhile, this approach has some drawbacks and they can be summarized in the following:

- The MADM methods doesn't have the same performance toward the different services (VoIP, Video Calls, web browsing), these methods have a good performance for VoIP service and a bad performance for the best effort services. This instability involves great problems because users use different services. (add references).

- SAW, WPM and TOPSIS have a relatively low complexity, meanwhile, other methods such as AHP, PROMETHE, VIKOR are complicated to implement, these method contains many complicated mathematical concepts such as eigenvector in AHP method. The eigenvectors concept are used in AHP as to weight the criteria, and this is another drawback of this approach. MADM methods do not have a common method to weigh the criteria which results on the use of different weighing method, thus different performances in term of the ranking results and the complexity of ranking process.

- The big problem of the MADM methods is the ranking abnormality called also ran reversal, indeed, the MADM suffers from the problem of ranking abnormality, most of them. In decision-making, a rank reversal is a change in the rank ordering of the preferability of alternative possible decisions when, for example, the method of choosing changes or the set of other available alternatives changes. The issue of rank reversals lies at the heart of many debates in decision-making and multi-criteria decision-making, in particular. So, simply, it is a phenomenon which occurs in the MDAM methods when an exact replica or a copy of an alternative was introduced or eliminated in the system.

  It is hard to tell if a particular decision-making method has derived the correct answer or not. Such methods analyse a set of alternatives described in terms of some criteria. They determine which alternative is the best one, or they provide relative weights of how the alternatives perform, or just how the alternatives should be ranked when all the criteria are considered simultaneously. This is exactly where the challenge with decision making exists. Often it is hard, if not practically impossible, to determine whether a correct answer has been reached or not. With other computational methods, for instance with a job scheduling method, one can examine a set of different answers and then categorize the answers according to some metric of performance (for instance, a project's completion time). But this may not be possible to do with the answers derived by most decision making methods. After all, determining the best decision making method leads to a decision making paradox.

  Authors in [46] has shown that the rank reversal problem occurs in most of the well-known MADM methods, this problem has been addressed in other works [47], [48] by modifying methods, but the original versions of the MADM methods suffer from the rank reversal phenomenon.

For all of these reasons, we can say that MADM are a good and acceptable solution to solve the NS problem, but the lack of general method that serves all kinds of services and the rank reversal phenomenon is a big problem. So, our objective is to propose a solution for the network selection problem which present a higher performances while it avoids the drawbacks of the MADM.

SAW method particularly is simple to understand and to implement, it have also a low computation complexity. So, we decided to enhance this method and make some modification to (1) have a good performance compared to the the original versions of the MADM methods, (2) eliminate the problem of rank reversal occurring in the legacy MADM methods , (3) keeping the low complexity of the original version of the SAW method.

3.4.2 The modified-SAW algorithm

The purpose of the proposed network selection procedure, named modified-SAW, is to ensure that the user gets a good quality of service during his his session and to promote the even distribution of the users on the networks. The proposed solution assigns the user to the best network from the available networks at the present instant while the selected network is not loaded. This process is repeated at several instances until the user's session service is ended. To evaluate our proposal, we will compare the performance of our solution with the existing MADMs solutions, [77].

So, when a user demands for a particular service, it sends a request to the operator which contains some information like the battery level and the service required. The other parameter needed for the network selection procedure are generated by the operator. Then, the operator triggers the process of ranking the available networks, after that, the operator forwards the result to the user who selects the best alternative. The result forwarded to the user contains the list of networks ranked from the best network to the worst one. Obviously, the user will choose the network having the best rank while the network is not loaded. A loaded network normally can not gets the best rank, simply because a loaded network has a higher delay and lesser throughput. Thus, it will propose a bad QoS performance. As a result, it's quite unlikely to get a loaded network as the best ranked network . But, in case of the best ranked network is a loaded one, the user take the next network on the result list and so on, until selecting a not loaded network. Figure.3.2.

In our description we have two sides: the mobile user which seeks for the best RATs and the operator who triggers the ranking process of the available networks. More details on the proposed solution are given in the following.

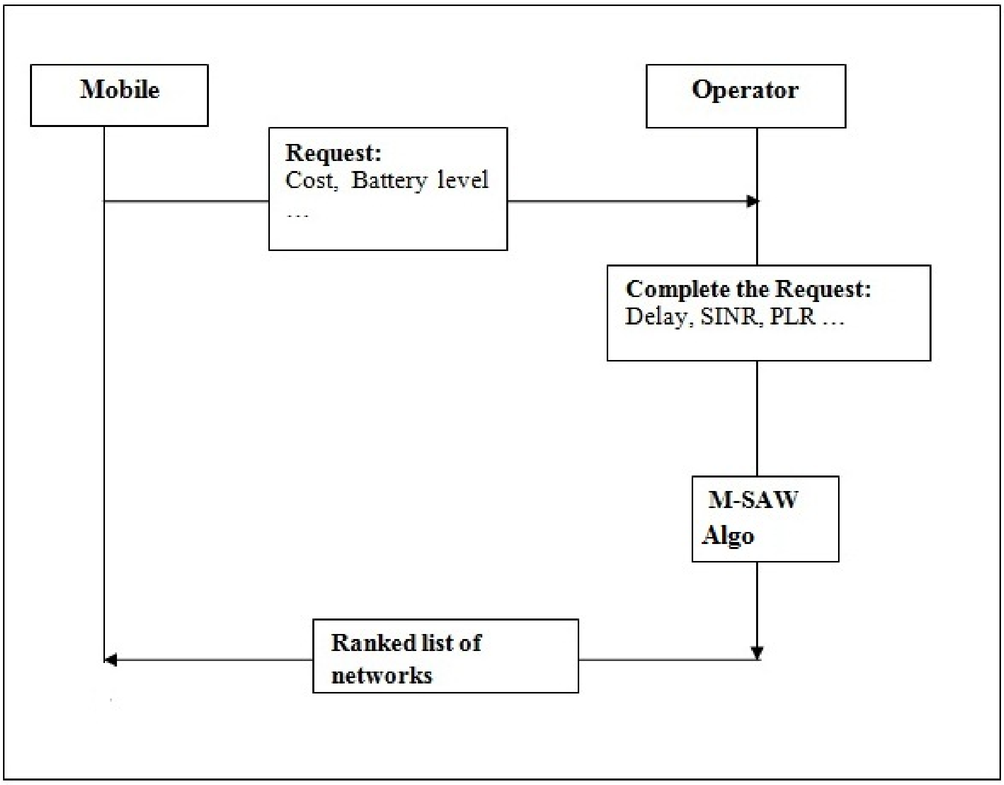

Figure 3.2: Process steps scheme

The idea behind the objective function is simple. We start by dividing the input matrix to a set of column vectors, so, we will get a group of vectors equal to the number of criteria.

For each vector, we rank the networks according to their data values. Then, the network receives a local income equal to the mathematical multiplication between the α minus, the rank value for the network and the weight value of the criterion where α is a fixed value equal to the number of criteria. This process is repeated for the other criteria equation 8. The total income value is equal to the summation of all the local incomes for each network equation 7.

We use the weigh value of each criterion to differentiate between the criteria. It is well-known that the delay and the PLR criteria are the most important comparing to the other criteria. For example, a network proposing a delay value of 30 ms and another proposing a delay of 100 ms don't have the same revenue here in this example having a small value of delay is good and offers a higher gain to the network. The process of evaluating the criteria is repeated until exhaustion of the criteria.

We formulate the system as a minimization function in which the lowest value for each criterion gives the best ranking order for the networks and then gives it the highest local gain. This process is repeated until all the criteria are evaluated. Its representation is as follows: For each network 'i', we do:

$$R_i = \sum_{j=1}^{m} income_{ij}. \quad (4)$$

$$income_{ij} = (\alpha - k_{ij}) \times w[j]. \quad (5)$$

$$k_{ij} = min\,(Vect_{ij}). \quad (6)$$

Where :

α: Is a fixed integer number equal to the number of alternatives.
k_{ij}: It represents the rank order of the network i for the criterion j.
$Vect_{ij}$: Is the column vector from the matrix mat where j is fixed.
j: Is the alternative and j is the criteria. $mat[n][m]$: Is the input matrix, Table 3.3.
More details can be found in Algorithm 1.

The proposed solution is equitable and stable. The stability of the solution consists on the non changing of the solution only if the input data are changed, this means that for one input matrix, the incomes value can not be changed at the end of the process only if the input data are changed.

The solution is also equitable, this concepts means that our process gives more relative incomes for the network with a good data value for the user and in the same time, the use of the weight value give more importance to the highly important criteria and differentiate between criteria.

The best case of one network 'i' is when it has the minimal value (we make the system as a minimization problem) for the criterion 'j' this means the highest local ranking, thus $k_{ij} := 0$ and the revenue is: $R_{ij} = \alpha \times w[j]$.

The worst case is when the network has the worst local ranking for the criterion, i.e. $k_{ij} := number\ of alternatives\ - 1$ i.e. $k_{ij} := \alpha - 1$, the revenue is $R_{ij} = w[j]$.

The use of the weight concept is taken from the MADM approach, this concept allows us to give more sense to our objective function which ranks the networks based their values for each criterion. The weight vector brings out the significant criteria and distinguish them from the other less important criteria depending on the application requirements, and therefore assigns higher incomes to the associated networks, this concept has advantages:

- Avoid the situation in which a network that has a good value for a non important criterion will have the same revenue as another network with a good value for an important criterion. This situation exists in the SAW method for example.

- The weight concept is the representation of application requirements in the system and that's how we distinguish between applications because each one has specific requirements. VoIP requires a minimum time delay and PLR for the video application, in addition to the requirements of VoIP service; it demands also a good throughput. For the best-effort applications, they accept the existing conditions, but the cost criterion is important. The information is transformed into digital values with the eigenvector method.

In our study, we use many parameters such as: cost, energy consumption, average throughput achieved, average delay, average PLR and network load. An example of our matrix is presented below. The parameters presented in our matrix are the margin values that the simulations bring to us.

The applications are represented by the weight values in the system. For this reason we have one vector weight for each type of application. Now, this vector is obtained based on the QoS important parameters for each application. For VoIP service, we know that the time delay, jitter and PLR are the significant parameters, therefore, we give higher values for these two parameters. For streaming video services, the important parameters are the data rate, PLR and even the time delay. For best effort application, the parameters have almost the same level of importance.

The modified-SAW function consists on: the decomposition of the input matrix in a column vectors, and use the *Sort function* to rank the networks of these vectors and return their indices ranked from the best network to the worst. The ranked indices are stocked in the *Tabind* vector. After that, we calculate the local income for each network using the equation 8. The process is repeated for all the vectors and like this the local incomes become a total incomes when the criteria is reached. The network selection flow chart is as follows, Fig.3.3.

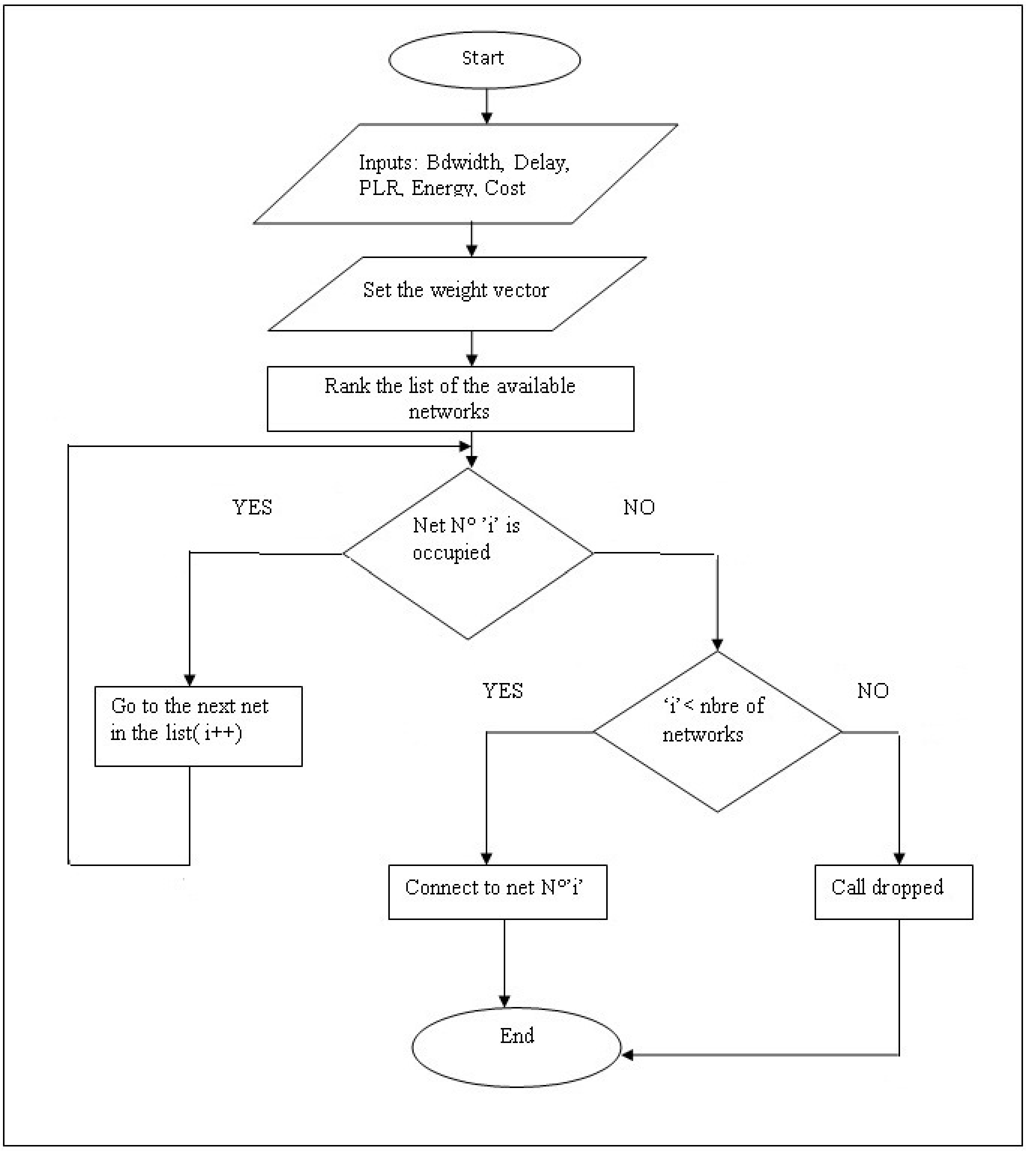

Figure 3.3: Network selection process diagram

Algorithm 1 Modified-SAW function

Data: The matrix mat [m][n] and the weighting vector w[], $\alpha = m$
Result: Income vector for each network
for $(j = 0; j < m; j++)$ **do**
  for $(i = 0; i < n; i++)$ **do**
    $tab[i] = mat[i][j];$
  end
  $tabind = Sort(tab);$
  for $(k = 0; k < Tabind.length; k++)$ **do**
    $income[tabind[k]] = income[tabind[k]] + (\alpha - k) \times w[k]$
  end
  $return\, income[];$
end

