Seamless handover in heterogeneous wireless networks

Gabriele Tamea

Dottorato di Ricerca in Ingegneria dell'Informazione e della Comunicazione

Sapienza Universitá di Roma

A thesis submitted for the degree of

Doctor of Philosophy

Advisor: Prof. R. Cusani

Ai miei genitori che mi hanno sempre supportato ed a Sara per essermi accanto.

J'aime les nuages... les nuages qui passent... l-bas... les merveilleux nuages!

C.Baudelaire, Le Spleen de Paris, Lètranger

Acknowledgements

I would like to acknowledge my advisor prof. Roberto Cusani for its tutorship during the study and to all the people of the "INFOCOM" department for the time spent together (prof. Biagi, Valentina, Tatiana, Nicola, Stefano and Lorenzo).
I would like to thank prof. Francesco Delli Priscoli for its advice and counsellings during such years.
Many thanks to the friends of life (Lucandrea, Daniele, Valerio, Simone and Alessandro).
Thanks to the people of present day Labreti and ex-Lab (Anna, Antonio, Marco, Fabio, Cristina and Simona).

Abstract

The present thesis introduces decision criteria for the handover among heterogeneous wireless networks also known as *vertical handover*. The goal is to select the best network that can support the required service(s) and avoid excessive switching among different networks in order to minimize service interruptions and power consumption.
Firstly we address the vertical handover between specific broadcast technologies using a single aggregate function (SAF), then we generalized the approach in order to consider heterogeneous technologies and their possible optimization at link layer.
To avoid unnecessary handover in the SAF approach we introduced a probabilistic approach that has minimal computational complexity while maintaining a trade-off between received bits and number of vertical handovers.
In the last part of the present thesis we propose an approach based on a multi-criteria decision technique that permits to consider the out-of-service parameter and includes parameters that are affected by a degree of uncertainty.

Contents

List of Figures

Chapter 1

Introduction

Continuity of Service for users on the move has been always a target for different telecommunication industry actors, mainly network providers and device producers, since the first generation of radio technologies.
New multimedia services with strict quality of service requirements have emerged like video on-demand or high resolution television. At the same time different technologies was launched on the market (e.g. IEEE802.11, IEEE802.16, Bluethoot) each with its own pros and cons, e.g. high data rate but reduced range.
With a plethora of services and technologies, the main challenge is no longer to be "always connected" to a service but instead to be "always best connected" (ABC). ABC paradigm has only recently been introduced to indicate the possibility for a user to be always connected to the "best" network using the "best" device that can support the desired service(s). It should be noted that the notion of "best" is relative and, in some cases, subjective to the context and may includes different aspects like cost minimization and quality of service constrains successfully respected.
Among the procedures that permit the fulfillment of ABC is possible to identify the Mobility Management and, more in particular, the handover among heterogeneous networks (i.e. Vertical Handover, VHO).
VHO can be considered as a generalized handover that includes the possibility to switch between networks basing on different technologies.
One of the most widespread and studied example, due to the complementary

technologies characteristics, is between UMTS and WiFi network. The former can provide connectivity for user on the move while the latter permits to achieve generally higher bit-rate.
VHO, as the ordinary handover, can be divided into different phases indipendently by the considered technologies, namely:

- *Handover Information Gathering*, it is the first phase where the available networks are identified and their characteristics discovered. In VHO it can include an harmonization between different characteristics so the network can be comparable and any unbalancement is avoided;

- *Handover Decision*, the terminal and/or the network can decide if an handover is executed and, in VHO case, towards which networks. If the decision is taken by only one side (i.e. network or mobile terminal), the other may assist it providing additional information or signalling support;

- *Handover Execution*, it is the final phase of the handover process and may not be included if the terminal remains in the original network. In VHO specific additional issues (like authentication, authorization and accounting) may require special care.

Summarizing VHO process is generally considered a more difficult task than the handover between the same technology due to the networks heterogeneity, i.e. same parameters may have different impacts on quality of service (and also on quality of experience).
The main contribution of this thesis is to define new decision strategies for VHO considering both the network as well the user expectations in terms of network usage and quality of service. Specific aspects and techniques for VHO considered in the present thesis are: network point of view, single objective function, ping-pong effect and decision process under uncertainty.
Network point of view, it is essential to provide a formalized description of a converged network, i.e. a "multi-technology network", so the operator may decide which parameters should be optimized depending by strategy. We have proposed a network description using linear programming and have developed different strategies using content adaptation (i.e. the possibility to change codec and/or

bit-rate for a service flow) that can maximize the network efficiency, the quality of services provided to the users or the service allocation fairness.

Single Objective Function, it is the simplest form of decision criteria, anyway the modelling of the decision function is an open reaserch point. Especially, the modelling of specific technologies (Digital Video Broadcasting for Handheld, DVB-H, and Universal Mobile Terrestrial System, UMTS) is considered in the present thesis. Another open research point that was introduced is the impact of other procedures (like power control or adaptive modulation and coding) on VHO.

Ping pong effect, it is a widely know problem in VHO as well as in ordinary handover. It consists in the execution of the handover procedure with an high frequency. The contribution of the present thesis is to limit the unnecessary handover using a probabilistic approach. More specifically different closed form equations for the probability to have an unnecessary VHO have been proposed and evaluated through numerical simulations.

Decision process under uncertainty, it assumes that the networks parameters may be not exactly known. We have proposed an extension of a well known multi-criteria decision process called Total Order by Similarity to Preference Solution (TOPSIS) that can support parameters affected by uncertainty. Specific parameters of the proposed algorithm have been introduced to permit a trade-off between the out of service probability, i.e. the probability that a network is below the minimum concerning a specific parameter, and the expected value.

After this introduction, the thesis is structured as follows:

in Chapter 2 the general problem of Vertical Handover is introduced and the solutions available in literature or at standardization level are presented;

in Chapter 3 the network operator point of view including a technique for network optimization based on transcoding is analyzed;

in Chapters 4, 5 and 6 the user point of view is considered. Specifically, Chapter 4 introduces the decision process using a single objective function, Chapter 5 deals with the specific sub-problem of VHO called ping pong effect, i.e. an high number of handover, that can arise in specific circumstances. Chapter 6 presents a decision process under uncertainty.

A concluding chapter ends the main part of the present thesis summarizing the achieved results.

An annex explains some state of the art and details additional results.

Chapter 2

Vertical Handover: problem statement and possible solutions

2.1 Introduction

The present chapter gives a general introduction to the Vertical Handover (VHO) issues and solutions already available in literature focusing specifically on the decision process, in fact decision strategies are defined in the rest of this thesis.
It is important to remind that the “vertical handover” can be defined as an handover when more than one technology is involved (see Fig. 2.1 and [34] for a taxonomy of handover procedure).

More specifically in 2.2 Mobile IP and IEEE 802.21 standards are considered. The former has included for its stability and widely usage, the former for its current development and growing interest in the research community. Some aspects (protocols description) of the support to mobility at protocol level was also included in 2.2.
A section is integrally devoted to the description of strategies for the decision process. In fact the decision to execute or not the handover is the most important part of the handover process, the one that can impact performances to an extended degree, reducing the quality of experience or impacting on user terminal aspects (e.g. battery lifetime).

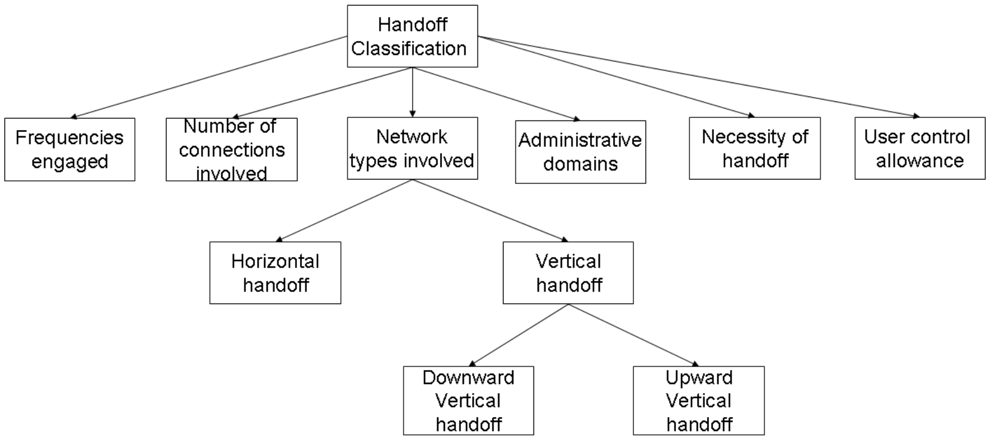

Figure 2.1: Handover Taxonomy

2.2 Standards

Standards permits to achieve interoperability and, in case of vertical handover, offer a common framework for interoperability. In the following the most widely used and accepted standards are described.

2.2.1 Media Independent handover Services

Media Independent handover services have been defined in the IEEE 802.21 standard with the scope to extend "... IEEE 802 media access independent mechanisms that enable the optimization of handover between heterogeneous IEEE 802 networks and facilitates handover between IEEE 802 networks and cellular networks. ". IEEE802.21 had defined the following elements:

- *a framework for service continuity.* In fact assuring service continuity during an handover is necessary from the user point of view. The framework based on media independent handover (MIH) reference models permits to handle heterogeneity across different link-layer technologies;

- *a set of functions within the network stack of the network elements* and a new entity *Media Independent Handover Function*;

Link event	Description
Link_Detected	Link of a new access network has been detected.
Link_Up	L2 connection is established and link is available for use. This event is a discrete event.
Link_Down	L2 connection is broken and link is not available for use. This event is a discrete event.
Link_Parameters_Report	Link parameters have crossed pre-specified thresholds.
Link_Going_Down	Predictive Link conditions are degrading and connection loss is imminent.
Link_Handover_Imminent	L2 handover is imminent based on changes in link conditions.
Link_Handover_Complete	L2 link handover to a new PoA has been completed.
Link_PDU_Transmit_Status	Indicate transmission status of a PDU.

- *a media independent handover service access point MIH_SAP* and the associated primitive. MIH SAP was used to provide services concerning: (i) detection of changes in link layer properties and, in case, execute the right trigger, (ii) a set of commands for the links that can be involved in handover, (iii) information about the different networks.

- *definition of new link-layer service access points SAPs and associated primitives for each link-layer technology.* It should be noted that IEEE802.21 is mainly oriented to provide support for other IEEE standards: IEEE802.3, 802.11, 802.16 are specifically described, but also 3GPP is included.

The link states that can trigger an handover are defined in the standard and reported in 2.2.1.

Any handover procedure that should support IEEE802.21 should be able to cope at least with the previous defined link event.
IEEE802.21 standard is structured with a lot of annexes, hereafter we had reported only the two that are important to understand some of the motivation of the present thesis:

- *QoS mapping.* In an heterogeneous scenario is necessary to map link layer parameters into common parameters. Moreover the standard defines how the information can be described using a statistical model, mainly based on mean, minimum and maximum values. Histogram can also be added;

- *Handover Procedures.* The annex includes the procedures executed in the cases of Mobile-initiated handover and Network-initiated handover. Additionally specific examples concerning the handover between 802.16 and 802.11 and the use of a proxy IPv6 Mobile IP are reported.

Concluding it is important to remark that the standard do not define handover control, policies, and other algorithms involved in handover decision making. Some of those aspects (mainly algorithms) was defined in the present thesis and can be integrated into a IEEE802.21 framework.

2.2.2 IP Mobility

In the present paragraphs we are interested in providing information about the existence and the basic mechanism of mobility support at IP layer using Mobile IP. For additional enhancement of Mobile IP (like IPv6 support (Mobile IPv6) , Fast Mobile IP or Hierarchical Mobile IP please refer to [15], [31] and [43]).
Mobile IP does not provide a solution to micro-mobility, i.e. like the one that was proposed in the present thesis, therefore Mobile IP was reported for completeness only.
In [38] protocol enhancements have been defined in the IP context to provide mobility at IP layer. The solution was based on the use of two different addresses: (i) an *home address* that identifies the mobile terminal not considering its current position and (ii) a *care of address* that represents the actual IP address of the mobile terminal. Assuming that an host is provided by a home address, Mobile IP can be divided into the following procedures for the management of care-of-address (CoA):

- *Agent Discovery*, the agent should determine if it is attached to the home network or to a foreign network. The procedure is accomplished using IP messages. More specifically, IGMP Router Discovery protocol is used;

- *CoA Registering*, Mobile Node sends a registration request to its Home Agent (in case through the Foreign Agent). The Home Agent creates an association to be able to determine the Foreign Network of the Mobile Terminal.

- *Tunneling to CoA* , Mobile Terminal uses its home network address. The corresponding node sends packets to the original network, such packets are tunneled to the foreign network using the foreign address.

Concluding it is important to note that Mobile IP is mostly suitable to provide nomadicity but can be applied also to cellular network when mobility is between different packet networks domains, each with its own address scope (public or private).

2.2.3 Session and Application Mobility

The most widely used protocol for mobility at application level is Session Initiation Protocol (SIP).
SIP support different mobility models including user and session mobility, the former correspond to the classical mobility while the latter includes a more complicate mobility approach than user mobility.
Concerning user mobility SIP with its functional elements permit to act as a proxy towards the user current location. Concerning session mobility we should note that it includes the possibility for a session to be moved between different terminals.
The two capabilities required for session mobility are (as described in [41]):

- *Device Discovery* - At all times, a user is aware of the devices that are available in his local area, along with their capabilities.

- *Session Mobility* - While in a session with a remote participant, the user may transfer any subset of the active media sessions to one or more devices.

SIP permits to manage mobility across different networks masquerading the complexity at underlining layer. Due to the fact that the protocol operates at application layer some triggers and a decision process is anyway needed for a complete

mobility support,moreover the time delay introduced by an approach at applicative layer can be greater than the one that can be supported by an application.
Another protocol that can support session mobility is Universal Plug and Play (UPnP) with Audio/Video extension. UPnP is currently being developed by the UPnP Forum and it is specifically targeted for local network: it can support only the devices inside a local area therefore its applicability to mobility is limited in some specific context, e.g. mobility inside the home network of an audio/video session between an hand-held device to a more comfortable video entertainment system.

2.3 Decision Process

Decision theory, i.e. the theory behind the decision process, is an interdisciplinary subject (comprising statistic, physiologic, political science) aimed at describing both how decisions should be made (*normative theory*) and how decisions are actually made (*descriptive theory*).
Decision theory can be applied in support of the decision process in the VHO and can involve any phase of the decision process that are, using a sequential model that is suitable for VHO like the one proposed by Brim [9]:

1. Identification of the problem, in case of VHO such phase can be realized off-line;

2. Obtaining necessary information, it is the determination of the network parameters including quality of service and contracts with the network operators;

3. Production of possible solutions, in VHO are the available networks that support at least a certain quality of service;

4. Evaluation of such solutions, the possible networks should be evaluated in terms of performances and/or stability;

5. Selection of a strategy for performance, the strategy may include received bits, out-of-services or more specific metrics;

6. Implementation of a decision, may include the attachment of the mobile terminal to the new network.

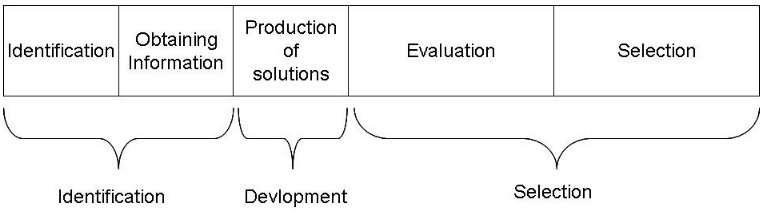

Figure 2.2: Bari's decision phases and relationship with general phases

The different phases are summarized in 2.2 where their relationship with usual phases is depicted.
The decision process in VHO can take into account different parameters, namely: *network*, *terminal*, *user* and *service related*.
Network related parameters include all the aspects related to network at all layers: coverage, bandwidth, latency, link quality (RSS, CIR, SIR, BER, etc.), monetary cost, security level, etc. Generally the user may or may not directly influence them depending by its context and capabilities.
Terminal-related are the aspects related to the physical terminal and may include static information (e.g. model and capabilities) or dynamic one (e.g. velocity, battery power, location information);
User-related parameters may be directly influenced by the user and include user profile and preferences;
Service-related depends by the service(s) that are currently provided or that are possible on a specific network including service capabilities, QoS, etc.

A good survey of the different decision strategies was given in [27] therefore in the following we summarized some of the necessary aspects detailed in the specific references.
The first, in temporal order, strategies defined for vertical handover dates back to 1999 and was *function based*. The function used was:

$$f_n = \sum_s \sum_i w_{s,i} p^{n_{s,i}} \tag{2.1}$$

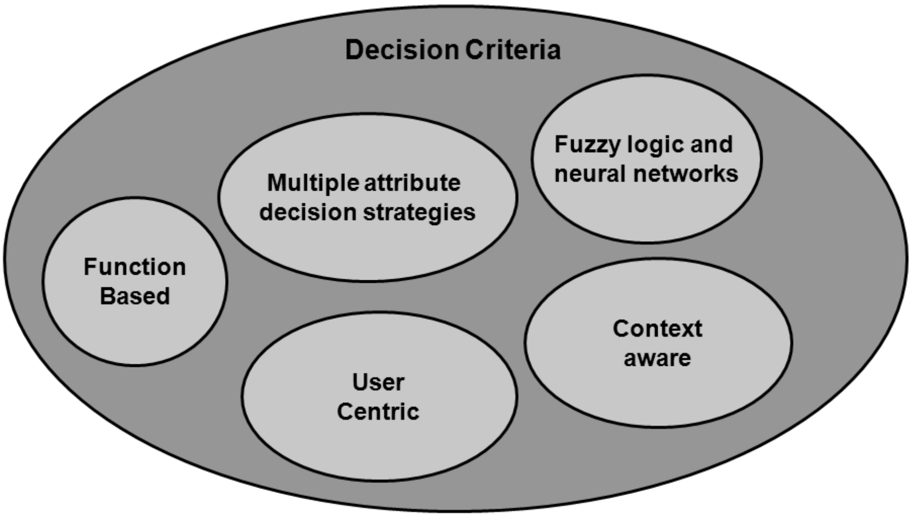

Figure 2.3: Decision Strategies for Vertical Handover

where $w_{s,i}$ is the weight assigned to the service with respect to the i-th parameter (with $\sum_i w_i = 1$) and $p^{n_{s,i}}$ is the cost in the $i-th$ parameter to carry out service s on network n. We can notice that f_n is a cost function that should be normalized, as an example if only specific parameters are consider it is possible to rewrite it as:

$$f_n = w_1 \ln\left(\frac{1}{B_n}\right) + w_2 \ln(P_n) + w_3 \ln(C_n) \tag{2.2}$$

where B_n is the bandwidth the network can offer,P_n the power consumption of using the network access device, and Cn the cost. The usage of a cost function is also considered in [10]

$$C = T_{WiFi}c_{WiFi} + T_{GPRS}c_{GPRS} \tag{2.3}$$

where T_i: the time spent by the user in the $i-th$ access network; $c_i(h)$: the fee per unit of time (second) that the operator of the $i-th$ access network charges to the user; C: the monetary cost faced by the user for a given communication session. The approach was implemented using a Mobile IP-like distributed mobility protocol to support the roaming of MNs in the Wi-Fi and GPRS domain.

Based on zone-based structure:

- satisfaction zone where the user is willing to pay to complete the service;
- tolerance zone with different utility function based on users risk attitude, generally a user can be described as risk adversal, risk neutral or with a risk attitude;
- frustration zone where the user is not willing to pay to complete the service.

Utility function based for specific service, e.g. file transfer.

Under the common name *Multiple Attribute Decision Making (MADM)* there are a set of technique that chose a solution from a set of solutions characterized in terms of their attributes. The techniques applied to VHO are: (i) *Simple Attribute Weighting Method*, (ii), *AHP - Analytic Hierarchy Process* and (iii) *Total Order by Preferences to Ideal Solution*.
In *Simple Attribute Weighting Method* the contributes from each attributes are added together after a normalization process, more formally the score of $i-th$ alternative is:

$$V_i = \sum_{j=1}^{n} w_j v_j(x_{i,j}) \tag{2.4}$$

The score with the highest value is the chosen one.
The *AHP - Analytic Hierarchy Process* [40] is based on comparison between different alternatives, more specifically the necessary steps are:

1. *build a hierarchy* including: the decision goal, the alternatives, and the criteria for evaluating the alternatives;
2. *priority assignment*, usually it is accomplished by making pairwise comparison among the elements of the hierarchy;
3. *judgments synthesize* to assign overall priorities to the hierarchy;
4. *consistency check* item *final giudment*

Total Order by Preferences to Ideal Solution was described in 7
Based on fuzzy logic and neural network, in [11] a decision process based on fuzzy logic is detailed. More specifically each parameter is described through a membership function μ .

It was calculated the membership function between two decision attributes A, B as a logical AND or OR. More specifically

$$
\begin{aligned}
\mu_{A\in B} &= \min(\mu_A, \mu_B) \\
\mu_{A\cup B} &= \max \mu_A, \mu_B)
\end{aligned}
\tag{2.5}
$$

triangular membership functions may be considered in order to use a multi-attribute decision making technique jointly with fuzzy logic.
Context aware are techniques that consider the "context" defined generally as all the information that may be useful for an handover decision therefore a context aware technique may employ any of the previous techniques.

2.4 Final Remarks

In this chapter we analyzed some of the aspects of the vertical handover problems focusing on decision theory and strategies.
The previous information are needed to better understand the algorithms that are described in the rest of this thesis.
Additional issue that are not include in this thesis and that may influence vertical handover performance are authorization and accounting mechanism and registration.

Chapter 3

Network operator point of view: a linear programming approach to network selection

3.1 Introduction

The present chapter analyzes the network operator point of view concerning vertical handover. In order to permit vertical handover it is necessary that the network operator deliver the same service through more than one networks that are under its own control, or that more network operators choose to deliver the same service. The former hypothesis is used in the present chapter (one operator with more than one network), more specifically one operator that has both a unicast as well as a unicast network is considered.
The general objective of the present chapter is to investigate into the mechanism that can be used to optimize a dual network, i.e. broadcast and unicast, when a single operator manages both and when it is possible to reduce the rate of the IP flows with transcoding techniques.
Broadcast technologies like analogue TV are considered due to the fact that they are widely used as the primary or, especially in remote areas, as the only source of information. They have the capabilities to reach a great number of users and use the spectrum efficiently at low cost for network operators. In recent years,

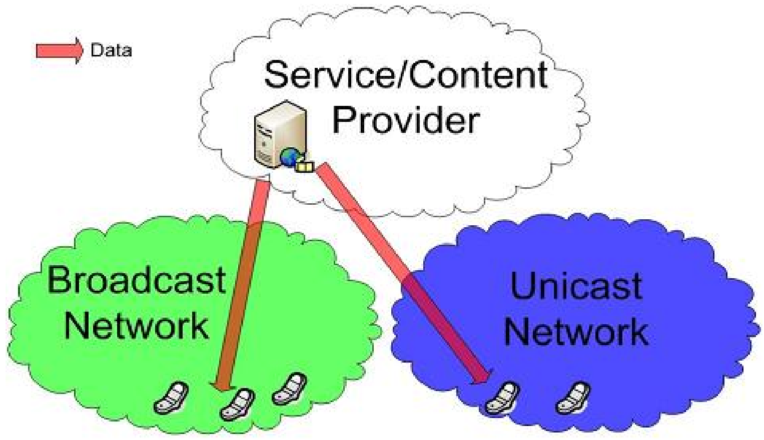

Figure 3.1: overall scenario

as the mobile terminals have grown in number and in features, digital broadcast technologies, like Digital Video Broadcast-Handheld (DVB-H, [17]), have been developed to provide digital video to mobile handheld devices [18] and also to provide the possibility to send IP-based content devices [16]. Interactive content delivery can be achieved through the integration with a unicast bidirectional network such as UMTS, providing the return channel for user generated signaling.
The integration of broadcast and unicast technology can be beneficial not only for enhancing a broadcast network with interactive capabilities but also to implement an optimal network resources utilization policy. Namely, the broadcast network can be used to convey the highly-demanded content to cover as wide audience as possible while unicast direct channels can be exploited to deliver programs with a lower audience. The number of contents delivered through broadcast channels can be optimized also with transcoding allowing to adjust the rate of single contents and achieve the best compromise between broadcast network capacity optimization, maximum number of served users and level of QoS perceived by the end users.
The overall scenario is depicted in 3.1 where a common service/content provider is highlighted along with the broadcast and unicast networks,

The chapter gives, in section 2.2., a description of the state of the art in digital broadcast and transcoding technologies The section 2.3 gives the problem formulation for a general network and describe two possible problem solving in a simplified case. Section 2.4 discusses efficiency and effectiveness of alternative approaches for content management and provides some simulation results. A concluding section (2.5) highlight the achieved results.