Algorithm 2 The sort function of the M-SAW algorithm

```
Data: vector tab[]
Result: Sort function
int t[][]
 double min;
 int ind, cpt = 0, deb = 0;
 int a = 0, nb = 0, val = 0;
 while (deb < t.length) do
     while (tab[a] < 0) do
         a++;
     end
     min = tab[a]; ind = a;
     for (int i = 0; i < tab.length; i++) do
         if (min > tab[i] and tab[i] >= 0) then
             min = tab[i]; ind = i; nb = 0;
         end
         else if (min == tab[i] and a != i) then
             nb++;
         end
     end
     for (int p = deb; p < deb + nb + 1; p++) do
         tab[hah] = -1; hah++;
         for (int h = 0; h < 2; h++) do
             if (h == 0) then
                 t[p][h] = val;
             end
             else
                 t[p][h] = ind + cpt; cpt++;
             end
         end
     end
     val++; cpt = 0;
     deb = deb + nb + 1;
 end
return t;
```

In this work, we use two examples of each type of network to bring out the rank reversal problem and solve it with the proposed method. The values of the matrix used in for processing Table 3.3 is generated randomly based on the Table 3.2.

	Bandwidth	**Delay**	**PLR**	**Energy**	**Cost**
Wi-Fi	1-11	100-150	0.2-8	-	1
3G	1-14	25-50	0.2-8	-	5
LTE	1-100	60-100	0.2-8	-	2

Table 3.2: The matrix model

The energy consumption parameter is related to the battery of the mobile device as well as the duration of the session of the application, i.e. with a high-level of the battery, the power consumption will not be as important in the system because the mobile equipment can ensure that the session will not be interrupted. In the opposite case, if the battery level is low and the duration of the session is long, this means that the session might get interrupted due to battery drain. So in the situation when the battery is low, this parameter is very important and will have a higher weight in the system to avoid depletion of the battery and therefore a break in the session.

Based on [78], the energy required to execute the user's request will be calculated for each RAT. The energy consumption parameter is set using the following equation [78]:

$$P[mJ/s] = \alpha_u \times th_u + \alpha_d \times th_d + \beta \quad (7)$$

α_u, α_d and β are parameters with different values from one RAT to another [78], th_u and th_d are uplink and downlink throughput. The power in mJ/s means that the energy depends on the session duration time.

3.5 Performance evaluation

In this section, we evaluate the performance of the proposed model and we compare it with the MADM methods described above based on the input data from table 3.2. In this study, we consider three types of services, VoIP, video service and the best effort service, these services are represented in the system by the weigh vectors. Indeed, each type of application has a special QoS requirements.

Many users believe that more bandwidth will resolve the problem. But, having more bandwidth though may not work any-more. In packet-switched networks, quality of service is affected by various factors. The most important are the technical ones which include: reliability, scalability, effectiveness, maintainability. Many things can happen to packets when they travel from the source to destination, in the following we present the important technical factors:

- **Low throughput** Due to varying load from the different users sharing the same network resources, the bit rate (the maximum throughput) that can be provided to a certain data stream may be too low for real-time multimedia services if all data streams get the same scheduling priority.

- **Latency** It might take a long time for each packet to reach its destination, because it gets held up in long queues, or it takes a less direct route to avoid congestion. This is different from throughput, as the delay can build up over time, even if the throughput is almost normal. In some cases, excessive latency can render an application such as VoIP or on-line gaming unusable

- **Jitter** Packets from the source will reach the destination with different delays. A packet's delay varies with its position in the queues of the routers along the path between source and destination and this position can vary unpredictably. This variation in delay is known as jitter and can seriously affect the quality of streaming audio and/or video.

- **Dropped packets** The routers might fail to deliver (drop) some packets if their data loads are corrupted, or the packets arrive when the router buffers are already full. The receiving application may ask for this information to be retransmitted, possibly causing severe delays in the overall transmission.

These are the most important factor which influence the application in a wireless network. So, we use the weigh vector to represent these requirements in the system.

The study is composed of two parts:

- In the first part we compare our proposal with the above-presented MADM methods in the normal case where no RATs disappear in the middle of the selection process.

- The second part is the case where one network disappears from the list of the available networks, this case allows us to prove that MADM methods suffer from the rank reversal phenomenon, and this problem does not occur in our proposed method.

The following matrix, represented by the table 3.3, is the data input matrix where the values are randomly generated from table 3.2; this matrix is used in our tests.

	Bandwidth	**Delay**	**PLR**	**Energy**	**Cost**
N(0)	1.730	105.85	7.94	1.00	0.2
N(1)	5.076	134.88	6.70	2.6	0.2
N(2)	6.849	43.98	284	6.26	1
N(3)	6.329	32.15	3.05	5.86	1
N(4)	66.66	95.15	6.32	12.78	0.4
N(5)	62.5	99.73	5.80	10.28	0.4

Table 3.3: The input matrix

$$(mat - \lambda \times I) \times w = 0 \tag{8}$$

The weight values of each type of application VoIP, video service and best effort are generated using the Eigenvector method 8. We decided to use the Eigenvector method because it is already used in the AHP process; So, to be consistent and fair, we decided to use the same method to have the same weight vectors for all of methods.

mat: is the input matrix, λ is the eigenvalue, I is the identity matrix and w is the associated eigenvector containing the weights values. The table 3.4 contains the weights vector for each type of application.

We start with the first part of this study, the ordinary case, i.e. having all networks available.

App	**Bwdth**	**Delay**	**PLR**	**Energy**	**Cost**
VoIP	0.047	0.486	0.371	0.047	0.047
Interactive	0.458	0.101	0.302	0.074	0.063
Best effort	0.299	0.146	0.146	0.108	0.299

Table 3.4: The weight vectors

3.5.1 Simulation 1: All RATs available

In this case, all network are available, this means that the networks in the covered area of the user do not disappear and remains available. In this situation we consider the MADM network selection algorithms presented in the above section and we compared with the proposed M-SAW algorithm. The objective of these simulations is to show which algorithm performs better, i.e., if our program is more efficient or one of the other algorithms is more suited. These simulation concerns the three type of applications, including VoIP, video service and best effort services.

- **Case 1: VoIP scenario**

The first case of these simulation concerns the VoIP application, this later is represented via a weigh vector which gives more importance to the delays and packet loss, the throughput in the case of VoIP is not too important because the packet are small and they can be transmitted with a relative low throughput, this is taken from [79].

The results in table 3.5 concern the first case VoIP. The analysis of the table 3.5 is done considering the table 3.3 as data input, it gives the following results:

Method	**Rank**					
SAW	N(1)	N(0)	N(2)	N(3)	N(5)	N(4)
TOPSIS	N(3)	N(2)	N(0)	N(1)	N(5)	N(4)
WPM	N(3)	N(2)	N(0)	N(1)	N(5)	N(4)
AHP	N(3)	N(2)	N(5)	N(4)	N(1)	N(0)
M-SAW	N(3)	N(2)	N(4)	N(5)	N(0)	N(1)

Table 3.5: Ranking results for VoIP

In this study we seek for the total order of the networks and not only the network ranked as the best one because, the first network is susceptible be loaded quickly after a given time and then it will be unavailable, i.e. overloaded, so, it is necessary to get an optimal total ranking of the networks.

Based on results in Table 3.5, the methods TOPSIS, WPM and M-SAW give the same order for the first two networks. For the third position, our M-SAW method chooses the N(4) but TOPSIS and WPM choose N(0) while AHP chooses N(5). Now we will compare performance of N(4) and N(0) and N(5) to see what method has made the right choice.

N(0) has a delay of 105,85 and a PLR of 7,94. The N(4) has a delay of 95,15 and a PLR of 6,32. So, N(4) is better than N(0) and this is an illustration that our method outperforms the TOPSIS, WPM, and AHP Figure.3.4 and Figure.3.5.

In the case VoIP, we showed that the proposed method M-SAW gives the best ranking order compared to the MADM methods, table 3.5 and Figure.3.4 and Figure.3.5.

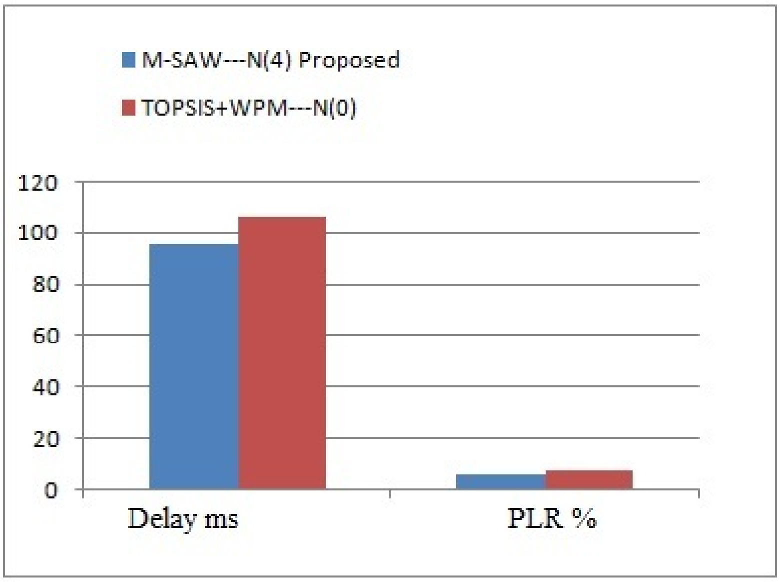

Figure 3.4: Delay and PLR comparison for N(0) and N(4)

The results in table 3.6 concern the second case (video service), using the matrix in table 3.3. From table 3.6, the SAW method has always a wrong ranking order. For the other methods, TOPSIS chooses N(5) and WPM, M-SAW choose N(4).

- **Case 2: Video scenario**

The second case of these simulation is dedicated to video service application, the associated weight vector gives more importance to the throughput, delays and packet loss in an equitable manner, the throughput is important because the packet are big and to be transmitted, they need a minimum required throughput, this is taken from [80].

Method	**Rank**					
SAW	N(1)	N(0)	N(2)	N(3)	N(5)	N(4)
TOPSIS	N(5)	N(4)	N(2)	N(3)	N(5)	N(0)
WPM	N(4)	N(5)	N(3)	N(2)	N(1)	N(0)
AHP	N(2)	N(3)	N(5)	N(4)	N(1)	N(0)
M-SAW	N(4)	N(2)	N(3)	N(5)	N(1)	N(0)

Table 3.6: Ranking results for video service

From Table 3.6, M-SAW and WPM select the N(4) and TOPSIS chooses N(5). N(4) has higher bandwidth and lower delay. The N(5) has a better PLR. But in this case of interactive scenario, the importance is given to the bandwidth and delay. So, the best choice is the N(4), Figure.3.5.

Furthermore, for the second order, the WPM method selects the N(5) and M-SAW selects N(2). N(5) has a bandwidth of 62.5, delay of 99,73 and a PLR of 5,80. N(2) has a bandwidth of 6,85, 43,98 ms delay and a PLR of 2,84.

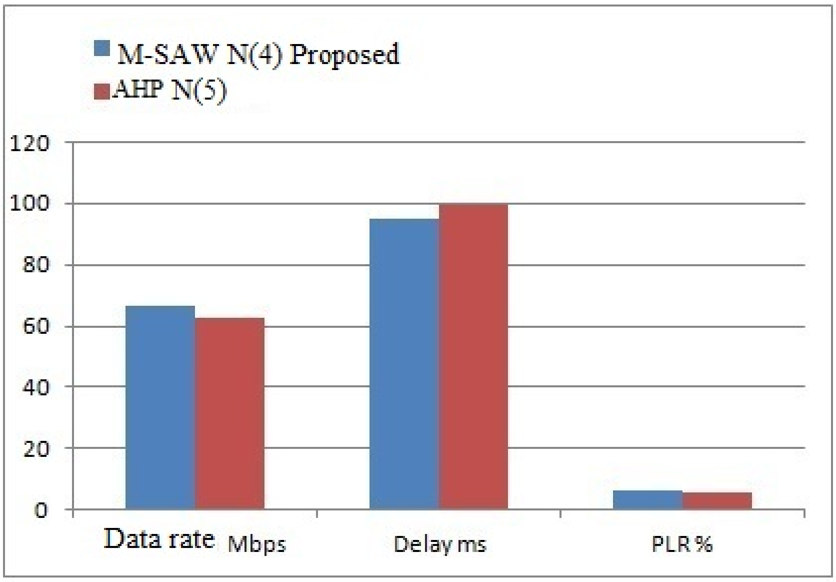

Figure 3.5: Comparison between N(5) and N(4)

Form these values, we see that N(5) has a larger bandwidth, but the bandwidth of N(2) is also good and can serve the interactive service (video service). In this case, we use the property of giving the user the minimum value that satisfy the application's requirement, this means that if the application requirements are satisfied, the network is considered as acceptable and the user can choose it.

For the other parameters (delay and PLR), N(2) is very good compared to N(5), Figure.3.6. Here we can see that the bandwidth parameter monopolizes the ranking decision in TOPSIS, i.e., the network having the best bandwidth forces the algorithm to neglect the huge delay and PLR. So, based on Figure.3.6, our method M-SAW brings the best choices with this service type.

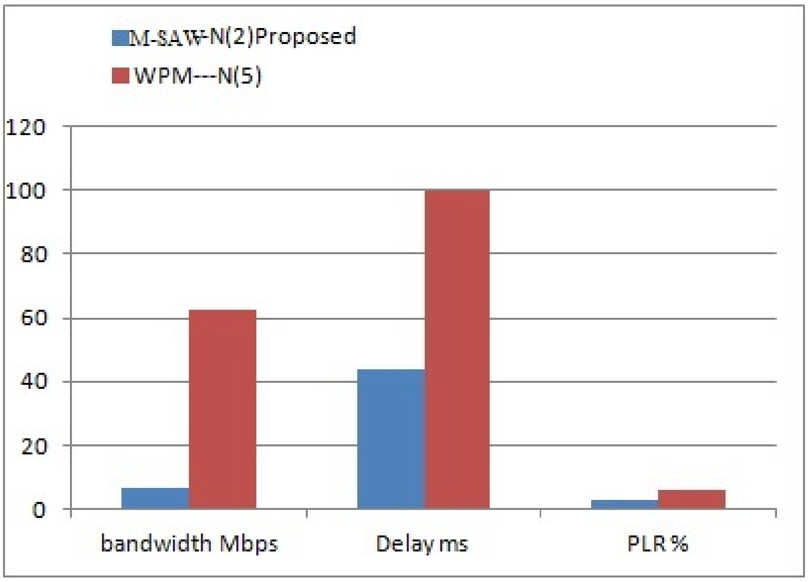

Figure 3.6: Comparison between N(2) and N(4)

- **Case 3: The best effort services scenario**

The third type of application used is the best effort application such as the mails and the navigation web, the associated weight vector does not give any special importance to any criterion, it gives equal values to all of the criteria with an equitable manner with a hope to get a higher throughput, the packets are transmitted and we have no idea on the received time neither the quality of the received packets.

Table 3.7 represents the results for the third case (data connections), based on table 3.3. WPM and M-SAW have the same first and second ranking orders; the third one, WPM chooses N(1) and M-SAW chooses the N(2). Form Figure.3.7, we see clearly that N(2) is better than N(1).

From all these cases we conclude that our proposed method M-SAW algorithm gives the best performances compared to the other methods. Our method chooses the best alternative at each stage of the process and eventually gives the best choices.

Method	**Rank**					
SAW	N(1)	N(0)	N(2)	N(3)	N(5)	N(4)
TOPSIS	N(5)	N(4)	N(2)	N(3)	N(1)	N(0)
WPM	N(4)	N(5)	N(1)	N(2)	N(3)	N(0)
AHP	N(5)	N(1)	N(4)	N(2)	N(3)	N(0)
M-SAW	N(4)	N(5)	N(2)	N(3)	N(1)	N(0)

Table 3.7: Ranking order for the best effort applications

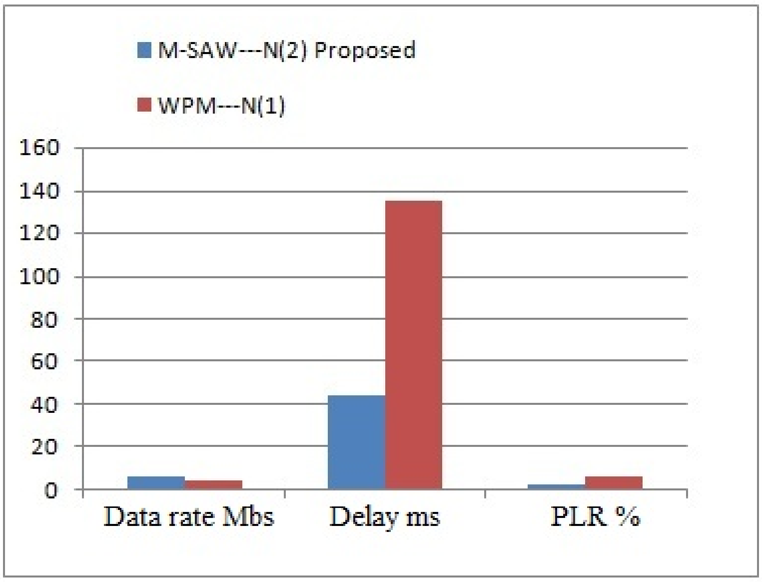

Figure 3.7: Comparison between N(2) and N(1)

In this subsection, we made a comparative study between our algorithm M-SAW and the MADM algorithms presented in the first sections. We simulate three cases VoIP, video service and the best effort applications, in the three cases, the proposed algorithm gives more accurate results, and at each time, it gives the adapted total rank of all the network. The figures presented in this subsection show that our algorithm gives the accurate total order of alternatives compared to others algorithm in all the cases.

3.5.2 Simulation 2: The rank reversal case

In this subsection, we investigate the rank reversal problem called also the ranking abnormality. This simulation has two objectives: first, we confirm that this problem occurs in the legacy MADM when we delete a line of the matrix, the second objective is to show that our proposed method avoid this problem. If the second objective is reached, we say that our proposed method is very well suited for the multiple criteria problem.

Now, we suppose the case that one network disappears, this case will allow us to study the rank reversal phenomenon. So, we will eliminate one network, for example the network N(4) from the table 3.3, and rank the remaining networks in the case of VoIP service, the results are shown in the table 3.8.

Results in table 3.8 show two things:
First, all the MADM (TOPSIS, SAW, WPM and AHP) methods suffer from the rank reversal phenomenon, this result matches the affirmation of the authors in [46], that all MADM methods suffer from the rank reversal phenomenon. This confirms that the legacy MADM are not suited for the network selection problem.

The second remark is that our method does not present the phenomenon of reversal's ranking when a network disappears. This result allows us to say the proposed M-SAW is adapted to solve the network selection more than the other MADM, we also add that the M-SAW has a low complexity, we just make some modification on the SAW method which is known with her low complexity too.

Method	**Rank**				
SAW	N(1)	N(3)	N(2)	N(0)	N(5)
TOPSIS	N(5)	N(3)	N(2)	N(0)	N(1)
WPM	N(5)	N(3)	N(2)	N(0)	N(1)
AHP	N(0)	N(1)	N(5)	N(3)	N(2)
M-SAW	N(3)	N(2)	N(5)	N(0)	N(1)

Table 3.8: Ranking order in the case of disappearing one network

To summarize, MADM give the ranking order for a multiple criteria problem, this order is not always optimal. It is possible that one method gives the correct best network and the user will try to connect to this network. But, it is not always possible to connect to the best network, it may be loaded given the number of users that select the best network. So, selecting the best network will easily load the best network and it does not accept connection requests. For this reason, it is important to have the correct ranking list of the available network, hence, if the best one is loaded, the user can connect to the second best network. Here, the algorithm must present the accurate ranking order and in the same time, it must avoid the ranking abnormality problem, because, if a network is loaded, it will be unreachable and deleted from the list of the candidate network. All these requirements are not present in the legacy MADM methods and at the same time are provided by our proposed method.

This process is done in the operator side to benefit from the operator's processing capacity and the permanent source of power which give us efficiency and speed at the same time. In addition, the operator has all the information concerning networks and users. This algorithm tries to find the most optimal total rank of networks and not only the best network from a list of networks. Having the right order of the network, allows us to be sure that at each instant, the user connects to the best available network among the existing ones.

A second advantage of this algorithm is that it works well in the normal case and in case where one network disappears. In the latter, the MADM methods present the problem in reverse order.

3.6 Summary

In the aim to find the best network at each instant, the idea was to rank the existing networks and get the optimal ranking order, in this case, the operator provides the users with the best network available in the ranked list of networks. In this chapter, we present our approach named modified-SAW in which the objective function is based on the relative ranking order of each alternative for each criterion at each round of the process, this basic idea allows us to get a greedy algorithm that gives good results. Results show that our proposed approach outperforms the existing used methods in the normal case, i.e. when all networks are available. Another test is done where a one of the networks disappears. Simulation shows that all the MADM methods present the rank reversal phenomenon, our proposed algorithm overcomes this phenomenon and stays coherent and brings the same ranking order whilst eliminating the disappearing network.

In the next chapter, we aim to study the complexity of the network selection algorithms, although our algorithm has a lower complexity and does not consume too much memory. Another perspective is to insert this part of network selection in the global handover vertical process and the consideration of the mobility case.

Chapter 4

Second contribution: Mobility-aware network selection

Contents

The coexistence of different Radio Access Technologies (RATs) in the same area has enabled the researchers to get profit from the available networks by the selection of the best RAT at each moment to satisfy the user requirements. In this chapter, we address a real world problem which is the frequent mobility of the users in heterogeneous networks. We present in this chapter a framework that allows the users to select the best networks for the whole call session especially from a mobility perspective. The framework consists of several steps, starting by the path prediction which is done using the Markov model order 2. The second step is to make the network selection on the zones of each predicted path, while the third step is to get the good RAT configuration for each predicted path. Finally, we use another function to select one of the best configurations to be used for all the possible used paths. The results show that our proposal performs very well by eliminating the unnecessary vertical handovers while maintaining a good Quality of Service (QoS).

4.1 Introduction

The next generation of wireless network is characterized by the existence of various wireless access network deployed such as, the basic WLAN IEEE802.11 family [1], UMTS, HSPA and the LTE [2]. The variety in radio access creates and constitutes the elements of the heterogeneous environment. This heterogeneous system [81] has different characteristics such as, network availability, coverage and Quality of Service (QoS). The main differences between these technologies is related to the speed of movement of the mobile node. One of the fundamental ideas of the heterogeneous systems is to give the users the ability to choose the best network among multiple existing Radio Access Technologies (RATs) based on several criteria, this process is called the Network Selection (NS) [1] [2]. The network selection procedure consists in selecting the best available network among those accessible [3]. The best network differs from one user to another due to the various parameters / preferences involved such as cost, QoS, energy consumed and even mobility, [4] [5].
Mobility is the key feature of wireless networks that differentiates them from the wired networks. On the other hand, the movement of the mobile user is a major cause for the QoS deterioration since the user might change the access point or even the network. Ensuring the required QoS for the mobile terminal is a very important field of research [82] which is the scope of this chapter. In the following, we propose to study the mobility of users and the network selection within a heterogeneous network. In particular, when a mobile terminal comes out of the coverage area of an access point in which it was connected, this situation is directly related to the physical mobility of the mobile user which might generate an interruption of connection when performing a vertical handover. To avoid these probable interruptions, the operator must take steps in advance to minimize the number of vertical handover especially the unnecessary ones. These interruptions might generate packet losses, transmission delay and even increase energy consumption. So, to eliminate these loses, we propose that the operator predicts the paths that are susceptible to be taken by the user with a certain probability using Markov model order 2. The prediction of the paths, allows us to take proactive measures reducing the degradation of QoS due to the unnecessary vertical handover. Subsequently, we consider the heterogeneous network including (Wi-Fi, 3G +, H + and LTE).
The main objective of this chapter is the adaptation of the network selection process in the case of high mobility of the users. The idea is to predict the paths that will be taken by the user with certain probabilities. After that, the operator makes an off-line network selection process considering the zones that constitutes the predicted paths. The third step consists in finding the best configuration of RATs for each predicted path after making the off-line network selection. Finally, the operator chooses the configuration of RATs that has the largest mean value for all the paths. This RATs configuration is the most likely the best suited solution to satisfy the users QoS requirements for the entire call session.
The rest of this chapter is organized as follows Section 2 presents the related work whereas in Section 3, we present our proposed approach, we discuss the Markov model used to predict the paths and we present the network selection algorithm used. After that, we give the methodology which allows us to reduce the vertical handover occurrences in the context of mobility while maintaining the required QoS. In Section 4, we provide an example to explain how our proposal does work, we have also compared theoretically the performance of the proposed approach with some other works. We finally conclude by a conclusion with perspectives.

[1]http://standards.ieee.org/
[2]http://www.3gpp.org/

4.2 Related work

In this section, we present the related work in the literature that treat the mobility issue, the prediction of states, the network selection and the vertical handover. In [83], the authors make a comparison between intrinsic (prediction made by the mobile user) and the extrinsic (prediction made by the Access point) methods based on a Markov model to predict the states of a mobile user. This work concluded that the intrinsic approach is more accurate and produced a higher performance. In [84], authors have compared the Markov model with other predictors like LZ predictors. They produced results showing that the Markov model is more accurate and suitable. After that, they proved that a Markov model order 2 is the most beneficial amongst the Markov models order k with $k \in N$.