3.2 Transcoding and content adaptation

Media rate adaptation is used in cellular networks to solve terminal heterogeneity as well as access condition issues. Namely, it is possible that the media format which is adopted by the media source originally generated does not match the specific capabilities of a mobile terminal and, therefore, content adaptation is required to adjust the media format to the terminal capability. In addition, when users can connect to the network through various data links with different capacities and traffic conditions, the load of multimedia flows can be adjusted to the capacity of the network which is currently used. Transcoding can be achieved through the so-called content adaptation nodes receiving one or more media flows as inputs and delivering them scaled with different formats and rates. Layered video [22] format makes it possible to easily adjust media content to end-user terminal capabilities. For example, MPEG-2 standard [26] defines scalable modes to provide support in various transmission scenarios.

Platforms able to transcode multimedia flows have been designed by standardization organization like Open Mobile Alliance OMA [36] and [1]. The OMA defines a standard transcoding interface, i.e. an interface between Multimedia Application Platforms and a Transcoding Platform, such interface is generic enough to allow other applications such as browsing, media download services, push service, etc. In [2] is described an architecture of the environment using the standard transcoding interface. Conversely, within 3GPP, the IP Multimedia Subsystem defines a Media Resource Function (MRF) for media management. Namely, MRF provides media related functions, which are further divided into Media Resource Function Controller (MRFC) and Media Resource Function Processor (MRFP). The MRFC acts as a controller for the MRFP that is in charge of performing the transcoding operations.

It is worth highlighting that both 3GPP and OMA specifications do not include

algorithms for transcoding, while a number of them have been already designed and belongs to four main families: requantization, spatial resolution reduction (SRR), temporal resolution reduction (TRR) and combination of SRR and TRR. An optimized requantization technique that exploits only operations on the frequency domain is described in [4], it is focused on MPEG-2. Some algorithms for spatial resolution reduction are described in [42] e [12]. Temporal Resolution Reduction can be realized skipping the frame to maintain the desired bit rate, some temporal resolution reduction approaches are described in [20]. Combining the spatial and temporal resolution reduction is a relative new approach and is described in [29] .
In the next sections is described how the concept of trancoding can be used to optimize a converged network, i.e. an integrated broadcast and unicast network, using a dynamic content allocation platform, a similar approach is developed in [21] but focus only on a specific technology, i.e. DVB-T. The approaches that are presented in this article abstract the specific transcoding algorithm and transcoder device which are used. They are general enough to be applied to a variety of broadcast/unicast network integration scenarios.

3.3 Problem Formulation

3.3.1 General converged network

We assume that a service provider can deliver a set of possible contents to its subscribers using either a broadcast (e.g. DVB-H) or a unicast (e.g. UMTS) network. Requests of contents can be made thorough a return channel (e.g. still via UMTS). All subscribers are provided then with a multihomed terminal having a unicast and a broadcast network interface card. The same content can be requested by a great number of subscribers simultaneously and, hence, they are suitable for delivery through a broadcast network. Namely, delivery of contents over the broadcast network allows great saving of resources, i.e. typically bandwidth, with respect to delivery of contents via dedicated unicast flows. However, we assume that the broadcast network has a limited capacity and in general cannot be used to deliver all the requested contents. The network provider goal is to minimize its overall costs, i.e. minimize the sum of the cost of using the unicast and the broadcast channel:

$$\min C_{network} = min(C_{broadcast} + C_{unicast}) \tag{3.1}$$

Knowledge of the actual requested contents can help the service provider optimize content allocation to broadcast channel. Namely, the strategy of the service provider becomes conveying the contents that have the higher number of requests via the broadcast network while using the unicast network to provide "niches contents" through individual channels. The service provider can deliver different multimedia contents from various sources simultaneously, i.e. active sources, which can carry information like ordinary TV programs or files. Each source can adjust its rate from a minimum to a maximum value. The level of QoS for delivery of a content will be higher for higher rates. Conversely, lower rates for delivery of contents will allow a higher number of delivered contents. Hence, the service provider allocates broadcast channels of variable rate to contents consuming the available broadcast capacity. It operates, on the one hand, to serve as much subscribers as possible, on the other hand, to deliver contents with the highest possible rate.
Let the set of content sources which are active at the time t be $S(t) = S_1, ..S_{NS}$. The generic source S_i has a maximum delivery rate $R_{max,i}$ and a minimum delivery rate $R_{min,i}$. Hence, the rate of the i-th source is $R_i \in [R_{min,i}, R_{max,i}]$ and it is possible to define a rate range for each content source S_i as follows:

$$R_i = R_{max,i} - R_{min,i} \tag{3.2}$$

Without loss of generality, we can assume that each user listens exactly to one source as a user listening to multiple sources can be easily modeled as multiple users, each listening to a single source, in an equivalent model. Let us consider then the partition of all the user set in NS subsets, where the $i-th$ sub-set consists of all the users that listen to the $i-th$ source. Namely, the NS sub-sets are defined as:

$$D_1 = \{D_1^1, ..., \}D_{NS} = \{D_1^{NS}, ..., \} \tag{3.3}$$

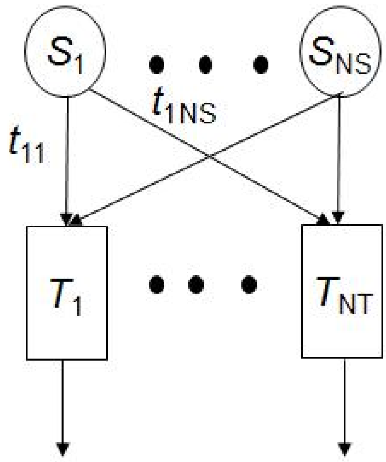

Figure 3.2: Relationship between sources (S) and transcoders (T)

Such sub-sets of 3.3 vary dynamically over time t depending on the user requests. In addition, from time to time, we also assume the a user can lose the connectivity with the broadcast network, as it often happens for mobile users. As a consequence, for each time t , the cardinality of each sub-set is:

$$|D_1(t)| = K_1(t), ..., |D_{NS}(t)| = K_{NS}(t) \tag{3.4}$$

For the sake of simplicity, we are omitting the dependency of such parameters from the time in the following of this work. Let us suppose that NT transcoders are available in the service provider network. We indicate with t_ij a binary variable that is 1 if the $i-th$ source is delivered to the $j-th$ transcoder for format adaptation and 0 otherwise.

We can assume that the $j-th$ transcoder can accept at most T_j input sources. The following constrains are to be satisfied: Where the first constrain express the fact that at most ''T_j' input sources can be accepted by a $j-th$ transcoder while the second constrain represent that a source is sent to only one transcoder.

$$\begin{cases} \sum_{i=1}^{NS} t_{ij} \leq T_j, \quad j = 1, ..., NT \\ \sum_{i=1}^{NS} t_{ij} = 1, \quad i = 1, ..., NS \\ t_{ij} \in \{0, 1\}, \quad i = 1, .., NS, \ j = 1, .., NT \end{cases} \tag{3.5}$$

3.3.2 Rate allocation strategies

Let us consider now a simplified case where we have exactly one broadcast link with a maximum capacity C, an ideal transcoder (one without constrains in terms of number of flows that can process simultaneously) and a set of NS sources. In such case, we can adopt two different approaches to optimize content delivery given the limited capacity C: 1. Minimum rate allocation first - we can try to first allocate broadcast channels to active sources covering the maximum number of requests with the minimum bandwidth; hence, bandwidth allocation for each broadcast channel can be increased up to filling all the available capacity; 2. Global maximization - we can try to allocate broadcast channels to active sources covering the maximum number of requests and maximizing the relevant bandwidth at the same time. The following sections details the two strategies.

3.3.2.1 Minimum rate allocation first (MRF)

This approach corresponds to find an optimal solution to the two optimization problems. Let us first consider the first problem, defined as follows:

$$\begin{cases} \max_{x_i} \sum_{i=1}^{NS} K_i x_i \\ \sum_{i=1}^{NS} R_{min,i} x_i \leq C \\ x_i \in \{0, 1\} \end{cases} \tag{3.6}$$

where the decision variable x_i is 1 if the $i-th$ source is allocated a broadcast channel and hence is transmitted using the broadcast network, 0 otherwise and hence is transmitted using the unicast network. Let us call $x*$ the optimal solution for the first problem and consider this second problem:

$$\begin{cases} \max_{y_i} \sum_{i=1}^{NS} K_i y_i \\ \sum_{i=1}^{NS} \Delta R_i y_i \leq C - \sum_{i=1}^{NS} R_{min,i} x_i^* \\ x_i = x_i^* & i = 1,..,NS \\ x_i \leq y_i & i = 1,...,NS \\ y_i \in 0,1 & i = 1,..,NS \end{cases} \tag{3.7}$$

where the variables $y_i, i = 1,,NS$ are real numbers in the interval $[0,1]$ and $R_i = y_i + R_{min,i}$ is the total rate finally allocated to the $i-th$ source.
With this approach it is possible to obtain first a fair channel allocation maximizing as many user requests as possible and afterwards it is possible to improve QoS perceived by all users exploiting the spare capacity.

3.3.2.2 Global maximization (global)

In this approach we combine the to problems of the MBF approach into a single problem to consider simultaneously the maximization of the number of served user through the broadcast channel and the maximization of the rate for the delivered contents. The problem can be formulated as follows:

$$\begin{cases} \max_{y_i} \sum\limits_{i=1}^{NS} K_i(w_1 x_i + w_2 y_i) & \\ \sum\limits_{i=1}^{NS} \Delta R_i y_i + R_{min,i} x_i \leq C & \\ w_1 + w_2 = 1 & \\ y_i \leq x_i & i = 1, ..., NS \\ x_i \in 0, 1 & i = 1, .., NS \\ y_i \in [0, 1] & i = 1, .., NS \\ w_1, w_2 \in [0, 1] & \end{cases} \tag{3.8}$$

where we have combined the two objective functions of the MBF using w_i variables.

3.3.3 Performance metrics

The allocation strategies are evaluated against the following performance metrics: efficiency, quality and fairness.
Efficiency (e) is the ratio between the users reached by the sources transmitted through the broadcast network and all the users. It represents how much the broadcast link is used to provide services to the users.
Quality (q) is an index expressing the degree of transcoding operated on the source. It indicates "how much" the sources are not transcoded. i.e. are provided to the users with the maximum quality.
Fairness (f) of the sources transmitted on the broadcast express how the sources transmitted through the broadcast channel are transcoded equally. summarizes

the different performance metrics.

$$e = \frac{\sum_{i=1}^{NS} K_i x_i}{\sum_{i=1}^{NS} K_i}, q = \frac{\sum_{i=1,x_i=1}^{NS} K_i y_i}{\sum_{i=1,x_i=1}^{NS} K_i} f = \frac{\sum_{i=1,x_i=1}^{NS} \sum_{j=1,x_j=1}^{NS} (K_i - K_j)^2 (y_i - y_j)^2}{\sum_{i=1}^{NS} K_i^2} \tag{3.9}$$

3.3.4 Simulations

3.3.4.1 Static Scenario

The two approaches for optimization of rate allocation, i.e. minimum bandwidth first and global optimization, have been assessed in terms of the metrics efficiency (eq. 8), quality (eq. 9) and fairness (eq. 10) and compared with a basic approach, i.e. static approach, in which transcoding is not used. Namely, in the static approach active sources are all transmitted at the maximum rate $R_{max,i}$ and the broadcast capacity is exploited trying to maximize the number of served users. We have used the matlab 7.0 simulator with the LP solve module to solve the linear programming problems in eq. (6)-(8). The first scenario considered was with a fixed amount of users, in such case the demand of contents was modeled with a Zipf's law [13], expressed by eq. 9:

$$f(k, s, N) = \frac{\frac{1}{k^s}}{\sum_{n=1}^{N} \frac{1}{n^s}} \tag{3.10}$$

It is important to note that under Zipf's law the vast majority of users are served by a relative small number of sources. Such behavior can be seen in where s^* is the s value that permits to serve at least the 90% of the users.