In [37] and [38] the authors have used the Simple Additive Method SAW to get a ranked list of networks, while in [34] the authors make a mix between game theory and SAW method. The main benefits of SAW are its simplicity and low complexity. On the other hand, it has two major drawbacks, first: a parameter can be outweighed by another one, second: the rank reversal phenomenon which represents a problem for the entire MADM approach. Furthermore in [37], the authors compare the performances of the vertical handover VHO using SAW and TOPSIS. The authors conclude that the TOPSIS method outperforms the SAW method. In general, we can say that the compensatory methods such as TOPSIS can avoid the problem where a parameter value can be overshadowed by another one. This is due to the usage of some trade-off between the criteria. This means that a poor value in one criterion is neglected due to a good value in the other criteria; this concept provides more credibility for whole NS process in comparison to the non-compensatory methods, which use the thresholds and do not permit any kind of trade-off between criteria. A comparative study between SAW and WPM in the context of vertical handover is done in [42]. The authors use the relative standard deviation as a metric of comparison and they obtained a conclusion that WPM is better than SAW. In [43], the authors use the WPM method in the context of heterogeneous systems. Their conclusion states that the WPM method is a more robust approach for dynamic decision making while it penalizes the attributes with poor quality to a greater extent. The authors of [85] have proven that Markov chain order 2 gives more prediction accuracy in the ranges of 70 % to 95 %. They also explained that using Markov model order k with $k > 2$ does not seem to bring an important improvement at the cost of a significant overhead in terms of computation and memory space for the learning and the processing of the mobility model.

The authors in [86] have made experiments on vertical handover between WLAN and GPRS. They mentioned that the vertical handover latency is a sum of detection time, configuration time and registration time. The latency values presented show that multiple vertical handovers can degrade the QoS, especially the unnecessary ones. In [87], the authors show that mobile IPV6 for Fast Handover is the suitable protocol to satisfy the QoS requirements. The obtained simulation results show that the mechanism reduces the vertical handover delay and packet loss better than ordinary PMIPv6 and FMIPv6. However, there is a slight trade-off in signalling overhead as the number of vertical handover per unit time is increased when compared to ordinary PMIPv6.

In the paper [88], the authors have proposed a network selection uplink scheduler algorithm and evaluated it with various mobility types. Two user profiles have been tested and it was shown that the end-to-end delay can be reduced if a parallel transmission profile is chosen. In this study the authors have used several 802.11n access points which doesn't represent a real heterogeneous system. The proposed scheme is based on two notions: the queue and the priority of the packets. It is known that the priority based systems are not suitable especially when using heterogeneous networks. Because, different RATs means different packets structures; thus, they cannot be treated with the same way. Consequently, the proposed method is interesting but it doesn't treat the real case of heterogeneous networks.

In [89], the authors have studied the usage of mobility-related factors for network selection in HWNs and provided a scheme to use them in the same way as other factors in an MADM based generic framework. The authors used the MADM method to rank the alternatives, the AHP method to get the weigh factors and then considered nine criteria : the average handover cost, monetary cost, bandwidth, power consumption, security level, bit error rate, burst error rate, jitter, traffic load and signal strength. Though the proposed work took in consideration multiple criteria, several of them are strongly correlated such as the bandwidth and the traffic load, the bit error rate is related to the burst error rate. So, there are some useless criteria. Regardless of the correlated parameters, considering nine parameters increases the complexity of this approach significantly which can delay the NS process. Hence, introducing further unnecessary communication delay. The second problem is, no criterion is representing the mobility. The mobility is a random concept which cannot be modelled with a static factor in a system. The authors conclude by saying that they demonstrated that mobility-related factors should be considered for network selection, and that their scheme achieved the goal of selecting the appropriate network for MTs with various mobility features. In our opinion, this is not necessarily true because of the random nature of the mobility.

Authors in [90] have proposed a dynamic management of vertical handover with advanced resources reservation. This feature is done by determining the future localisation of the mobile terminal. The authors have considered only the WLAN and the UMTS without presenting a real and clear concept to make the selection between the networks.

In this work, we applied all these ideas and combined them to get a novel approach for Network Selection suited for high mobility. Of course our approach is a probabilistic one and it is based on the prediction of the path that will be used by the user. It will consider all the potential paths and gives the best configuration of RATs to provide user with the best QoS.

4.3 Proposed scheme

This section represents the main part of this work. Indeed, the objective of this work is to provide the mobile user with the best QoS during the call session. The primary reason of QoS degradation is the motion of the users, the existing works try to grant users the requested QoS instantly without ensuring the continuity of this QoS during the call session. To attain our objective, we start with the prediction of path with some probabilities. We use these probabilities to estimate the geographical path (in term of cells) that will be taken by the mobile user. The second step is to make the network selection on the predicted paths to classify the RATs at each state of the predicted paths. The third step consists in collecting all the configurations (the initial and alternative) of RATs that represents all the paths predicted. The final step is using the mean value formula to decide which configuration to use for all the paths.
The proposed solution consists on 4 steps; the first 2 steps are the most important which are the path prediction and the network selection on the paths predicted. The second step finishes by providing a list of ranked RATs. The succession of the RATs having the same

ranking order on each zone constitutes a configuration. We have added two additional steps 3 and 4 to optimize the vertical handover occurrence in each configuration.

4.3.1 Path prediction

We are interested in the on-line predictors, which examine the user mobility history, extract the current context, and predict the next location. Once the next location is known, the history is updated and the next context is added to the history. Then, the predictor prepares to extract the next location and so on. Basically, two famous families of predictors exists: the order-k Markov predictors and the LZ-based family. In [84], the authors compare the two families and give the following conclusions:

- The LZ-based family has the same performance with the order-1 Markov model.
- The order-2 Markov process outshines the order-1 process.
- The order-2 Markov predictors worked as well as or better than the more complex compression-based predictors, and better than the higher order Markov predictors. Particularly, O(3) (or above) Markov predictors doesn't improve over O(2) Markov since it reduces the accuracy by failing to make predictions most of the time.

So, in the following, we explain the Markov process theoretically, we give the mathematical description and we present an example to understand the whole process and how we apply our approach in a real world scenario.
A Markov process is a stochastic process based on the Markov property, this means that the useful information for the prediction of the future is completely contained in the present state of the process and it doesn't depend on previous states. Hence, the system is called memory-less [91], Equation 1. It was proven in [84] and [85] that the order 2 Markov chains give more accurate predictions than simple Markov chains. Accordingly, in our work, the path prediction is done using the Markov model order 2.
The prediction models are used in several domains such as, the prediction of weather conditions, chess game, poker ...etc. In this work we are going to use this model to predict the movement of the mobile users. There are plenty of previous studies that used the Markov model to predict the path of users [90], [83]. However, in addition to path prediction, our methodology introduces three steps to provide the users with a successful and efficient selection of networks that enable them to be best served during the call session.

$$P = P(L_{next}|L_{current}) \tag{1}$$

The order-k Markov predictors assume that the next location can be predicted from the sequence of the k most recent symbols in the location history [84]. At each instant, a user has a unique location, i.e., is connected with a single base station. The location is formalized as a string representing the locations where the user was/is connected. The set of all possible locations is noted 'A', one location from 'A' is 'a. Each user has a set of the visited (precedent) locations, $L_n = L_1...L_n$. The stochastic Markov process of order k is:

$$\begin{aligned} P(L_i = a|L_1, ..., L_{i-1}) = P(L_i = a|L_{i-k}, ..., L_{i-1}) \\ \forall\, a \in A,\ i, k > n \end{aligned} \tag{2}$$

L_i, means the location at an instant i. The distribution of the possible next location can easily be learned on-line. It is assumed that the prediction agent is regularly notified by the movement of the mobile user, the efficiency of the prediction is correlated with the notification periods. L^m is the location history of the mobile m, the estimation formula of order-n Markov model is:

$$P(L_i = a|L_{i-n}, ..., L_{i-1}) = \frac{O(L^m{}_{i-n}, ..., L^m{}_{i-1}, a; L^m)}{O(L^m{}_{i-n}, ..., L^m{}_{i-1}; L^m)} \quad (3)$$

Where $O(a; B)$ is the number of times the substring a occurs in the string B. These values constitute the transition matrix.
The location history is a sequence of location observations, for example, the user's location recorded once every ten seconds where, the $a_i \neq a_{i+1}$ for all $0 < i < n$. The predictor uses the location's history L as a sequence of abstract symbols without interpreting them.
To explain how the Markov process is applied, we propose a simple example: We consider an environment associated with a location history L="ADAHEFCBABCFHGEDABCFHGDAB CEHGDADGEHFCBA", see Figure 4.1.
Assuming that the mobile user is in zone 'A' at time t_0, after previously being in zone 'D' (previous location). Next, we will explain how the Order-2 Markov process works, especially how to extract the probabilities of the paths.
Based on the formula 3, the part $L^m{}_{i-n}, ..., L^m{}_{i-1}$ of the formula is known for the Markov process, in this example at the first step, it is equal to 'DA', and the process will try to find the 'a' part (the next location).
Table 4.1 contains the entire process and shows how to get the paths and their probabilities.

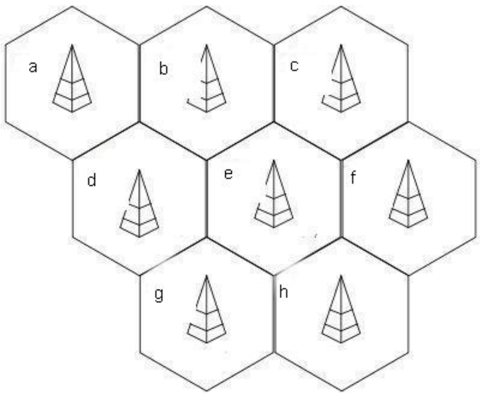

Figure 4.1: Example of a zone map

- c: is the concatenation of the current zone and the past zone, the order-2 Markov process. As its name indicates, this process needs 2 states to predict the next state.
- $O(c; L)$: is the number of times that the c occurs in the location history L.
- a: is the next state, the state that we want to predict.
- ca: is the concatenation between the values c and a.
- $P(ca)$: the probability of ca which is equal to $O(ca; L)/O(c; L)$.
- $P(Sub - path)$: the probability of the sub-path, the path is the succession of the zones predicted from the first round. The probability of the sub-path is equal to the

Round	c	O(c;L)	a	O(ca;L)	P(ca)	Sub-path	P(Sub-path)
1	αA	3	B	2	2/3	αAB	0.55
	αA	3	D	1	1/3	αAD	0.35
2	AB	3	C	3	1	αABC	0.55
	AD	2	G	2	1	αADG	0.35
3	BC	3	F	2	0.66	αABCF	0.396
	BC	3	E	1	0.33	αABCE	0.247
	DG	2	H	1	1/2	αADGH	0.192
	DG	2	E	1	1/2	αADGE	0.157
4	CF	2	H	2	0.9	αABCFH	0.396
	CE	1	H	1	0.9	αABCEH	0.247
	GH	1	E	1	0.9	αADGHE	0.192
	GE	1	H	1	0.9	αADGEH	0.157
5	FH	1	G	1	0.9	αABCFHG	0.396
	EH	1	G	1	0.9	αABCEHG	0.247
	HE	1	F	1	0.9	αADGHEF	0.192
	EH	1	F	1	0.9	αADGEHF	0.157

Table 4.1: The order-2 Markov process

product of the probabilities of the predicted zone with the probability of the older sub-path (the sub-path of the previous round). For example, in the third round the first row, the probability of the sub-path is equal to the product of the probability of the predicted zone ($P(ca) = 2/3$) and the probability of the sub path of the round two ($P(sub - path) = 0.396$).

The Markov process presented can be modelled by the scheme in Figure 4.2. This schematic representation of Table 4.1 shows the order-2 Markov transition from a state to another. It starts with a common state, and it discovers the next state using the location history. The values of the transitions are the probabilities of moving from the current state to the next state.
We remind that, we began from an arbitrary state and we use the location history L with the formula 3, we calculate the probabilities of the transitions. Finally, the probability of a path is the mathematical multiplication of the probabilities of the transitions in this path.

4.3.2 Network selection

In this step, the network selection part is presented; we discuss the parameters involved in the process and we explain the steps of the NS procedure. Subsequently, we propose an algorithm used to sort the available networks.
The context of this work is to maintain the user best served with the best available RAT in case of high mobility. To do so, it is essential to minimize the number of vertical handover i.e., eliminating the unnecessary vertical handover to reduce the delay and PLR caused by it.
A vertical handover is the process where the mobile user changes the radio transmitter, or access media used to provide the requested services, while maintaining a predefined QoS[3].