We have repeated the simulation for different values of the exponent s, i.e. between 0 and 10 using a step of 0.1, i.e. we have 100 values on the x-axis where we indicate s as the number of steps instead of the real value of s. When s equals 0, the Zip'f law is equivalent to a uniform distribution, denoting all contents requested with exactly the same popularity. Conversely, when s is greater than 0 the Zip'f law is strictly monotone decreasing and the first contents become more

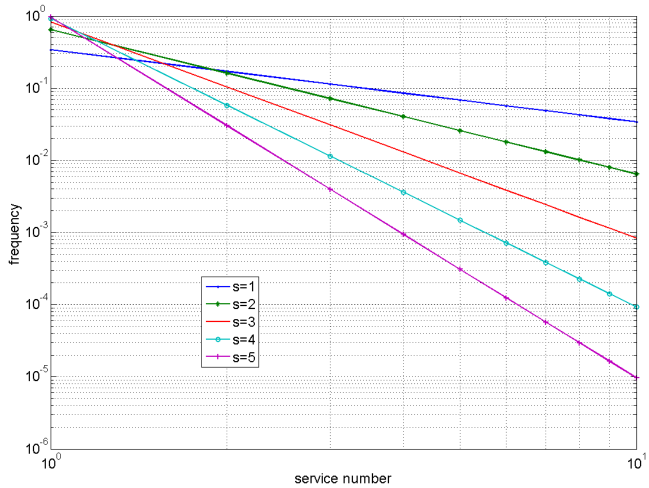

Figure 3.3: Zipf's distribution in log-log scale

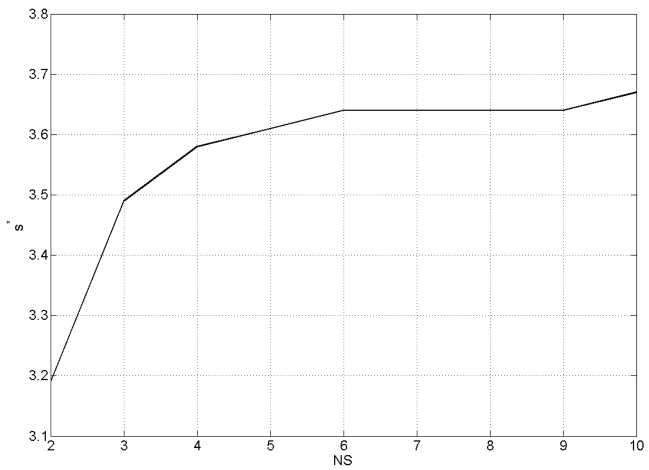

Figure 3.4: s^* value for more than 90% served user vs number of sources

and more popular with increasing values of the parameter s. Summarizing the parameters used for the simulation are:

- Capacity, a normalized capacity of 1 is chosen in order to avoid any normalization issue;

Parameter	Value
Capacity	1
Service Number	10
Maximum rate	1
Minimum rate	0.1
Service Popularity Distribution	Zipf
Global approach parameters	$1 \geq w1, w2 \geq 0$

- Service Number, a total of 10 services are considered;
- Maximum rate, equals to 1 to permit the possibility to have transcoding on the network
- Minimum rate, 0.1 to permit all the sources with the minimum rate value on the network;
- Service Popularity Distribution, a Zipf's distribution is chosen with the already described values ;
- Global approach parameters ranging between 0 and 1.

3.3.4.1 Summarizes the used parameters.

We can notice in 3.5 as the efficiency of the MRF approach grows linearly due to the growing unbalancement of service distribution. The solution of MRF in such case is the same for different values of s^* due to the fact that the MRF allocates the minimum rate to half the sources to fill the broadcast network capacity. Conversely, the static approach is the worst and tends to the maximum efficiency for increasing values of the s parameter, which means that only one type of content is requested. The global optimization is a compromise between the two approaches.

We can see an analogous behavior in 3.6 where the global approach has an intermediate performance between static and MRF. We can note that in 3.6 quality of static approach is always 1 due to the fact that no transcoding is operated. Quality of the MRF approach is equals to 0 due to the fact that maximum

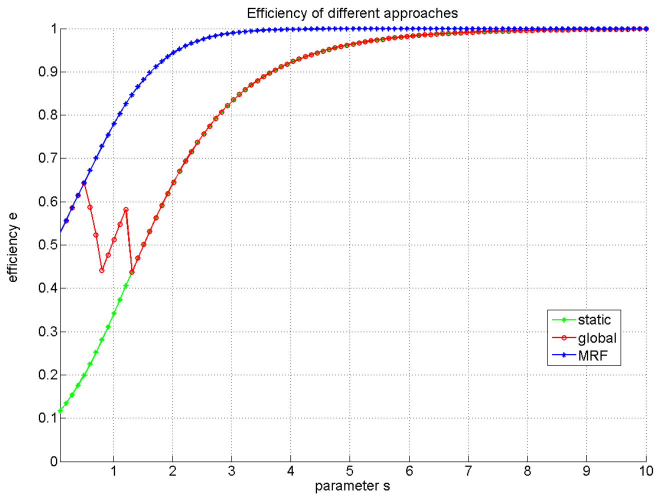

Figure 3.5: Efficiency

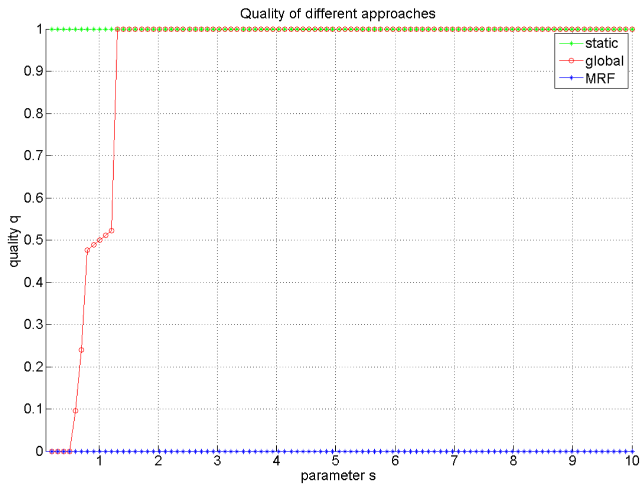

Figure 3.6: Quality

transcoding (i.e. minimum rate) is operated on all the admissible sources.

The fairness of the different approaches is depicted in 3.7. Global presents a worst fairness due to the fact that chose a trade-off between quality and efficiency not considering a “fair” solution in terms of transcoding.

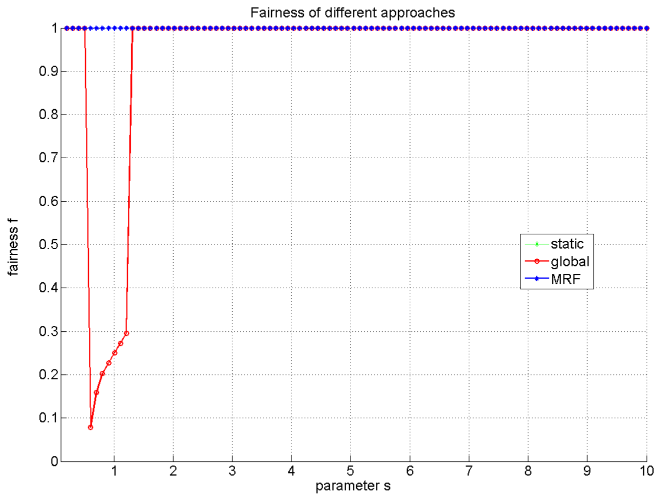

Figure 3.7: Fairness

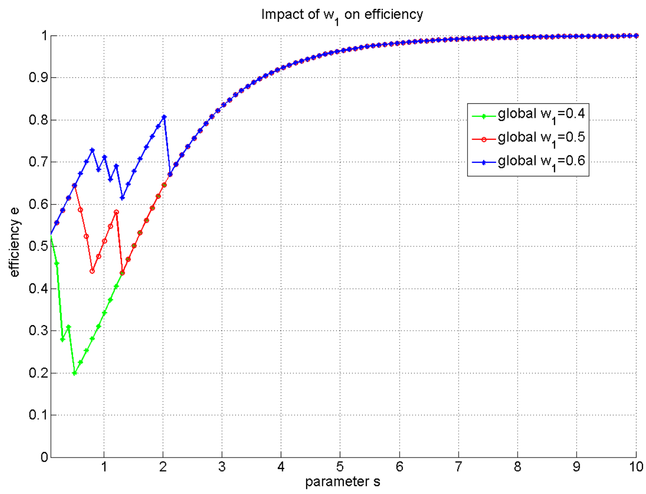

Figure 3.8: Impact of w_1

The impact of w_1 is shown in 3.8 where we can notice that the efficiency grows with the increasing of w_1. Such behavior is due to the fact that w_1 express the importance given to the efficiency in the linear target function of 3.8.

3.3.4.2 Dynamic Scenario

To simulate a dynamic behavior the following two states and NS+1 Markov's chain are used to represent user coverage and service request respectively (3.9 and 3.10).

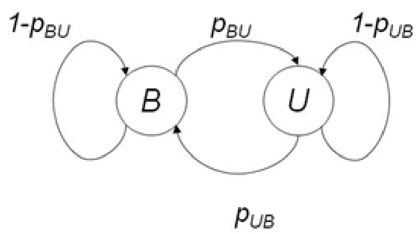

Figure 3.9: User coverage

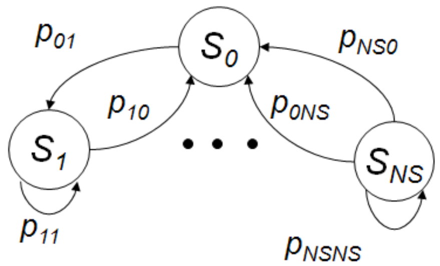

Figure 3.10: Service Request

We chosen to execute the approaches (i.e. static, dynamic or global) every T_{steps} seconds using the information acquired during the previous time slot. The results of the efficiency for different values of capacity are depicted in 3.11

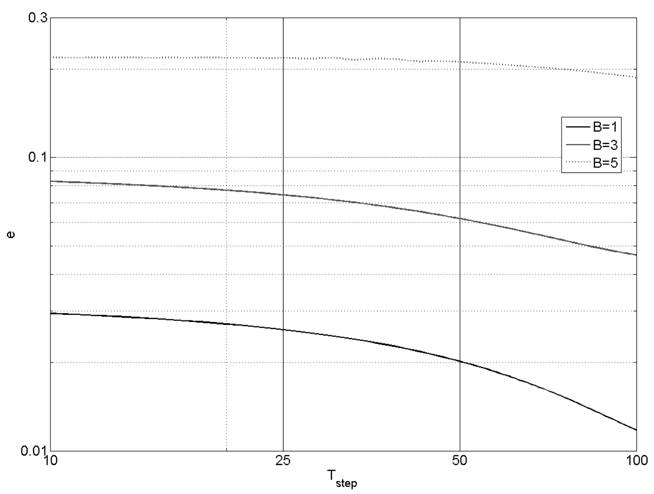

Figure 3.11: Dynamic

We can note that efficiency is very low due to the changing user behavior and connectivity.

3.4 Final Remarks

In this chapter we have introduced a model for resource allocation in a converged broadcast and unicast bidirectional network using transcoding. We have proposed two different approaches for the rate allocation policy: the former, i.e. MRF, which permits to achieve a high efficiency at the expense of quality while the latter, i.e. global, that permits to maximize the quality at the expense of the efficiency. As to the fairness the static and MRF approaches permits to have a maximum fairness while the global is a good compromise taking into account also the other performance metrics. Additionally we have evaluated the different approaches in a dynamic scenario, i.e. with users changing services and coverage. In such scenario it is possible to optimize the network if the correct sources request trend is forecasted.

Chapter 4

Single objective function

4.1 Introduction

Single objective function is the simplest decision strategy that can be adopted. In the present chapter, in the first part, we discuss its applicability to novel technologies, e.g Digital Video Broadcasting for Handheld (DVB-H) and Universal Mobile Terrestrial System (UMTS) with multicast and broadcast extension (UMTS-MBMS). In the second part we analyze the limits of the single objective function including the possibility of optimization of specific radio technology parameters.

4.2 Vertical handover between DVB-H and UMTS

4.2.1 Introduction

Nowadays, mobile terminals begin to be built with embedded multiple access technologies. In fact the same service can be provided using different access technologies as long as the terminal is capable of receiving and managing the content. If we consider services that need a great amount of bandwidth and that are targeted to a vast number of users simultaneously, it becomes favorable, for both the service provider and the users, to minimize the provision costs using a broadcast technology like DVB-H [17]. However DVB-H coverage is currently not so widespread in Europe as well as outside and might remain limited to metropolitan

areas. Consequently, at least for a preliminary phase, another back-up technology should be considered as a support to extend DVB-H network access. Furthermore, in order to render delivery of content via DVB-H interactive an additional network has to be combined with DVB-H to provide the return channel for delivery of control messages.
In the present chapter, we consider the Universal Mobile Telecommunication System (UMTS) as a valid back-up solution to compensate for DVB-H network lack of coverage and of interactive return channel.
Managing mobility across heterogeneous network is a challenging task. Namely, performance in different networks cannot be compared directly as they considerably differ in terms of bandwidth. A quality function balancing various aspects including bandwidth and packet loss in heterogeneous network is generally needed to allow direct comparison. In addition, a mechanism to handover between them has to be provided.
The aim of this chapter is to investigate into such mechanisms with a novel approach to evaluate the performance of a generic network able to compare different networks with different parameters. Based on such approach a quality function is introduced and two algorithms based on it are discussed: the former is a reactive algorithm while the latter is an extension of the former based on a proactive approach.

4.2.2 Technology Background and integration

4.2.2.1 DVB-H and UMTS

DVB-H is based on the Digital Video Broadcasting-Terrestrial (DVB-T) standard and it is aimed to provide digital video to mobile handheld devices. DVB-H and DVB-T flows can be transmitted on the same network, in fact they use the same transmission bands: VHF (174-230 Mhz) and UHF (470-838 Mhz). The major enhancement to DVB-T introduced in the DVB-H standard are: 4K mode, time-slicing and multi-protocol encapsulation-forward error correction. DVB-H includes also a way to transport IP packets in MPEG TSs (Transport Streams), i.e. IP Datacast architecture ([16]). Since IP has become a de-facto standard for information transport in this chapter we assume that the DVB-H networks uses

IP as network protocol and DVB-H as wireless link layer using multi protocol encapsulation (MPE).
UMTS is the most spread third generation mobile communication systems and its specification has been created by 3GPP [1].

4.2.2.2 Integration of DVB-H and UMTS

The services provided in a telecommunication network requires high level of interactivity and accessibility. Broadcast networks generally lacks interactivity but can provide broadband access to many user simultaneously. Conversely, the cellular technology is able to provide both high levels of interactivity and quality of service but exploit frequency bands less efficiently than broadcast by assigning resources individually to users. The combination of broadcast and unicast network to form a hybrid network could provide both broadband access to many users and individual interactive channels.
If DVB-H is considered as a broadcast technology and UMTS as a cellular one their integration can be obtained at the IP protocol architectural level by introducing a common gateway node that enables inter-working between the two networks [47].
Various scenarios where an integration of DVB-H and UMTS can be advantageously exploited are possible. In this chapter we are concentrating on the compensation for DVB-H network lack of coverage through a UMTS back-up network. Namely, we consider the scenario in which a user is moving out of the coverage area of DVB-H and exploits the UMTS network via the so called "Vertical Handover" process.

4.2.3 State of the art in Vertical handover between DVB-H and UMTS

Algorithms that specifically consider DVB-H and UMTS are [35] and [28]. In [35] a bicasting technique is adopted. The main idea is that a single IP datagram stream is encapsulated and sent to the DVB-H networks well as to the UMTS network. Such approach does not need frequency scanning and synchronization time to tune the signal of the new cell.

In [28] a paging message scheme for switching the network interface between UMTS and DVB-H is proposed. Such approach is energy efficient because the terminal can go to an idle mode when it is not receiving any data.
The above-mentioned approaches suffer from one of the two drawbacks:

- they fails in modeling only parameters specific of a wireless network like received signal strength, without considering user preferences, and service offer, which might be different form network to network.

- they use a non-standard approach to the network integration ([35] requires modifications in the user terminal to be able to correlate both flows and [28] requires a new paging message).

The use of the proactive approach defined in the IEEE802.21 standard combined with an algorithm of the first type is a very new research topics, an example can be found in [49] where the authors predict the next k-values of the Received Signal Strength Indication (RSSI) but they do not take into consideration all the possible parameters necessary for an handover.
In this chapter we propose a general quality function specialized for the DVB-H and UMTS and propose a mobility management approach with a quality function measuring network performance against some key performance aspects, including QoS, power saving and user preferences. We also extend a basic reactive vertical handover approach with a proactive one, using polynomial interpolation to optimize the performances in terms of the defined quality function.

4.2.4 Theoretical Basis

In the present thesis we contribute with a theoretical basis for optimization of mobility performance of UMTS/DVB-H dual-mode terminals. Namely,

- a general quality function for measurement of network performance in dual-mode UMTS/DVB-H terminals is first defined. In our approach a linear quality function modeling power saving, data loss and user preferences is considered.

- hence, we investigate into two algorithmic approaches for vertical handover between UMTS and DVB-H aiming at optimizing such function, for on-line implementation, with reasonable computational overhead.

- mobility performance are assessed against a differential quality function metric. This approach is general to be applied to a vast category of dual-mode terminals. In the following we are considering integration of DVB-H and UMTS in particular.

4.2.5 Quality function

Let k be the discrete time variable, with $k = 0, ., \infty$. Let us define for the DVB-H and UMTS networks the network quality function $Q_{NET}(k)$ (where the subscript NET is equal to DVB-H when we refer to the DVB-H network and is equal to UMTS when we refer the UMTS network) as a sequence of values in [0,1] for each $k = 0, ., \infty$ as follows:

$$\begin{cases} Q_{NET}(k) = \sum_l w_l f_l[x_l(k)] \\ \sum_l w_l = 1 \\ x_l(k) \in D_l \end{cases} \tag{4.1}$$

where each of the function f_i is a function of k through $x_i(k)$ and defined over an admissible domain D_i and contributes to the quality function $Q_{NET}(k)$ with a weight w_i. Each of the $f_i(k)$ function expresses one of the possible contribution to the quality function $QNET(k)$. In our approach we consider three contributions $f_i(k)$ to $Q_{NET}(k)$, i.e. power saving (PS), bit error rate (BER)and user preferences(UP):

$$Q_{NET}(k) = w_{PS} f_{PS}(k) + w_{BER} f_{BER}(k) + w_{UP} f_{UP}(k) \tag{4.2}$$

which will be detailed in the next paragraphs. Let us introduce the difference quality function $\Delta Q_{NET}(k)$, defined as follows:

$$\Delta Q_{NET}(k) = Q_{DVB-H}(k) - Q_{UMTS}(k) \tag{4.3}$$

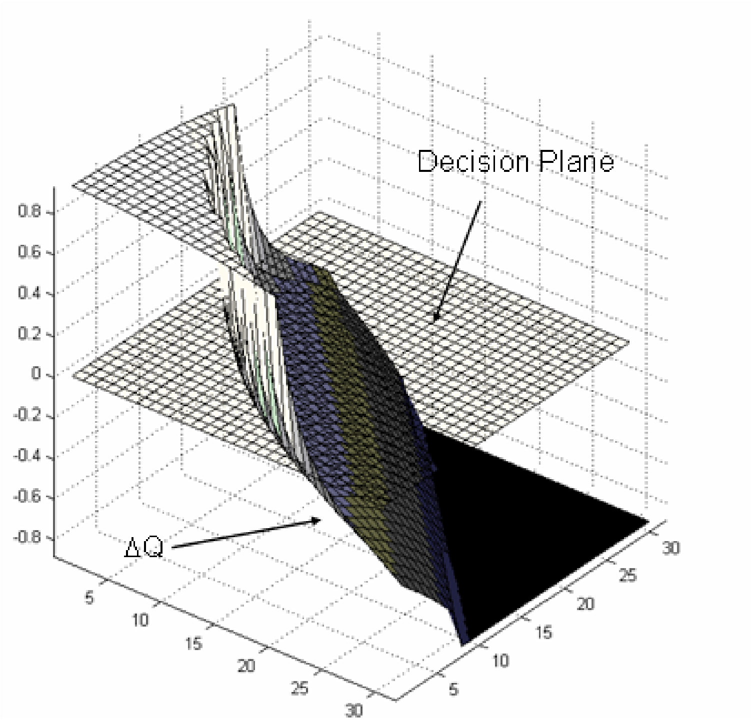

Figure 4.1: ΔQ over a sample scenario

Fig. 4.1 shows a simple scenario in which the values of the difference quality function are reported over a 30x30-zone grid representing a geographical area divided into square zones where DVB-H and UMTS network connectivity is provided simultaneously. Q is a surface dependent on the user position in the horizontal plane $Q = 0$, i.e. Q decision plane. When Q is negative (i.e. below the decision plane) the UMTS provides better performance than the DVB-H network, while when it positive (i.e. Q above the decision plane) DVB-H exhibits better performances.

4.2.5.1 Power Saving

In mobile portable devices it is essential to save as much power as possible in order to assure long life for the battery charge.
In DVB-H time slicing permits to reduce power consumption according to the

following equations:

$$B_d = \frac{B_s}{0.96 B_b} \tag{4.4}$$

$$O_t = \frac{B_s}{0.96 C_b} - B_d \tag{4.5}$$

$$PS_{DVB-H} = \left(1 - \frac{0.96(B_s + S_t)C_b}{B_s}\right) \tag{4.6}$$

where B_s is the burst size (bits), B_b the burst bit rate (bits per second), C_b the constant bit rate (bits per second), i.e. the average bit rate required by the elementary stream when no time slice is used, O_t the off time interval (seconds), i.e. the time between bursts and S_t is the synchronization time. The correction factor 0.96 compensates for the overhead caused by transport packet and section header. We choose f_{PS} equal to PS_{DVB-H}.

In UMTS, in order to save mobile terminal (MT) power the Discontinuous Reception (DRX) mechanism is adopted, i.e. if there is no packet transmission for a certain time the MT is turned off for a sleep period. The MT sleep period contains at least one DRX cycle after which the MT must wake up for a short period of time so that it can listen to the paging information from the network. In the UMTS extension, the Discontinuous reception transmission DRX is activated for all the mobile terminals receiving the same multicast flow. A formula to calculate power saving in UMTS is derived in [48]:

$$PS_{UMTS} = \lim_{T_off \to \inf} Pr[T_{off}] = \frac{P_3^{\inf} E[T_3^{HLD}] + P_3^{\inf} E[T_4^{HLD}]}{\sum_{i=1}^{4} P_i^{\inf} E[T_i^{HLD}]} \tag{4.7}$$

where T_{off} is the holding time in which the MT keeps a sleep mode, $E[TiHLD]$ is the expected holding time for the state S_i, where S_3 and S_4 are sleep periods related to the two different DRX cycles. Power saving probabilities typically range between 0.299 to 0.53 depending on the packet inter-arrival time. The $f_P S$ is set to PS_{UMTS} and ranges in the interval [0.299,0.53].