[3] http://www.3gpp.org/specifications/67-releases

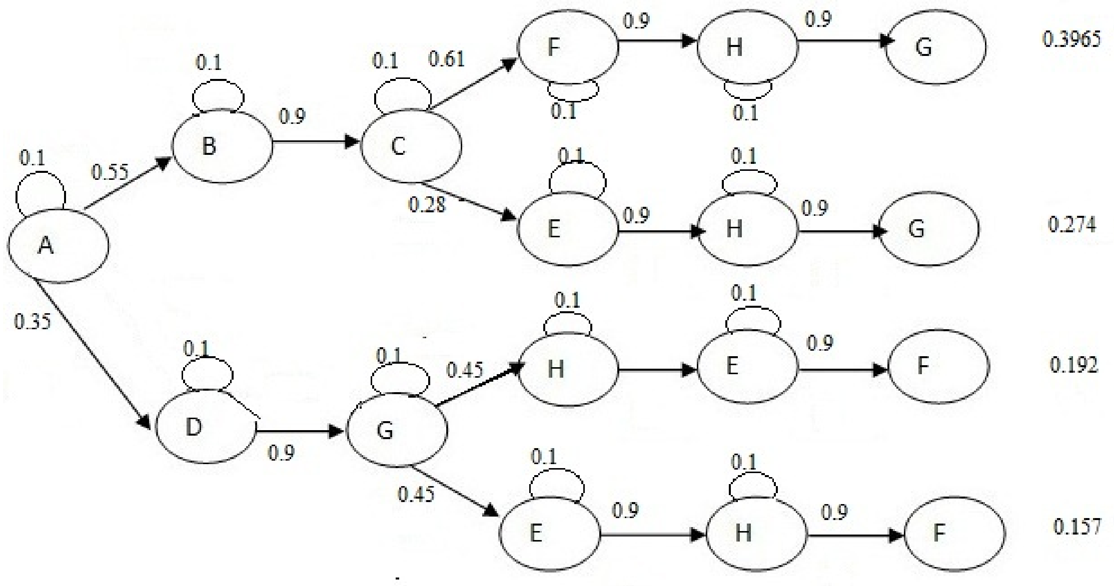

Figure 4.2: The Markov process scheme

A mobile user in heterogeneous radio environments is under the coverage of several different networks having different characteristics: access medium, link layer...etc. [92]. Therefore, the vertical handover decision coupled with the network selection process have become even more complex and crucial step towards achieving seamless mobility in the fulfilment of the vision of being "Always Best Connected" [93].
Making the vertical handover only for the sake of maintaining connectivity during mobility is not the only objective instead the vertical handover process is coupled with the network selection process to get the maximum benefit from all available networks. Unnecessary vertical handover actually amplifies the signalling overhead and delay. It also forces the mobile user to consume more energy to achieve a successful VHO [94]. To limit such negative effects on the QoS, we aim to minimize the number of vertical handovers.
For starters, our proposed algorithm sorts the networks and provides the user with the best network. The algorithm is essentially based on the following equations:

$$R_i = \sum_{j=1}^{m} score_{ij}. \quad (4)$$

$$score_{ij} = (\alpha - k_{ij}) \times w[j]. \quad (5)$$

$$k_{ij} = min\,(Vect_{ij}). \quad (6)$$

α: is a fixed real number equal to the number of alternatives.
k_{ij}. it represents the rank order of the network i for the criterion j.
$Vect_{ij}$: is the column vector from the matrix mat where j is fixed.
i: is the alternative and j is the criteria
$mat[n][m]$ is the input matrix. More details can be found in Algorithm 3.
The best case of one network 'i' is when it has the minimal value (we design the system to be a minimization problem) for the criterion 'j' this means the highest local ranking, thus $k_{ij} := 0$ and the revenue is: $R_{ij} = \alpha \times w[j]$.
The worst case is when the network has the worst local ranking for the criterion, i.e. $k_{ij} :=$

$\alpha - 1$, the revenue is $R_{ij} = w[j]$.
The use of the weight concept is taken from the MADM approach. This concept allows us to give more sense to our objective function which ranks the networks based on their values for each criterion. The weight vector brings out the significant criteria and distinguishes them from the other less important criteria depending on the application requirements. Therefore, it assigns higher incomes to the corresponding networks. This strategy has the following advantages:
- Avoids the situation in which a network that has a good value for a non important criterion will have the same revenue as another network with a good value for an important criterion. This situation exists in the SAW method for example.
- The weight concept is the representation of application requirements in the system and that's how we distinguish between applications because each one has specific requirements. VoIP (Voice over IP) requires a minimum time delay and PLR for the video application. In addition, VoIP service requires also a good throughput. For the best-effort applications, the existing conditions are accepted, but the cost criterion is very important. This information is translated into digital values with the eigenvector method.
In our study, we use many parameters such as: cost, energy consumption, average throughput achieved, average delay, average PLR and network load. An example of our matrix see Table 4.3. These parameters displayed in the corresponding matrix are the margin values that the simulations produced.

The following steps (3 and 4) are vertical handover optimization oriented. A vertical handover VHO is a period of time in which the user is not connected to any RAT; in this tiny period of time there is no connection leading to some dropped packets which is known packet loss. Furthermore, these packet loss means more delay to receive the data correctly. So, to eliminate these drawbacks, we added these two steps to try to get a configuration with an acceptable QoS without making many unnecessary vertical handovers Figure 4.3.
The objective of the steps three and four of the proposed architecture is to compare the performance of C_init and C_alt. The flowchart shows that in 2 from 3 cases , C_init is selected as the best configuration for this path, and in the remaining case, the C_alt is chosen.

Algorithm 3 Modified-SAW Function

Inputs: $mat[m][n]$, $vector\ w[], \alpha = m$
for $(j = 0; j < m; j++)$ **do**
  for $(i = 0; i < n; i++)$ **do**
    $tab[i] = mat[i][j]$;
  end
  $tabind = Sort(tab)$;
  for $(k = 0; k < Tabind.length; k++)$ **do**
    $score[tabind[k]] = score[tabind[k]] + (\alpha - tabind[k]) \times w[k]$
  end
end
Return $score[]$

4.3.3 RATs configurations

This section is the third step of the proposed framework where we use the results of the network selection step to generate the configurations. A configuration is a list of successive RATs where each RAT is the best one on its zone. For example, we have three zones named A, B and C with three available RATs on each zone R1,R2 and R3. Ideally, the best configuration of RATs in these zones is the succession of the best RATs on each zone. One configuration amongst others for these zones could be selecting the same RAT on all the zones, which will circumvent the vertical handover caused by the user mobility. However, this choice is not necessarily the optimal configuration of RATs in term of the global QoS for the user. The objective of this part is to collect all the possible RATs configurations. In the upcoming section we will discuss the process of selection between the collected configurations.

For each path that the user is susceptible to take, we make an off-line network selection to get the RATs configurations. The obvious one is the combination of the best RAT for each zone which constitutes a given path which is named the initial configuration. Moreover, we generate also the alternative configurations.

Conversely, selecting the initial configuration might introduce multiple vertical handovers. Essentially, our idea aims to generate alternative configurations that minimize or eliminate the vertical handover occurrences. Generally, there are two alternative configurations as we are able to change a RAT with another RAT which has the next highest score in RAT ranking of each zone.

When having the initial and alternative configurations, it is possible to compare the alternative configurations with the initial one based on the delay, Packet Loss Rate (PLR) and especially adding the jitter caused by the VHOs. We choose these parameters as they are the most essential ones for a best performance during the service session. The evaluation function depends also on the service application. Indeed, for the VoIP service, the delay is more important than packet loss. Same for the video service, these parameters have a similar importance.

In the following, a configuration is represented by C, so the initial configuration is named C_init and C_alt for the alternative configurations.

We remember that the vertical handover latency is the duration from handover initiation to handover completion which is equal to the amount of time from when the mobile is disconnected from the old RAT to the time when the mobile receives the first packet from the new RAT. The transmission delay and the handover delay are perceived by the mobile user, it is well known that the transmission delay is the amount of time taken by the packet to reach from source to destination. It is the sum of transmission delay, propagation delay and processing delay [95]. These delays are summed to give us the end to end delay (e2e). The VHO latency is modelled in our situation as jitter, indeed, when a VHO occurred, the user perceives an augmentation of the transmission delay which represents the jitter. The challenge is whether to select the initial configuration which have the suited values of delay and PLR while it is susceptible to generate jitter due to the possible VHO or to select an alternative configuration with higher delay and PLR but without any jitter because of the absence of any VHO. These different metrics have a common influence on the mobile user which is the deterioration of the QoS for the user so, we will try to connect the user with the configuration which generate the lesser degradation of QoS.

An alternative configuration must minimize the vertical handover occurrences, i,e., the number of vertical handovers in the alternative configuration must be lesser than the number of vertical handovers in the initial configuration. To select one configuration among others to

be used for one path, we use the formulas 7, 8 and 9. Whereas, if the formulas 7 and 8 are satisfied, we select the C_X otherwise, we select the C_Y.
Where:
$C\_X[i].delay$ and $C\_X[i].plr$: They represent the value of delay and PLR for the zone i of the alternative X.
$nb\_VHO\_(C\_X)$: This parameter means the number of vertical handovers in the configuration C_X. It is simply the number of RATs in which, the present and the previous RAT are different.
VHO_delay and VHO_plr: Are the latency and the PLR introduced when the vertical handover occurs. Both values are extracted from previous related work [86], [95], [96].

$$C\_alt[i].delay <= 150ms. \tag{7}$$

$$C\_init[i].delay \ + \ VHO\_delay > 150ms. \tag{8}$$

$$(C\_alt[i].delay - C\_alt[i+1].delay) <= |30|ms \tag{9}$$

These equation are applied for each zone. Where i represents one zone, 150 ms represents the maximum value of delay tolerated for the real time application. 30 ms is the maximum jitter tolerated and beyond it the QoS will be affected [86], [95], [96]. The previous equations focuses on delay only, because based on the PLR resulted from a VHO is in the range of 2%, so we have ignored the PLR in our process.
For each path among the predicted ones, these equations 7, 8 and 9 are used to select only one configuration and to eliminate the other configurations. At the end of this step, we get one configuration of RAT and the best configuration is considered. In the next section, we are going to propose a process that will take only one configuration to use it for the entire predicted paths.

4.3.4 Selection of the adequate configuration

The final step of the process is to choose one configuration to be used in the entire predicted paths. This configuration is probably the best suited for the whole session duration. To do that, the idea is to take all zones from Z and for each one of them, we select the appropriate RAT using the following algorithm. Z is the list which contains the zones susceptible to be visited by the mobile user.
For each path p_i, we have a configuration C_pi (see step three of the process).

Algorithm 4 Performance function

```
Inputs: p_i, C_p i Local variables: for (z in L) do
    for (i < n) do
        if (z in p_i) then
            p_i[z] ++;
        end
    end
end
return max (p[z])
```

This algorithm take as inputs the predicted paths and the selected configuration of RAT in each path, the output is a zone with a RAT. The process consists of selecting all zones which are susceptible to be visited by the mobile user and associate to those zones the RAT which have been selected by the maximum number of paths. In the following we make a recap of the proposed methodology and, give the general algorithm that summarizes the described process.

The internal process of Algorithm 5 is composed of four steps:

- The first step is the path prediction.

We use in this step the order-2 Markov process. The algorithm takes as inputs two parameters:

$Mob-position, Loc-history$ which are the mobile position and the location history of this mobile user. The result of this step is the combination of paths ch_i and their probabilities prb_i.

- The second step consists in making the network selection on the predicted paths.

In this step we make a network selection on each zone of each predicted path. We start by creating the matrix system $mat[][]$ for each zone in which, the rows are the RATs and the columns are the criteria like bandwidth, delay, PLR, energy consumption and cost. The result of this step is a matrix $Rk_i[][]$ that contains the ranking order of RATs for each zone on the path ch_i.

- The third step, we collect the configurations and we select only one configuration for each path.

First, the initial configuration that we generate is a trivial one because we just take the RATs with the first ranking order. The alternative configurations are a mix between the RATs with the first ranking order and those with the second ranking order in which, the number of vertical handover is minimized.

After collecting these configurations, the objective of the next step is to select only one configuration for each path and to eliminate the others using the equations 7 and 8 and 9.

- In the final step, we create a new configuration called the adequate configuration which contains on each zone the RAT which have been selected by the maximum number of paths. The following flowchart brings clarifies more the proposed network selection process.

4.4 Performance evaluation

We consider the same example presented earlier in Figure 4.1. Supposing that the mobile user is in the cell a, the first step is to use the O(2) Markov model function to get the possible paths that the user can take with their probabilities. Hypothetically, one of the paths predicted is $ch_i = "a- > b- > c- > e- > h"$ with a given probability. The next step is to make an off-line network selection on all the zones constituting the path ch_i. The first task of the step 2 in Algorithm 5 is to generate the input matrix (Table 4.3) based on the Table 4.2. The input matrix contains data that the RATs on each zone provide.

	Bwidth	Del(ms)	PLR(%)	Energy	Cost
Wi-Fi	1-11	100-150	0.2-8	-	1
3G	1-14	25-50	0.2-8	-	5
LTE	1-100	60-100	0.2-8	-	2

Table 4.2: The model matrix

Algorithm 5 Mobility-aware network selection

Inputs: $Mob - position, Loc - history$
Local variables: ch_i, prb_i, $mat[][]$, $Rk_i[][]$, $C\_init[]$, $C_alt[]$
Step 1: Path prediction
O(2)Markov Model ($Mob - position$, $Loc - history$)
Results :(ch_i, prb_i).
Step 2: Network selection

```
foreach ch_i do
    foreach zone on ch_i do
        Construct the input matrix mat[][]
        Gain Function (mat[][])
    end
    Results :( Rk_i[][] )
end
```

Step 3: RAT configuration and evaluation

```
foreach Rk_i[][] do
    cfg_init[] The initial configuration
    cfg_alt[] Two alternative configurations
end
if nb_VHO_(C_init[])==0 or nb_VHO_(C_alt[]) >= nb_VHO_(C_init[]) then
    Choose C_init[].
end
else
    for i ∈ C_alt ; C_alt[i] != C_init[i] do
        if formula 7 and formula 8 and formula 9 then
            Choose C_init[].
        end
    end
    Choose C_alt[].
end
```

Step 4: Selection of the configuration
Apply Algorithm 4.

The values presented in Table 4.3 are generated randomly based on Table 4.2, which contains range values for each criterion generated using NS3 simulator [74]. Other works have provided the same QoS values for these networks [97], [98].
Furthermore, when we execute the step 2 of Algorithm 5, we make the network selection. Algorithm 3 takes as an input the data matrix and a weight vector that represents the service application (voice, video or web navigation). The network selection results are presented in the Table 4.4.