4.2.5.2 Quality of service

Quality of service is generally measured in terms of bit error rate, which is a function of the received signal-to-noise power ratio and the modulation scheme which is used.
In UMTS a BPSK modulation is used, while DVB-H can use QPSK, 16-QAM and 64-QAM. For an L-QAM modulation the following well-known formula applies:

$$BER = \frac{2}{log_2 L}\left(1 - \frac{1}{\sqrt{L}}\right) erfc\left(\sqrt{\frac{3E_b \log_2 L}{2N_0 L - 1}}\right) \tag{4.8}$$

Some critical values exists for the bit error rate depending on the transported application and network which is used. In fact if the bit error rate is below a certain threshold, i.e. target value , which depends on the service and on the considered technology, the network quality function is zero, i.e. the user cannot access the service with the given technology. For f_{BER} we can than consider the following linear function:

$$f_{BER} = \begin{cases} 0 & BER \geq \tau \\ 1 - \frac{BER}{\tau} & BER \leq \tau \end{cases} \tag{4.9}$$

4.2.5.3 User Preferences

User preferences allow a user to select a network on the basis of his subjective quality perception of the network, but also on the overall service offer which might differ significantly from network to network. The overall service/content offer can be different from network to network and better match different user tastes and profiles.
We can take into account user preferences in the considered case by introducing two parameters, i.e. u_{DVB-H} and u_{UMTS}, which represent two indicators of the matching of user tastes with service offered in the DVB-H or UMTS network. u_{DVB-H} and u_{UMTS} contribute to the DVB-H and UMTS network quality

function respectively, with the following properties:

$$
\begin{aligned}
&u_{DVB-H} + u_{UMTS} = 1 \\
&0 \le u_{DVB-H}, u_{UMTS} \le 1
\end{aligned}
\tag{4.10}
$$

then f_{UP} is equal to the user preference value u_{DVB-H} or u_{UMTS} when the terminal is in the DVB-H or UMTS network respectively.

4.2.5.4 Vertical Handover algorithms

Let us consider a user moving on a geographical area covered by DVB-H and UMTS simultaneously which is divided into zones as described in previous section. For each zone a value Q is defined. Every instant of time k the user changes a zone to an adjacent one a different $Q(k)$ is calculated.
When the reactive algorithm is used to govern handover decisions based on the observation of $Q(k)$, the network is simply changed at the instant k whenever $Q(k)$ function turns from negative to positive or vice versa . The reactive algorithm can be extended with a proactive approach in order to keep the level of $Q_{NET}(k)$ as high as possible during handover intervals. Namely, a proactive handover aims to predict Q and anticipate handover decisions in order to obtain performance gain. Therefore, samples of the quality functions $Q_{DVB-H}(k)$ and $Q_{UMTS}(k)$ are regularly calculated for every k and used to assess their quality trend and predict the function values on the next time intervals.
It is essential to keep the complexity of the prediction algorithm low while keeping satisfactory the forecasting accuracy, which depends on the number of samples that are considered for the prediction. The received power changes, in fact, depends on the mobile terminal position, so does the bit error rate and the packet loss.
Let us assume that a handover from DVB-H to UMTS is performed at time h, we define the handover quality function $Q_{HD}(k)$ as the $Q_{NET}(k)$ calculated as

$Q_{DVB-H}(k)$ for $k = 0, .., h-1$ and as $Q_{UMTS}(k)$ for $k = h, ., \infty$, that is

$$Q_{RH} = \begin{cases} Q_{DVB-H}(k) & k \leq h \\ Q_{UMTS}(k) & k \geq h \end{cases} \tag{4.11}$$

where the subscript HD is equal to PH when the proactive handover approach is used and is equal to RH when the reactive handover approach is used. The instant h is different for the proactive and reactive approach, and is generally smaller for the proactive approach. This is equivalent to saying that proactive handovers are generally anticipated with respect to reactive handovers.
We refer to the handover initiation time for the proactive and reactive approach with the term h_{PH} and h_{RH} respectively. We denote the duration of an handover with D, which is equal for both approach.

4.2.5.5 Differential quality function metric

The objective of the vertical handover algorithms is to optimize mobility performance based on the maximization of the quality function. Effective mobility performance gain should be than assessed through a suitable metric. Such metric can be derived by directly comparing the two handover quality function $Q_{PH}(k)$ and $Q_{RH}(k)$. It is worthwhile noticing that these two functions are identical for almost all samples, except those relevant to the handover intervals, that is:

$$Q_{PH}(k) \neq Q_{RH}(k) for k \in [h_{PH}, h_{PH} + D] \tag{4.12}$$

We can consider the relative increment RI of $Q_{RH}(k)$ in the interval $[h_{PH}, hRH + D]$ when the proactive extension is defined as follows:

$$RI = \frac{\sum_{k=h_{PH}}^{h_{RH}+D} Q_{PH}(k) - \sum_{k=h_{PH}}^{h_{RH}+D} Q_{RH}(k)}{\sum_{k=h_{PH}}^{h_{RH}+D} Q_{RH}(k)} \tag{4.13}$$

which is the difference between the quality function series of the two approaches divided by the quality function series of the reactive approach and it can be used to obtain a relative increment of performance with the proactive approach.

Performance gain of the proactive approach vs. the reactive one tends to be high when rapid variations of the quality function are experienced. We measure the degradation speed DS during a handover with the following quantity:

$$DS = \frac{\sum_{k=h_{PH}}^{h_{RH}+D} \Delta Q(k)}{h_{RH} + D - hPH} \tag{4.14}$$

which is the mean value of Q in the interval between proactive and reactive handover including handover duration, i.e. $[h_{PH}, h_{RH}+D]$. We introduce then the handover quality metric HQM as the ratio between RI and DS to obtain a metric to evaluate performance gain of the proactive approach which be independent of the degradation speed for the quality function during a particular handover. HQM is given by the following:

$$HQM = \frac{RI}{\frac{\sum_{k=h_{PH}}^{h_{RH}+D} \Delta Q(k)}{h_{RH}+D-hPH}} \tag{4.15}$$

4.2.6 Simulation Results

The reactive and proactive vertical handover algorithms based on the difference quality function Q have been tested through simulations in a vertical handover scenario from DVB-H to UMTS.

The simulated scenario consists of a region of 1000x1000 zones of 1 square meter. Some portions of the region are covered by the DVB-H signal, others by the UMTS one and some are covered by both signals. A vertical handover takes place in the region with double UMTS and DVB-H coverage.
As far as the quality function parameters are concerned we have given a weight equal to 0.8 to the bit-error-rate contribution to the quality function and 0.1 to the other two contributions related to power saving and user preferences respectively. The values summarized in Table 4.1 determines the portions of the region in the simulated scenario.

In such scenario the user moves in an area where UMTS and DVB-H connectivity are both present. The user crosses the curve where Q turns from positive to negative in the decision plane and executes a vertical handover from DVB-H to UMTS. We have assessed:

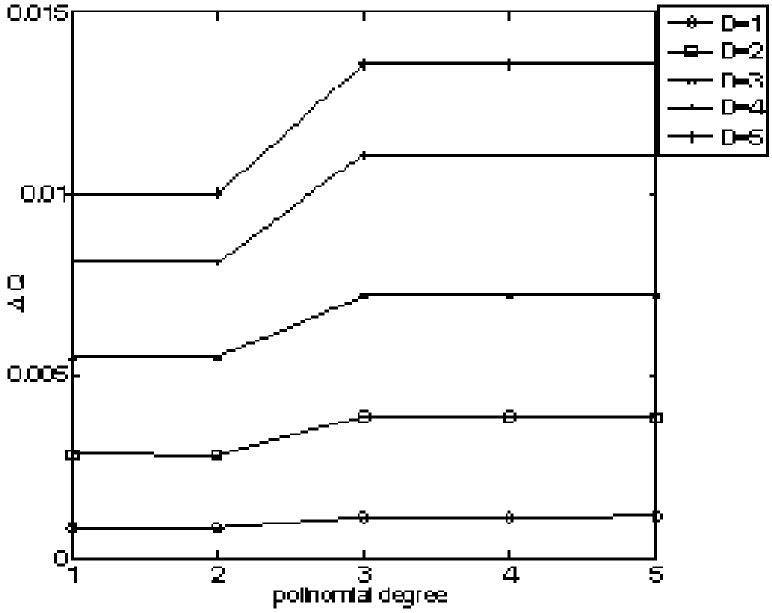

Figure 4.2: ΔQ for different handover duration and various interpolation degree

Table 4.1: Network Parameters

Parameter	UMTS	DVB-H
Base Station Power [dBm]	38-43	59-61
Power Save [%]	0.42	0.91
Noise Figure [dB]	7	5
Bandwidth [Mhz]	5	8
Thermal Noise	-174	-174
Attenuatuion Model [dB, distance in Km]	Okomura Hata	Cost 231
Transmission Frequency [Mhz]	2140	700

- the dependency between Q and the degree of the polynomial interpolation for prediction of future changes of Q;
- the quality gain, in terms of the HQM parameter, vs. the handover duration D.

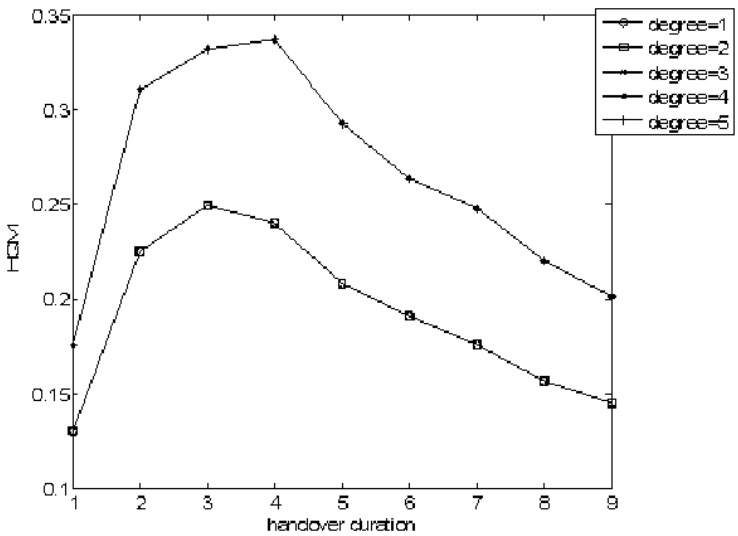

Figure 4.3: HQM for different handover duration and various interpolation degree

Statistics for the first assessment are shown in Fig. 4.2. This depicts the Q function vs. polynomial interpolation degree. A separate curve is provided for a number of handover duration intervals D. Each curve is obtained considering the mean values of Q over different crossing angles of the $Q = 0$ curve.
It is worth highlighting that there is no gain for a degree above 3 also for high handover time D because it is possible that the prediction fails for some crossing angles.

Fig. 4.3 depicts the HQM parameter vs. handover duration. A number of curves for difference polynomial interpolation degree are shown (notice that curves 1 and 2 as well as 3,4 and 5 overlap onto two distinct profiles). Mean values of HQM over different crossing angles a reported. Depending on the polynomial interpolation degree a maximum value for HQM can be reached when $D = 3$ (degree 1-2) or $D = 4$ (degree 3-5), which is approximately equal to 25% (degree 1-2) or 35% (degree 3-5). We can finally conclude that polynomial interpolation of the 3rd order exhibits the best performance, which are obtained when $D = 4$.

Table 4.2: Network Attributes and Parameters Relationship

Attributes	Parameters
Allowed bandwidth	modulation, coding, bandwidth
Packet delay	ARQ scheme, power
Packet drop	queue management

4.3 VHO as a QoS constrain problem

4.3.1 Problem definition

Generally, as described in the previous sections, VHO may be formulated as a cost minimization problem: each network is associated with a cost that depends by the values of the M parameters that it is currently experiencing.
If we indicate with x_i the point $x_i \in \Re^M$ that is the representation of the $i-th$ networks and with $x_{i|l}$ the $l-th$ parameters value the problem can be formulated as follows:

$$\begin{aligned} &\arg\min_{x_i} c(x_i) \\ &x_{i|l} \geq k_l \;\; i = 1, ..., N, l = 1, ..., M \end{aligned} \tag{4.16}$$

where $c(x_i)$ is the cost of using the $i-th$ network and k_l express the QoS constrain respect to the $l-th$ parameters. It is important to note that k_l may be also equals to 0 in such case no bound is active on the $l-th$ parameter.
As an example for the cost function it is possible to use the distance between x_i and 1 assuming normalized all the parameters values. In such case each point x_i is inside a unit hypercube, with $\mathbf{0} = [\mathbf{0}, ..., \mathbf{0}]$ representing the lowest bound and $\mathbf{1} = [\mathbf{1}, ..., \mathbf{1}]$ the highest.