4.4.1 Validation of the proposed scheme

So, before we verify if our selection process is more efficient than other existed ones, we must present an example to more explain the process and how it works. The C_init is the configu-

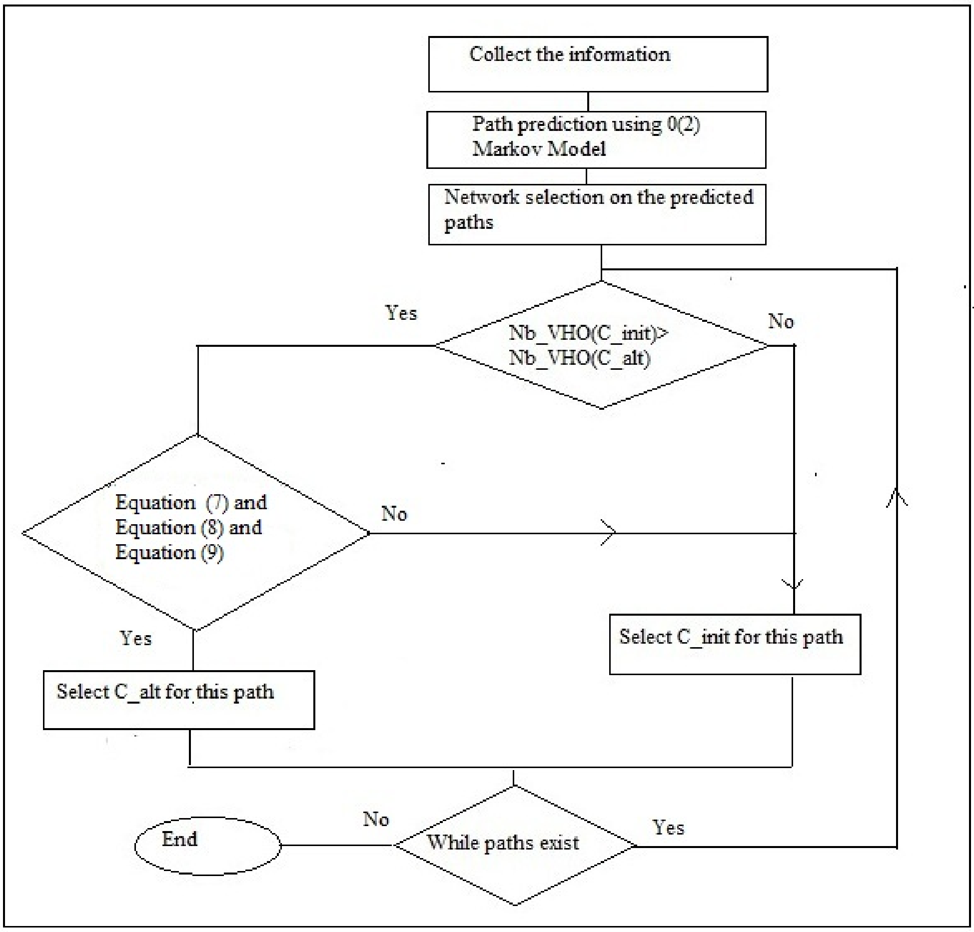

Figure 4.3: Proposed architecture

ration containing the best RAT for each zone, in our example, $C\_init = [R1, R2, R2, R1, R1]$. This configuration offers the best QoS for the user at each zone of the path ch_i. Nonetheless, this implies that the user has to make 2 vertical handovers which can caused the decline of the QoS of the entire path i.e., the entire session.
The alternative configurations are selected from the second row of the results Table 4.4. We have chosen the second row because generally the networks of this latter have an acceptable QoS performances in comparison with the ones in the third row.
So, based on C_init and the second row of the Table 4.4, we get two alternative configurations. The first one is $C1\_alt = [R1, R1, R1, R1, R1]$ and the second one is $C2\_alt = [R2, R2, R2, R2, R2]$. Remember that each C_alt must minimize the number of vertical handovers; Hence, we cannot accept any alternative configuration which has a number of vertical handovers greater than or equal to the number of vertical handovers in C_init.
Previous works revealed that making a vertical handover introduces additional delay of 86 ms and PLR of 3% of packets sent in the period of vertical handover [86]. The authors in [99] have provide a vertical handover latency in the range of 50 to 100 ms and packet loss rate range from 1% to 3%. So, we have decided to take an average value based on those references which means, a vertical handover latency of 50 ms and a packet loss rate equal to 2%. These value are used later in our process.
From Table 4.3, we can see and affirm that for zone A, R1 is more suited than R2 (lesser delay and Plr), so $C1\_alt$ is the alternative configuration which will be compared to the C_init.

Zones	RATs	Thr(Mb/s)	Delay(ms)	PLR (%)	Energy(J)	Cost
zone 'A'	R0	10.86	115.54	1.59	5.33	0.2
	R1	4.42	31.23	1.23	4.44	1.0
	R2	35.46	80.26	2.55	51.94	0.4
zone 'B'	R0	9.61	103.87	1.71	4.76	0.2
	R1	6.09	30.63	2.62	5.69	1.0
	R2	32.46	82.80	1.44	51.09	0.4
zone 'C'	R0	8.47	120.51	2.13	4.19	0.2
	R1	1.82	40.81	2.93	2.48	1.0
	R2	50	79.95	1.63	61.83	0.4
zone 'E'	R0	4.52	120.20	2.72	2.33	0.2
	R1	4.78	40.19	1.12	4.71	1.0
	R2	41	68.11	2.84	53.743	0.4
zone 'H'	R0	10.52	102.83	2.89	5.16	0.2
	R1	8.13	41.54	2.13	7.22	1.0
	R2	30.5	69.10	2.68	39.84	0.4

Table 4.3: Input matrix of the zones A, B, C, E, H

Thus, the selection of the best configuration in the third step (Algorithm 5) is based on reducing the number of vertical handovers while ensuring a good QoS for the user. Otherwise, the initial configuration is kept.
The comparison between the configurations is performed in term of the number of vertical handovers and the QoS offered to the users. When comparing the C_init and $C1\_alt$, we confirm that $C1\_alt$ has no vertical handover. Thus, it satisfies the first condition. Secondly, we see if the changes of RATs in the second and third positions of the C_init don't degrade the QoS or at least give similar performance values.
For our example, we have to compare C_init with C_alt. The network selection step has selected the C_init as the suited configuration for the user. The steps 3 and 4 consist on making some deep verification. In C_init we have two VHO, these 2 VHO affect the QoS of the user, because in this lapse of time (the VHO time) there is no connection, which means some packets will be lost and others will have more delays. Thus, the QoS in this lapse of time will be affected. So, we make this study to minimize the VHO occurrence especially when it is unnecessary. So, we have added the delays and PLR caused by the VHO with the transmission delay and PLR. In this example we have concluded that the C_alt is more suited but in other cases it might be not.
Figure 4.4 shows that $C1\_alt$ produces a lower delay by comparison with C_init or at least the same delay. In addition, we have two VHO in C_init which generates more latencies. We remark that C_init provides higher jitter more than the tolerated values, Figure 4.5.
If we select the network R2 in zone B, the user will have more a variation of delay and packet loss which is known as jitter. Thus, with this information we can say that choosing R1 instead of R2 is the right choice for the user. From this, C_alt is better than C_init. As a result, it will be wise for the user to not choose the initial configuration. Now, we must compare C_alt and $C1\_alt$.
The Figure 4.4 shows that R1 is better than R2 in terms of QoS in the zone 'B' and 'C' and have the same result on the other zones 'A', 'E' and 'H'. From this we conclude that the best configuration for the user is definitely C_alt because it eliminates the vertical handovers and do not degrade the QoS for the user in all the zones of the path because it presents a

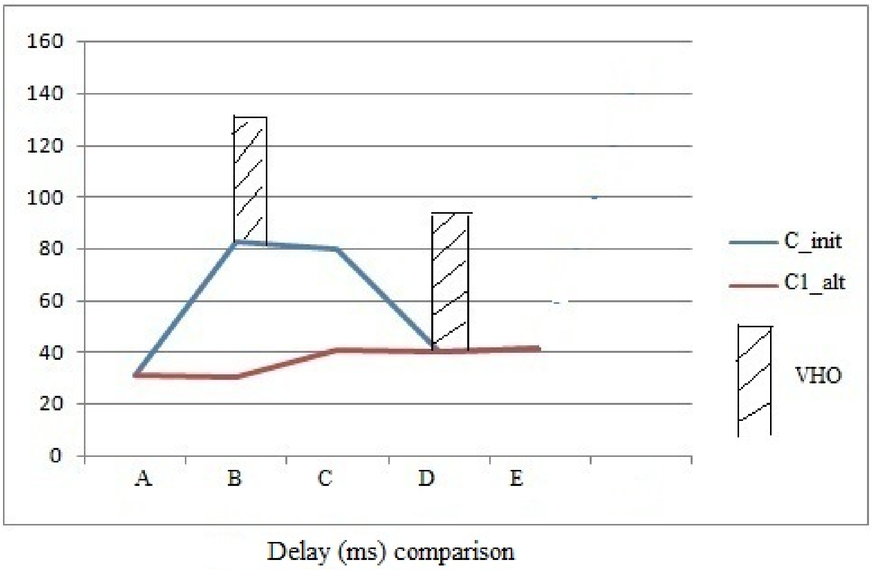

Figure 4.4: Instant delay comparison : C_init vs C_alt

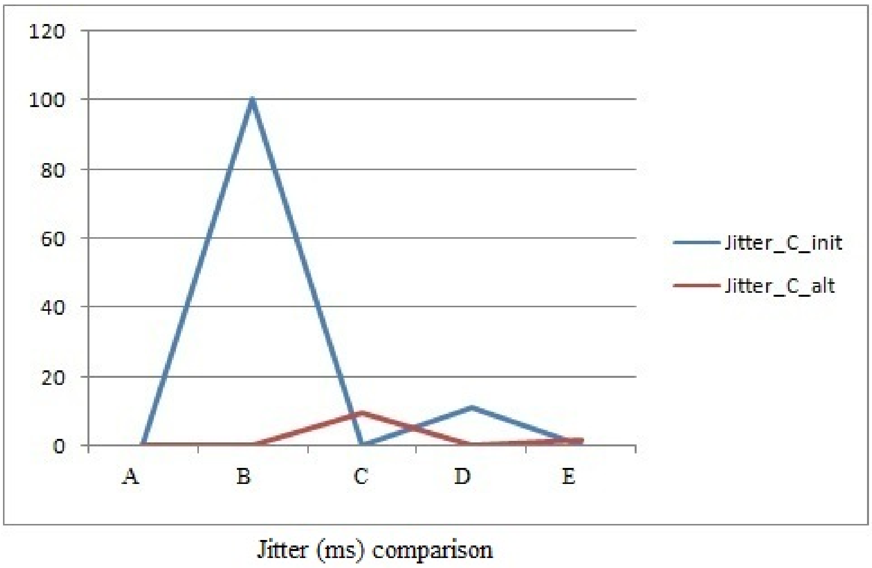

Figure 4.5: Jitter : C_init vs C_alt

low jitter.
The Figure 4.4 represents the instant delay for both configurations C_init and C_alt, the figure shows clearly that choosing the alternative configuration provides a stable and lesser delay without big variation in comparison with the initial configuration which provides a larger delays at each zone with higher jitter.
The same process must be performed for the other possible paths. Indeed, for each selected configuration, we apply the proposed process to choose the best configurations and using it for any chosen path.
The results of this example show clearly that our selection mechanism improves the performance. Indeed, for each path, we get the best configuration that minimizes the unnecessary vertical handover while we maintain a good QoS.

4.4.2 Comparison with other works

After we have present an example of the proposed framework, it is time to compare our proposal with the existed works. The results in Table 4.4 provide the ranking order of RATs

	A	B	C	E	H
Proposal	R1 R2 R0	R2 R1 R0	R2 R1 R0	R1 R2 R0	R1 R2 R0
TOPSIS [37]	R0 R2 R1	R0 R2 R1	R0 R2 R1	R0 R2 R1	R0 R2 R1
WPM [43]	R2 R0 R1	R0 R2 R1	R0 R2 R1	R0 R2 R1	R0 R2 R1
AHP [44]	R2 R0 R1	R0 R2 R1	R0 R1 R2	R0 R2 R1	R0 R2 R1

Table 4.4: Network selection results for the path ch_i

from zone 'A' to zone 'E' using different methods, in the following, we will compare these results.
For zone 'A', our proposal ranks the RATs like R1, R2, R0. The other methods give other ranking orders such as TOPSIS that ranks the RATs as R0, R2, R1; WPM and AHP give a ranking list equal to R2, R0, R1.
We will compare the QoS offered by R0, R1 and R2 for all zones to know which method allows the suitable selection of RATs.
From Figure 4.6, Figure 4.7 and Table4.4, we deduce that our proposal is more suited for

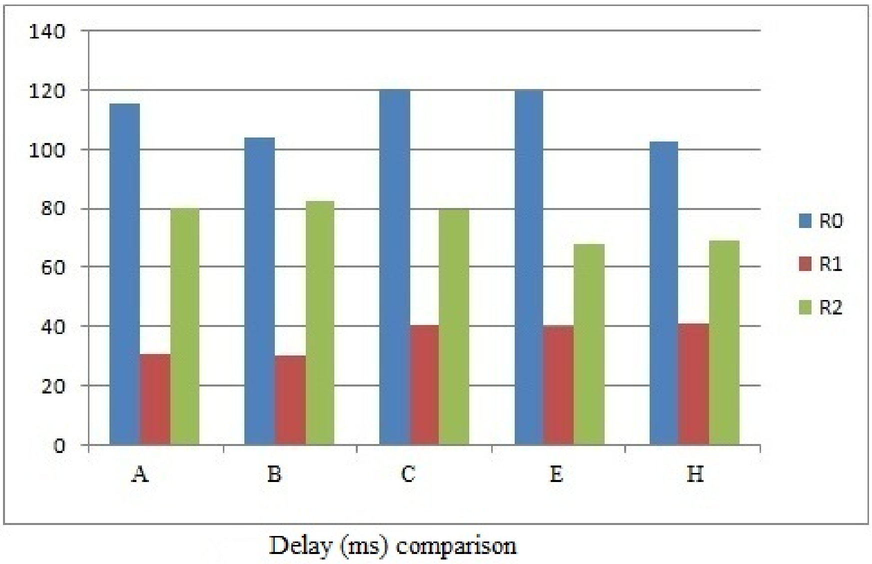

Figure 4.6: Delay offered by R0, R1, R2 on each zone

the users and gives more accurate ranking of RATs which provides a better QoS. In zone 'A' for example, our proposal selects R1 as the appropriate RAT. Whilst, TOPSIS chooses R0,

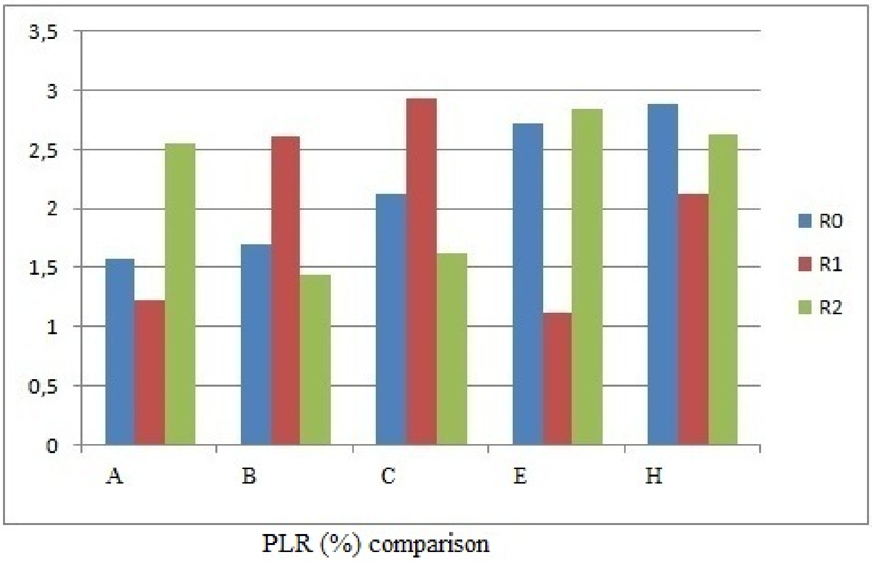

Figure 4.7: PLR offered by R0, R1, R2 on each zone

WPM and AHP select R2. Figure 4.6 and 4.7 show that R1 is more suited in zone 'A' and so on for the other zones.
From these figures, we conclude that our selection mechanism offers the accurate rank for the RATs. These methods such as TOPSIS, WMP and AHP have been used by several works like [37], [43], [44]. In the table 4.5, we present the advantages and inconvenient of some important works in this field.

In this section we have displayed the originality of our proposal with an example. In this example, any other QoS improvement oriented method should have chosen the initial configuration which invoke more unnecessary vertical handovers. However, our intelligent approach allows the user to have a high QoS while minimizing the vertical handover occurrences in the system.
As a comparison with other works, we haven't find works using the same approach. In the Table 4.5 , we give a comparison between our proposal and other works.

4.5 Summary

The objective of this study was to propose a network selection in the case of high mobility. The proposed approach begins with the path prediction using Markov model order 2. We used the latter because it was proven that this predictor provides the best performance for our case study. Subsequently, we apply a network selection across the zones of the paths predicted, we have proved that our selection mechanism performs very well and outperforms other previous proposals. The third step is the evaluation of the RATs configuration and finally, we use the mean value to get the configuration most adapted with all the predicted paths. This approach gives the user a lesser number of vertical handovers while it maintains a good or at least an acceptable QoS.
As a perspective, it will be interesting study the computation time of the whole process and try to enhance the prediction model to swap from a probabilistic model to a certitude model, also, we consider to realize a real implementation of the proposal.

Work	Advantages	Inconvenient
Paper[88]	- Reducing the end-delay	-Only 802.11n networks. -One queue with packet prioritization. -No consideration of handover.
Paper[89]	- Offering the required QoS.	- Many correlated parameters. - Only RSS to model the mobility. - No consideration of handover.
Paper[90]	- Prediction of mobility using Markov model. -Consideration of handover.	- No real network selection is presented - Only WLAN and UMTS.
-Prediction of mobility using Markov model. approach	- Many steps mean more time to get the optimal configuration of RATs. -Consideration of handover. -Reducing delay and PLR .	Our - Probabilist model.

Table 4.5: Theoretical comparison

General conclusion and perspectives

In this work we were interested in the analysis and the synthesis of network selection problem, it consists on the automatic choice of the suitable access network from the available RATs list; the decision process is called the Network Selection NS. The network selection process depends on several parameters such as QoS, mobility, cost, energy battery life . . . etc. Several methods and approaches have been proposed in this context, the objective was to satisfy the users with the best QoS and/or maximize the re-usability of the networks.

In the first chapter, we have presented generalities concerning mobile wireless networks. We began by giving definitions, then we detailed the networks concerned in this thesis, notably, WLAN networks very well known by Wi-Fi, we also talked about cellular networks and we focused specifically on the 3GPP family with all the Generations from the analogue era passing through GSM to the LTE-A. In the section 3, we bring out the difference between the traffic flows while giving the QoS criteria required by this type of flow. The last section was devoted to the handover process, which is a crucial element in mobile wireless networks.

The second chapter have consisted on a survey on the network selection methods used to solve this problem, so, in order to achieve the ABC concept, users with multi mode mobile equipment must be able to select a target network that meets their needs. This assumption implies a multi-technology environment. This chapter presents a review of the popular used method in this field; we first talked about the network selection problem and the utility functions, after that, we presented the most popular theories used basically the MADM methods, game theory and fuzzy logic. In this study, we managed to provide some critics to the discussed methods; regarding the MADM, we highlighted on the well known problem of rank reversal and the influence of one parameter over another in some cases, the major problems of the game theory can be the congestion situation in the case of the game between users, and the relatively high complexity computation. Furthermore, fuzzy logic needs another method to make decisions; it is generally coupled with the MADM or the genetic algorithms. Obviously, other approaches might be existed in the literature, but in reality, researchers focused on the exposed methods, which represent the core of the treatment, so the other approaches are used only in combination with the exposed methods. In this chapter, we give our critics of all mentioned methods seeking to make these solutions more robust and flexible.

The third chapter is the first contribution, indeed, in the aim to find the best network at each instant, the idea was to rank the existing networks and get the optimal ranking order, in this case, the operator switches the users with the best network available in the ranked list of networks. In this paper, we present our approach based on linear local weighed sum, the objective function is based on the relative ranking order of each alternative for each criterion at each round of the process, this basic idea allows us to get a greedy algorithm that gives good results. Results show that our proposed approach outperforms the existing used methods in the normal case, i.e. when all networks are available. Another test is done where a one of the networks disappears, simulation shows that all the MADM methods present the rank reversal phenomenon, our proposed algorithm overcomes this phenomenon and stays coherent and brings the same ranking order with eliminating the disappeared network.

The chapter four is dedicated to the second contribution in this work, it consists on the selection of the best network in the case of the user mobility. The objective of this study was to propose a network selection in the case of high mobility. The proposed approach begins with the path prediction using Markov model order 2, we used the latter because it was proven that this predictor is the most performer and beneficial, after that, we apply a network selection across the zones of the paths predicted, in the third step, we evaluate the RATs configuration and finally, we use the mean value to get the configuration most adapted with all the predicted paths. This approach gives the user lesser number of handover while it maintains a good and an acceptable QoS.

Our contribution are theoretical and proven with simulations, so one the first perspective is to apply these contribution in real systems to see the performance of our proposals in reality. A second perspective is to use the Quality of Experience QoE as criteria of network selection instead of the QoS. A last perspective would be to enhance the predicted model and swap from a probabilistic model to a certitude model.

4.6 List of publications

1- Survey On Scheduling And Radio Resources Allocation In Lte
F Bendaoud, M Abdennebi, F Didi
International Journal of Next-Generation Networks, 2014

2-A modified-SAW for network selection in heterogeneous wireless networks
F Bendaoud, F Didi, M Abdennebi
ECTI Transactions on Electrical Engineering, Electronics, and Communications, 2017

3- Mobility-aware network selection in heterogeneous wireless environment
F Bendaoud, M Abdennebi, F Didi
Journal of communications and networks (submitted)

4- Network Selection schemes in Heterogeneous Wireless Networks
F Bendaoud, M Abdennebi, F Didi
ICIPEE'14

5- Allocation des ressources radio en LTE
F Bendaoud, M Abdennebi, F Didi
International Conference on Telecommunication and Application, 2014

6- Network selection using game theory
F Bendaoud, M Abdennebi, F Didi
3rd International Conference on Control, Engineering, Information, Technology CEIT, 2015

7-Performances des algorithmes d'ordonnancement dans LTE Uplink
F Bendaoud, M Abdennebi, F Didi
Revue Méditerranéenne des Télécommunications 4 (2), 53-57, 2014

Bibliography

[1] Wang Yafang, Cui Huimin, and Zhang Jinyan. Notice of retraction network access selection algorithm based on the analytic hierarehy process and gray relation analysis. In *New Trends in Information Science and Service Science (NISS), 2010 4th International Conference on*, pages 503–506. IEEE, 2010.

[2] Farooq Bari and Victor Leung. Multi-attribute network selection by iterative topsis for heterogeneous wireless access. In *4th IEEE Consumer Communications and Networking Conference*, pages 808–812, 2007.

[3] Fayssal Bendaoud, Marwen Abdennebi, and Fedoua Didi. Network selection using game theory. In *Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on*, pages 1–6. IEEE, 2015.

[4] Mohamed Lahby, Silki Baghla, and Abderrahim Sekkaki. Survey and comparison of madm methods for network selection access in heterogeneous networks. In *New Technologies, Mobility and Security (NTMS), 2015 7th International Conference on*, pages 1–6. IEEE, 2015.

[5] Jianqing Fu, Jiyi Wu, Jianlin Zhang, Lingdi Ping, and Zhuo Li. A novel ahp and gra based handover decision mechanism in heterogeneous wireless networks. In *Information computing and applications*, pages 213–220. Springer, 2010.

[6] Josephina Antoniou and Andreas Pitsillides. 4g converged environment: Modeling network selection as a game. In *Mobile and Wireless Communications Summit, 2007. 16th IST*, pages 1–5. IEEE, 2007.

[7] Kun Zhu, Dusit Niyato, and Ping Wang. Network selection in heterogeneous wireless networks: Evolution with incomplete information. In *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, pages 1–6. IEEE, 2010.

[8] Matteo Cesana, Nicola Gatti, and Ilaria Malanchini. Game theoretic analysis of wireless access network selection: models, inefficiency bounds, and algorithms. In *Proceedings of the 3rd International Conference on Performance Evaluation Methodologies and Tools*, page 6. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008.

[9] Theodore S Rappaport. Introduction to wireless communication systems. *Wireless Communications: Principles and Practice*, pages 1–24, 1996.

[10] Theodore S Rappaport et al. *Wireless communications: principles and practice*, volume 2. Prentice Hall PTR New Jersey, 1996.

[11] Dharma P Agrawal and Qing-An Zeng. *Introduction to wireless and mobile systems.* Cengage Learning, 2015.

[12] Andrea Goldsmith. *Wireless communications.* Cambridge university press, 2005.

[13] Mischa Schwartz. *Mobile wireless communications.* Cambridge University Press, 2004.

[14] Todor Cooklev. *Wireless Communication Standards: A Study of IEEE 802.11, 802.15, 802.16.* IEEE Standards Association, 2004.

[15] Ieee standard. https://www.ieee.org/publications$_{s}tandards/publications/books/index_{b}ooks.html$, 20

[16] Marja-Leena Lahti. Ieee 802.11 wireless lan. In *Proceedings of the Seminar on Internetworking on Ad-Hoc Networking*, 2000.

[17] Timo Halonen, Javier Romero, and Juan Melero. *GSM, GPRS and EDGE performance: evolution towards 3G/UMTS.* John Wiley & Sons, 2004.

[18] Christophe Chevallier, Christopher Brunner, Andrea Garavaglia, Kevin P Murray, and Kenneth R Baker. *WCDMA (UMTS) deployment handbook: planning and optimization aspects.* John Wiley & Sons, 2006.

[19] Michel Mouly, Marie-Bernadette Pautet, and Thomas Foreword By-Haug. *The GSM system for mobile communications.* Telecom publishing, 1992.

[20] Gunnar Heine and Matt Horrer. *GSM networks: protocols, terminology, and implementation.* Artech House, Inc., 1999.

[21] Harri Holma and Antti Toskala. *HSDPA/HSUPA for UMTS: high speed radio access for mobile communications.* John Wiley & Sons, 2007.

[22] Marceau Coupechoux and Philippe Martins. Hspa+: les évolutions de hspa. In *Vers les systèmes radiomobiles de 4e génération*, pages 137–174. Springer, 2013.

[23] Stefania Sesia, Matthew Baker, and Issam Toufik. *LTE-the UMTS long term evolution: from theory to practice.* John Wiley & Sons, 2011.

[24] Erik Dahlman, Stefan Parkvall, Johan Skold, and Per Beming. *3G evolution: HSPA and LTE for mobile broadband.* Academic press, 2010.

[25] Dapeng Wu. Qos provisioning in wireless networks. *Wireless communications and mobile computing*, 5(8):957–969, 2005.

[26] Dapeng Wu and Rohit Negi. Utilizing multiuser diversity for efficient support of quality of service over a fading channel. In *Communications, 2003. ICC'03. IEEE International Conference on*, volume 3, pages 2202–2207. IEEE, 2003.