4.3.2 Limits

As already said in literature the networks are identified with a single point, in fact it is assumed that other procedures may correctly identified the “best” set parameters for each network. It should be noted that the notion of “best”

parameters may depend by the target function of each procedures that may not be the same as for VHO.
The mobile terminal may decide several parameters on the wireless link that directly affects network performances, e.g. power vs bit error rate, therefore each network may be represented not only be a unique point, i.e. x_i but by a set of points with different cardinality. Such set may be generally theoretically infinite but for practical purpose only a limited parameters values may be used therefore such set X_i may have cardinality n_i, i.e. $X_i = (x_{i,1}, x_{i,2}, ..., x_{i,n_i})$. An example of parameters that affect network attributes is reported in Table 4.2.
If a network evaluation criteria is decided as a function of $x_{i,j}$, i.e. $c(x_{i,j})$ the problem defined in eq. 4.16 may be formulated as follows:

$$\begin{aligned} &\arg\min_i \min_j c(x_{i,j}) \\ &x_{i,j} \geq k_l \; i = 1, ..., N, j = 1, ..., n_i, l = 1, ..., M \\ &x_{i,j} \leq 1 \end{aligned} \tag{4.17}$$

Using eq. 4.17 the mobile terminal selects the mode, i.e. the combination of parameters, of the network with the minimum distance to the optimal point.
As an example of 4.17 in Fig. 4.4 we can note the different points that characterize a IEEE802.11a network using goodput and delay as attribute and different modes and power as parameters. The calculus of goodput and delay is based on the analysis carried out in [39].
It should be noticed that the networks may be not generally totally ordered:

$$\begin{aligned} &\exists j_1, j_2, j_3 | c(x_{i1,j1}) \leq c(x_{i2,j3}) \\ &\wedge c(x_{i2,j3}) \leq c(x_{i1,j2}) \end{aligned} \tag{4.18}$$

where $\wedge$ represents the logical and between conditions. A general order may be applied to network and modes:

$$x_{i1,j1} \preceq x_{i2,j1} \preceq ... \preceq x_{in,jm} \tag{4.19}$$

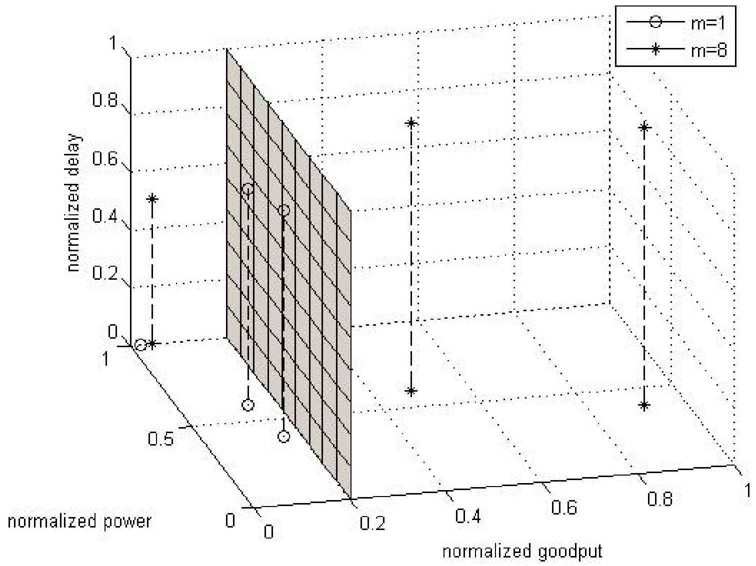

Figure 4.4: $x_{i,j}$ for IEEE802.11a for mode 1 and 8 with a constrain on normalized goodput equals to 0.2

where the symbol $\preceq$ represents a total order relationship between points, i.e. network and modes.
The distances are calculated at a specific time instant k_i therefore the order at time k_1 may be different to a order at k_2. To avoid a ping pong effect a specific handover cost between network i and j may be introduced $c_{i,j} i, j = 1, ..., N$. Let us suppose that the mobile terminal is attached to network $i1$ with the mode $j1$ therefore $d(x_{i1,j1})$ is the lowest among all the distances at time $k1$.At time instant $k2$ the new lowest distance may be $d(x_{i2,j1})$ then the mobile terminal executes a VHO if and only if:

$$c_{i1,i2} d(x_{i2,j1}) \leq \min_j d(x_{i1,j}) \tag{4.20}$$

in 4.20 two costs are compared but $c_{i1,j}$ is assumed equals to 1.

4.3.3 Logical representation for VHO

In the previous section we defined the problem considering the possible optimization of network parameters, in Fig. 4.5 a possible logical scheme to implement the procedure described in Par. 4.3.2 is depicted.

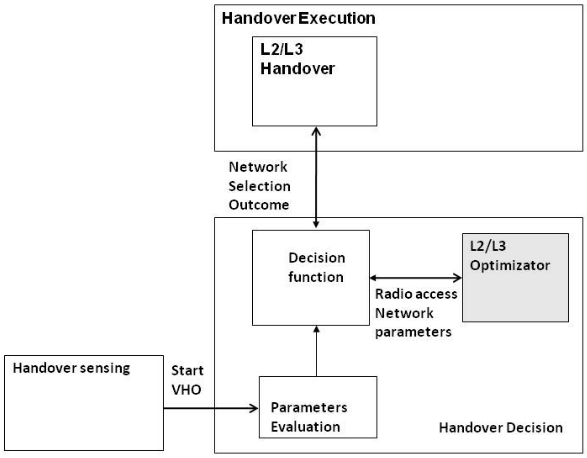

Figure 4.5: Logical representation for VHO

It is important to highlight that we do not propose a full architecture reference model but only define a logical representation of necessary blocks VHO (for models available in literature please refer to [19]).

It should be noted that the logical scheme defines the three phases of the VHO as functional blocks but they may be implemented differently, i.e. jointed or separately. It can also integrated into different functional architecture, as an example the one of IEEE802.21. The three functional blocks are:

- a *handover sensing* module that detects changes in QoS parameters, to avoid ping-pong effect([45]) the handover may be triggered only when QoS parameters are not sufficient.
- *handover decision* that takes the decision upon the network that should be selected basing not only on network parameters but also by information provided on network optimizator modules that establish the parameters for each network and how they may influence the handover decision.
- *handover execution* including layer two and three procedures that are necessary to accomplish the handover.

4.4 Geometric Strategies for VHO

In the following sections two strategies are described: the former utilizes the power as a decision criteria while the latter minimize the number of VHOs maintaining the mobile terminal connected to the chosen network as long as possible.

4.4.1 Minimum Power Strategy (mP)

Power may directly influence bit error rate and battery saving therefore it may be considered a key factor not only in VHO but also on all the procedures that the mobile terminal (MT) should execute.
Two power allocation are generally used in the wireless link: uplink, from the access point to the MT and downlink, from the MT to the access point. In the following we indicate with P_i the uplink power that should be allocated to network $i = 1, ..., N$ to achieve a minimum quality of service expressed as a function of bit error rate. In 4.21 we indicate the selection of the network with the lowest power consumption.

$$\begin{cases} \min_i P_i \\ P_i \geq 0 \\ P_i \leq P_{i_{max}} \\ f_l(P_i) \geq k_l \end{cases} \tag{4.21}$$

where $f_l(P_i)$ indicates the QoS constrain of the $l-th$ parameter as a function of power.
It can be noticed that the network are totally ordered by their power usage:

$$\begin{aligned} & i_1 \preceq i_2 \preceq ... \preceq i_N \Rightarrow \\ & P_{i_1} \leq P_{i_2} \leq P_{i_3} \leq ... \leq P_{i_N} \end{aligned} \tag{4.22}$$

If adaptive modulation and coding is considered each technology can be identified by different modes. We indicate with $P_{i,j}$ the power that can be allocated to technology i using the mode j

$$\begin{cases} \min_i P_{i,j} \\ P_{i,j} \geq 0 \\ P_{i,j} \leq P_{i_{max}} \\ f_l(P_{i,j}) \geq k_l l = 1, ..., M \end{cases} \tag{4.23}$$

It should be noticed that the network may be no longer totally ordered, i.e.

$$\begin{aligned} &\exists j_1, j_2, j_3 \| P_{i_1,j_1} \preceq P_{i_2,j_3} \\ &\wedge P_{i_2,j_3} \preceq P_{i_1,j_2} \end{aligned} \tag{4.24}$$

It is possible to consider a safeguard interval that permits to compensate eventual losses $k_l^{'} = k_l + \Delta k_l l = 1, ..., M$ then it is possible to compute the minimum power for each attribute as: $P^*_{i,j\|l} = \arg\min[f_l(P_{i,j}) \geq k_l^{'}]l = 1, ..., M$ and the maximum considering all the parameters $P^*_{i,j} = \max_l P^*_{i,j\|l}$.

4.4.2 Minimum handover strategy mH

In such strategy the mobile terminal stays connected to its network if the QoS bounds are satisfied by at least a network mode independently by its efficiency. More specifically the mobile terminal do not follow the "best" network but use the one that is currently available. The "best" network is used only as the first attachment point.

The rationale of such strategy is to avoid any VHO, in fact the VHO may have a cost in terms of used resources, mainly power consumption for signaling, that do not compensate any beneficy that may arise due to the use of a better network. Summarizing the mobile terminal should execute the following steps, described in Table 4.3:

- $t = 0$ determine the network according to 4.16 and attach to the network;
- if the network satisfy the QoS bound mobile terminal stay connected to the network;

Table 4.3: Minimum handover strategy MH

1.	$t = 0$ determine the best network $i = \arg\min_{x_{i,j}} c(x_{i,j})$ $x_{i,j} \geq k_l \ i = 1, ..., N, j = 1, ..., n_i, l = 1, ..., M$ $x_{i,j} \leq 1$ $\Rightarrow mt(0) = i;$
2a.	if $\exists j \| x_{i,j} > k_l l = 1, ..., M$ $\Rightarrow mt(t) = i$
2b.	otherwise $i = \arg\min_{x_{i,j}} c(x_{i,j})$ $x_{i,j} \geq k_l \ l = 1, ..., N, j = 1, ..., M$ $x_{i,j} \leq 1$ $\Rightarrow mt(t) = i;$

- if the network do not longer satisfy all the QoS bound for all the possible modes a new network may be computed: if no network is available the connection is maintained to the original network otherwise the mobile terminal switch to the new network.

4.5 Numerical results

In this section we evaluate the two strategies: minimum power and minimum changes in terms of number of VHO and used power in a scenario composed by 2 UMTS BS and 5 IEEE802.11 AP, the mobile terminal path is the same for all the scenario. An example of a random scenario is depicted in Fig. 4.6.

We considered the presence of an AMC that tries to maximize throughput and the presence of a L2/L3 optimizator (without AMC in the rest of the section). In Table 4.4 the main simulation parameters are reported, it should be noted that AMC is possible using WiFi technology while it is not applicable to UMTS therefore in the analyzed scenario only one technology supports it, such approach permits to highlight the impact of AMC on VHOs.