[27] Uyless Black. *Voice over IP.* Prentice-Hall, Inc., 1999.

[28] Carmen Egido. Video conferencing as a technology to support group work: a review of its failures. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, pages 13–24. ACM, 1988.

[29] Mariem Zekri, Badii Jouaber, and Djamal Zeghlache. A review on mobility management and vertical handover solutions over heterogeneous wireless networks. *Computer Communications*, 35(17):2055–2068, 2012.

[30] Henrik Persson and Johan M Karlsson. An approach to structure the handover terminology. *Proceedings of SNCNW'05*, 2005.

[31] Edson H Watanabe, Daniel S Menasché, Edmundo De Souza e Silva, and Rosa MM Leao. Modeling resource sharing dynamics of voip users over a wlan using a game-theoretic approach. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*. IEEE, 2008.

[32] Mohammed Alkhawlani and Aladdin Ayesh. Access network selection based on fuzzy logic and genetic algorithms. *Advances in Artificial Intelligence*, 8(1):1, 2008.

[33] Evangelos Triantaphyllou, B Shu, S Nieto Sanchez, and Tony Ray. Multi-criteria decision making: an operations research approach. *Encyclopedia of electrical and electronics engineering*, 15:175–186, 1998.

[34] Yass K Salih, Ong Hang See, Rabha W Ibrahim, Salman Yussof, and Azlan Iqbal. A user-centric game selection model based on user preferences for the selection of the best heterogeneous wireless network. *annals of telecommunications-annales des télécommunications*, 70(5-6):239–248, 2015.

[35] Aggeliki Sgora, Dimitrios D Vergados, and Periklis Chatzimisios. An access network selection algorithm for heterogeneous wireless environments. In *Computers and Communications (ISCC), 2010 IEEE Symposium on*, pages 890–892. IEEE, 2010.

[36] Lazim Abdullah and CW Rabiatul Adawiyah. Simple additive weighting methods of multi criteria decision making and applications: A decade review. *International Journal of Information Processing and Management*, 5(1):39, 2014.

[37] K Savitha and C Chandrasekar. Trusted network selection using saw and topsis algorithms for heterogeneous wireless networks. *arXiv preprint arXiv:1108.0141*, 2011.

[38] Quoc-Thinh Nguyen-Vuong, Yacine Ghamri-Doudane, and Nazim Agoulmine. On utility models for access network selection in wireless heterogeneous networks. In *Network Operations and Management Symposium, 2008. NOMS 2008. IEEE*, pages 144–151. IEEE, 2008.

[39] David L Olson. Comparison of weights in topsis models. *Mathematical and Computer Modelling*, 40(7):721–727, 2004.

[40] Bojan Bakmaz, Zoran Bojkovic, and Miodrag Bakmaz. Network selection algorithm for heterogeneous wireless environment. In *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–4. IEEE, 2007.

[41] Mohamed Lahby, Leghris Cherkaoui, and Abdellah Adib. An intelligent network selection strategy based on madm methods in heterogeneous networks. *arXiv preprint arXiv:1204.1383*, 2012.

[42] K Savitha and C Chandrasekar. Vertical handover decision schemes using saw and wpm for network selection in heterogeneous wireless networks. *arXiv preprint arXiv:1109.4490*, 2011.

[43] Peyman TalebiFard and Victor CM Leung. Context-aware mobility management in heterogeneous network environments. *JoWUA*, 2(2):19–32, 2011.

[44] Dong-ming Shen, H Tian, and L Sun. The qoe-oriented heterogeneous network selection based on fuzzy ahp methodology. In *Proc. of UBICOMM2010 (The Forth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies)*, pages 275–280, 2010.

[45] Chin-Tsai Lin and Pi-Fang Hsu. Adopting an analytic hierarchy process to select internet advertising networks. *Marketing Intelligence & Planning*, 21(3):183–191, 2003.

[46] Ying-Ming Wang and Ying Luo. On rank reversal in decision analysis. *Mathematical and Computer Modelling*, 49(5):1221–1229, 2009.

[47] Árpád Huszák and Sá Ndor Imre. Eliminating rank reversal phenomenon in gra-based network selection method. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–6. IEEE, 2010.

[48] Yong B Shin, Seungho Lee, Sun Gi Chun, and Dalsang Chung. A critical review of popular multi-criteria decision making methodologies. *Issues in Information Systems*, 14(1):358–365, 2013.

[49] Martin J Osborne and Ariel Rubinstein. *A course in game theory.* MIT press, 1994.

[50] Dimitris E Charilas and Athanasios D Panagopoulos. A survey on game theory applications in wireless networks. *Computer Networks*, 54(18):3421–3430, 2010.

[51] Tansu Alpcan, Tamer Basar, and Subhrakanti Dey. A power control game based on outage probabilities for multicell wireless data networks. *IEEE transactions on wireless communications*, 5(4):890–899, 2006.

[52] Shamik Sengupta, Santhanakrishnan Anand, Mainak Chatterjee, and Rajarathnam Chandramouli. Dynamic pricing for service provisioning and network selection in heterogeneous networks. *Physical Communication*, 2(1):138–150, 2009.

[53] Yan Zhang and Mohsen Guizani. *Game theory for wireless communications and networking.* CRC press, 2011.

[54] Kimaya Mittal, Elizabeth M Belding, and Subhash Suri. A game-theoretic analysis of wireless access point selection by mobile users. *Computer Communications*, 31(10):2049–2062, 2008.

[55] Amol Sahasrabudhe and Koushik Kar. Bandwidth allocation games under budget and access constraints. In *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*, pages 761–769. IEEE, 2008.

[56] Stavroula Vassaki, Athanasios D Panagopoulos, and Philip Constantinou. Bandwidth allocation in wireless access networks: Bankruptcy game vs cooperative game. In *2009 International Conference on Ultra Modern Telecommunications & Workshops*, pages 1–4. IEEE, 2009.

[57] Mohammod Sazid Zaman Khan, Shaiful Alam, and Mohammad Rezaul Huque Khan. A network selection mechanism for fourth generation communication networks. *Journal of Advances in Information Technology*, 1(4):189–196, 2010.

[58] Dimitris E Charilas, Ourania I Markaki, and Panagiotis T Vlacheas. Admission control as a non-cooperative multi-stage game between wireless networks. In *Systems, Signals and Image Processing, 2009. IWSSIP 2009. 16th International Conference on*, pages 1–5. IEEE, 2009.

[59] Olga Ormond, Gabriel-Miro Muntean, and John Murphy. Network selection strategy in heterogeneous wireless networks. In *Proc. Information Technology and Telecommunicaions Conference (IT&T)*, 2005.

[60] Dusit Niyato and Ekram Hossain. A cooperative game framework for bandwidth allocation in 4g heterogeneous wireless networks. In *2006 IEEE international conference on communications*, volume 9, pages 4357–4362. IEEE, 2006.

[61] Ramona Trestian, Olga Ormond, and Gabriel-Miro Muntean. Reputation-based network selection mechanism using game theory. *Physical Communication*, 4(3):156–171, 2011.

[62] Manzoor Ahmed Khan, Umar Toseef, Stefan Marx, and Carmelita Goerg. Game-theory based user centric network selection with media independent handover services and flow management. In *Communication Networks and Services Research Conference (CNSR), 2010 Eighth Annual*, pages 248–255. IEEE, 2010.

[63] Mainak Chatterjee, Haitao Lin, and Sajal K Das. Non-cooperative games for service differentiation in cdma systems. *Mobile Networks and Applications*, 10(6):939–946, 2005.

[64] Josephina Antoniou, Vicky Papadopoulou, Vasos Vassiliou, and Andreas Pitsillides. Cooperative user–network interactions in next generation communication networks. *Computer Networks*, 54(13):2239–2255, 2010.

[65] Lotfi A Zadeh. Fuzzy sets and information granularity. *Advances in fuzzy set theory and applications*, 11:3–18, 1979.

[66] Jelena Godjevac. *Idées nettes sur la logique floue*. PPUR presses polytechniques, 1999.

[67] Jerry M Mendel. Fuzzy logic systems for engineering: a tutorial. *Proceedings of the IEEE*, 83(3):345–377, 1995.

[68] Shubha Kher, Arun K Somani, and Rohit Gupta. Network selection using fuzzy logic. In *Broadband Networks, 2005. BroadNets 2005. 2nd International Conference on*, pages 876–885. IEEE, 2005.

[69] Sung-Bae Cho and Jin H Kim. Multiple network fusion using fuzzy logic. *Neural Networks, IEEE Transactions on*, 6(2):497–501, 1995.

[70] Cengiz Kahraman, Ufuk Cebeci, and Ziya Ulukan. Multi-criteria supplier selection using fuzzy ahp. *Logistics information management*, 16(6):382–394, 2003.

[71] Ying-Ming Wang and Taha MS Elhag. Fuzzy topsis method based on alpha level sets with an application to bridge risk assessment. *Expert systems with applications*, 31(2):309–319, 2006.

[72] Eng Hwee Ong and Jamil Y Khan. Dynamic access network selection with qos parameters estimation: A step closer to abc. In *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pages 2671–2676. IEEE, 2008.

[73] János Fülöp. Introduction to decision making methods. In *BDEI-3 Workshop, Washington*. Citeseer, 2005.

[74] Network simultor 3. https://www.nsnam.org/, 2011.

[75] G Piro. Lte-sim-the lte simulator, 2013.

[76] Teerawat Issariyakul and Ekram Hossain. *Introduction to network simulator NS2*. Springer Science & Business Media, 2011.

[77] Fayssal Bendaoud, Fedoua Didi, and Marwen Abdennebi. A modified-saw for network selection in heterogeneous wireless networks. *ECTI Transactions on Electrical Engineering, Electronics, and Communications*, 15(2):8–17, 2017.

[78] Junxian Huang, Feng Qian, Alexandre Gerber, Z Morley Mao, Subhabrata Sen, and Oliver Spatscheck. A close examination of performance and power characteristics of 4g lte networks. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 225–238. ACM, 2012.

[79] Christopher S Lewis and Steve Pickavance. *Selecting Mpls Vpn Services*. Cisco Press, 2006.

[80] Christina Hattingh and Tim Szigeti. End-to-end qos network design: Quality of service in lans. *WANs, and VPNs, CiscoPress, USA*, page 44, 2004.

[81] Krzysztof Gierlowski. Ubiquity of client access in heterogeneous access environment. *Journal of Telecommunications and Information Technology*, (3):3, 2014.

[82] Fatema Sabeen Shaikh. *Intelligent proactive handover and QoS management using TBVH in heterogeneous networks*. PhD thesis, Citeseer, 2010.

[83] Jean-Marc François and Guy Leduc. Prédiction de mobilité par le mobile ou par le point d'accès: comparaison sur base de traces réelles. In *Colloque Francophone sur l'Ingénierie des Protocoles-CFIP 2006*, pages 12–p. Hermès, 2006.

[84] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. Evaluating location predictors with extensive wi-fi mobility data. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 1414–1424. IEEE, 2004.

[85] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, page 3. ACM, 2012.

[86] Abdessadeq Fettouh, Najib El Kamoun, and Abdelaziz El Fazziki. Improving vertical handover performance of real time applications over heterogeneous wireless networks. *International Journal of Computer Applications*, 62(21), 2013.

[87] Linoh A Magagula, H Anthony Chan, and Olabisi E Falowo. Pmipv6-hc: Handover mechanism for reducing handover delay and packet loss in ngwn. In *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, pages 1–5. IEEE, 2010.

[88] Cristian Roman, Ruizhi Liao, Peter Ball, and Shumao Ou. Mobility and network selection in heterogeneous wireless networks: User approach and implementation. *Network Protocols and Algorithms*, 8(2):107–122, 2016.

[89] Lusheng Wang and David Binet. Madm-based network selection in heterogeneous wireless networks: A simulation study. In *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on*, pages 559–564. IEEE, 2009.

[90] Badr Benmammar and Francine Krief. Gestion dynamique du handover horizontal et vertical basée sur le profil de mobilité de l'utilisateur. In *Colloque GRES 2005: Gestion de REseaux et de Services*, 2005.

[91] Julian Keilson. *Markov chain models—rarity and exponentiality*, volume 28. Springer Science & Business Media, 2012.

[92] Qingyang Song and Abbas Jamalipour. An adaptive quality-of-service network selection mechanism for heterogeneous mobile networks. *Wireless Communications and Mobile Computing*, 5(6):697–708, 2005.

[93] Eva Gustafsson and Annika Jonsson. Always best connected. *IEEE Wireless communications*, 10(1):49–55, 2003.

[94] Hyun-Ho Choi and Dong-Ho Cho. Takeover: a new vertical handover concept for next-generation heterogeneous networks. In *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, volume 4, pages 2225–2229. IEEE, 2005.

[95] P Payaswini and DH Manjaiah. Simulation and performance analysis of vertical handoff between wifi and wimax using media independent handover services. *International Journal of Computer Applications*, 87(4), 2014.

[96] L Nithyanandan and I Parthiban. Vertical handoff in wlan-wimax-lte heterogeneous networks through gateway relocation. *International Journal of Wireless & Mobile Networks*, 4(4):203, 2012.

[97] Ilya Grigorik. *High Performance Browser Networking: What every web developer should know about networking and web performance.* " O'Reilly Media, Inc.", 2013.

[98] Maheen Hasib. *Analysis of packet loss probing in packet networks.* PhD thesis, Queen Mary, University of London, 2006.

[99] Toan Nguyen-Duc and Eiji Kamioka. An energy-efficient mobile-controlled vertical handover management for real time services. *Journal of Computer and Communications*, 4(17):59, 2016.