In Fig. 4.7 the mean and variance of the power usage for both mH and mP are depicted. Integrating the AMC in the VHO permits to reduce the mean power

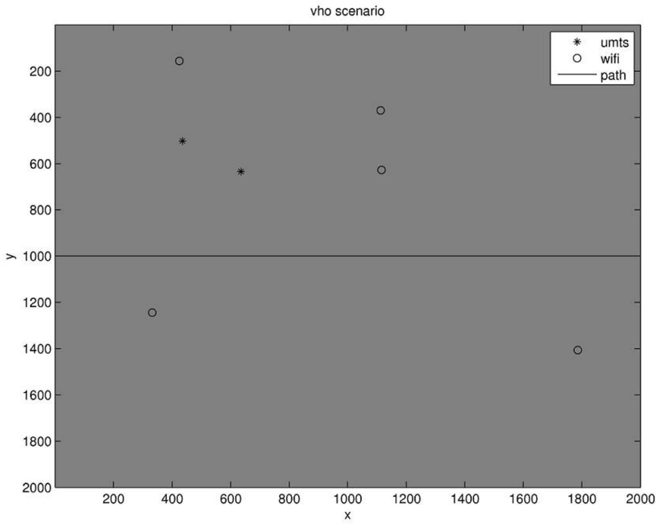

Figure 4.6: Random scenario example

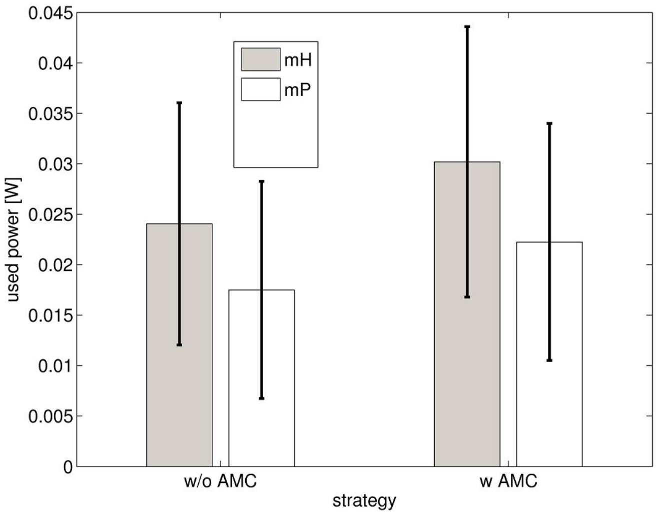

Figure 4.7: Power usage in random scenarios

Table 4.4: Simulation Parameters

Parameter	UMTS	WiFi
Maximum Power [mW]	125mW	125mW
AMC	N/A	64-QAM/16-QAM/PSK
Attenuation Model	Two-Slope	ITU-T Indoor

usage from 0.0302 mW to 0.0240 mW for mH and from 0.0222 mW to 0.0175 mW for mP (both have the same reduction amount of 20%). Additionally the following issue can be noticed:

- using AMC has an impact over power consumption due to the fact that the "best" modulation is used. The same impact can be seen both in mP and mH strategy therefore if adaptive modulation and coding is not designed jointly with handover a wasting of resources may happen.
- mP strategy has a decrease of 20% with respect to mH.

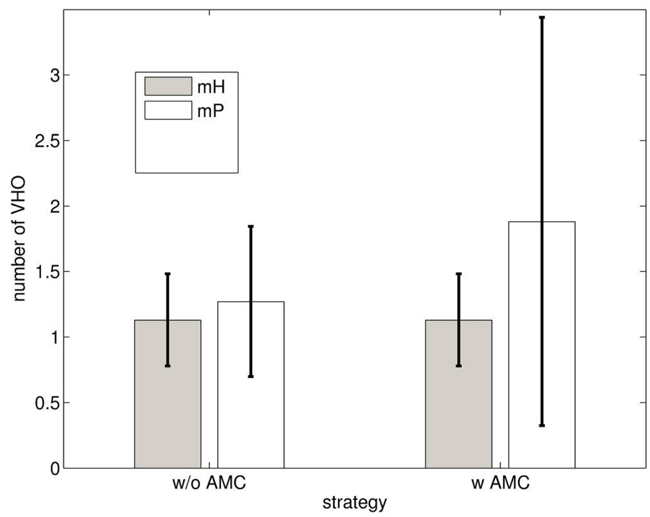

Figure 4.8: number of VHOs

In Figure 4.8 the number of VHO is depicted with and without AMC. In such case the number of VHO for mH is the same with and without AMC, in fact mH stays connected to the network independently by power issues but the handover is triggered only by connectivity problems, i.e. connectivity loss toward AP.
mP with an integrated AMC permits to reduce the number of VHO from a mean of 1.8800 to a mean of 1.2700 (around a 30% reduction) due to the fact that integrating AMC can permit to avoid ping-pong effect, it can also be noted by the variance reduction in the number of VHOs.

4.6 Final remarks

In the present chapter we discuss the modeling of UMTS and DVB-H parameters including power saving and quality of service, such modeling has been employed in the definition of a reactive and proactive vertical handover procedure.
The usage of a proactive approach, based on linear interpolation, permits to achieve better results than a reactive one.
Considering the technology as the only measured parameters is limiting and may hide a possible optimization achieved through the exploitation of adaptive modulation and coding.
In the second part of the present chapter, we discuss the already defined limits of a single objective function approach and defined strategies to overcome such limitations: the former aimed to minimize the used power and the latter aimed to minimize the number of vertical handover.
Numerical results to assess the impact of a jointly procedure of adaptive modulation and coding and vertical handover have been presented.

Chapter 5

Reducing the ping pong effect in vertical handover: a probabilistic approach

5.1 Introduction

Current wireless and cellular systems aim to provide limited and dedicated services. Next Generation Networks (NGNs) environments will be depicted by a heterogeneous network set, consisting of different overlapping access networks [5]. In such scenario seamless connectivity represents a strong requirement for mobile users, who want to reach on-the-move connectivity for multimedia services (i.e. video-conferencing, video-on-demand, Internet browsing, on-line gaming, and so forth).

Vertical Handover (VHO) mechanism aims to maintain service connectivity for mobile users in heterogeneous environments where multiple wireless network co-exist [5]. A VHO process moves a user service from a Serving Network to a Candidate Network, in order to keep users always connected to the best available network.

The importance and usage of VHO techniques has been rapidly increasing in the last few years, due to the growing diffusion of multi-mode Mobile Terminals (MTs), equipped with several network interface cards (i.e., IEEE 802.11, UMTS,

GPS, etc.). The main challenges to support an efficient VHO execution take part in the handover initiation phase, which focuses on (i) the choice of the best available Candidate Network, and (ii) the decision strategy driving the switch from the Serving Network to the Candidate Network [5]. The handover initiation is traditionally based on physical parameters, such as the Received Signal Strength, and the Signal-to-Noise and Interference Ratio, [54]. Though the goal of connectivity switching should improve or maintain equivalent user Quality-of-Service level, quick and frequent VHOs can strongly affect the mobile terminal power consumption, and cause the unwanted so-called ping-pong effect [30]. It is easy to understand this effect, by supposing how frequent connectivity switches would occur when a mobile user is moving along on the bounder-line of a wireless cell.
A properly designed handover initiation algorithm should decrease the number of unnecessary handovers, mitigating, and controlling the ping-pong effect. Traditionally, techniques for ping-pong avoidance require additional information, like channel estimation [24] or location information [25], although they are not always available. Other approaches relay on clustering of different users and adaptive hysteresis, as described by Lee et al. in [33]. The effectiveness of their technique has been evaluated in terms of number of vertical handovers only, and no network parameter has been investigated. In [14] Chi et al. propose a technique for limitation of the ping-pong effect, by exploiting the estimation of the Wrong Decision Probability, i.e. the probability that an unnecessary handover is initiated, and then executed. A similar concept has been assumed in [45], where the Wrong Decision Probability calculus is based on a linear correlation between goodput performances measured at different time samples. The goodput is used as a common metric among heterogeneous networks, while the linear correlation allows a simple and efficient calculus.
In this chapter our contribution is to extend the previous work in [45] and in [3], by assessing further correlation approximations for Wrong Decision Probability (WDP) (i.e., not just linear, but also uniform, and exponential), in order to obtain the best approximation for the WDP, with respect to the limitation of the ping-pong effect. Comparison between three proposed WDP distributions will be described. The chapter is structured as follows. Section 5.2 deals with recent related work about the handover probability, along with our contribution. In

Section 5.3 we give basic definitions in order to introduce the WDP, while in Section 5.4 the general descripion of the proposed algorithm is included. The three probability distributions for WDP are introduced (i.e., uniform, exponential, and linear) in 5.5. Simulation results in terms of network performance and limitation of ping-pong effect are reported in Section 5.6, through a comparison between the proposed three WDP approximations. Finally, in Section 5.7 we conclude the chapter.

5.2 Related Work

We use a novel approach, where the handover probability is the criterion to perform correct vertical handover decisions (i.e., WDP). Our technique works into a VHO algorithm, which prevents wrong VHOs, and limits the ping-pong effect by the use of WDP parameter [45]. The Wrong Decision Probability represents a VHO decision criterion to execute right handovers. This approach should result differently from many previous techniques for VHO decisions, since in the literature many authors analyze the performance of a handover algorithm in terms of handover probability [32], [50],. In [32] Lal et al. introduce a handover algorithm based on the so-called adaptive hysteresis margin, which reduces the probability of unnecessary handovers, and maintains a constant Quality-of-Service (QoS) level. The handover probability is evaluated as a performance result of the proposed handover algorithm. As the same, in [53] Zhang et al. analyze the impact of the main wireless channel characteristics upon the handover probability, and system performance are evaluated in terms of the probability to perform a vertical handover. In [52] Zhang explicitly defines the handover probability for a Rayleigh fading scenario. His results shows the direct link between handover probability, and physical characteristics of wireless networks. Finally, in [50] Zarai et al. describe a novel technique to reduce the handoff probability in next generation wireless networks. The authors consider the use of an adaptive resource reservation scheme providing QoS, as well as physical parameter (i.e. Received Signal Strength).

We don't use the handover probability as a metric for VHO process, but as a criterion to avoid unnecessary VHOs. The concept of WDP is described, and analyzed through three different probability distributions, in order to validate which of them gives best performance in terms of number of VHOs, and Cumulative Received Bits.

5.3 Wrong decision Probability Theoretical Approach

In this section, we introduce the Wrong decision Probability (WDP) for VHO execution. A multi-network environment is considered in order to describe our referred scenario. Moreover, we give a first definition about the Delta-goodput discrete time stochastic process, for which we evince the WDP definition.
The general definition have already been introduced in [45], in the present section we will give a more formal approach to the WDP.
Let us consider a heterogeneous scenario, composed by two wireless networks, denoted by i and j, respectively. We assume $GP_i[k]$ and $GP_j[k]$ as two discrete-time stochastic processes, which represent the goodput assessment at time instant k for network i and j, respectively.

Definition 1. *Delta-goodput discrete time stochastic process: given two goodput discrete-time stochastic processes, sampled at time instant k in network i, and j (i.e. $GP_i[k]$, and $GP_j[k]$), the delta-goodput discrete time stochastic process is the difference between $GP_i[k]$ and $GP_j[k]$, such as:*

$$\Delta GP[k] = GP_i[k] - GP_j[k] \tag{5.1}$$

Definition 2. *Wrong Decision Probability: given a value g^* for the Delta-goodput discrete time stochastic process, sampled at time instant k (i.e. $GP[k] = g^*, g^*$), the Wrong Decision Probability (WDP) is the conditional probability that $GP[k+1] < 0$ when $GP[k] = g* > 0$, or equivalently, the conditional probability that*

$GP[k+1] > 0$ *when* $GP[k] = g^* < 0$, *such as*

$$WDP(g^*) = \begin{cases} P(\Delta GP[k+1] < 0|\Delta GP[k] = g^* & g^* > 0 \\ P(\Delta GP[k+1] > 0|\Delta GP[k] = g^* & g^* < 0 \end{cases} \tag{5.2}$$

Notice that the Wrong Decision Probability is dependent on the actual value of $GP[k] = g*$. In the rest of this chapter, for sake of compactness we will simply write the WDP without g^* dependence.
The WDP gives information about sign changes of delta-goodput process (i.e. from positive to negative, and vice versa). In 5.2 the value $g^* = 0$ can be included in the WDP definition, but in such case we should assume a sign transition from positive to negative values, crossing the zero, and vice versa.
The name of wrong decision is due because if a vertical handover occurs at time instant $k+1$, where $GP[k+1] < 0$, while $GP[k] = g^* > 0$, then the handover will be an unnecessary handover. Since the goal of a VHO mechanism is to obtain network performance maximization, in the case of fast Delta-goodput sign changes, performance are strongly affected. The calculus of WDP strictly depends on the Delta-goodput process conditional probability p_{cond} (i.e. $P(GP[k+1] < 0|GP[k] = g^* > 0)$, and $P(GP[k+1] > 0|GP[k] = g^* < 0)$). In the following, we shall give a formal definition for the conditional probability p_{cond}, as follows:

Definition 3. *Delta-goodput Conditional Probability: The Delta-goodput conditional probability pcond represents the probability distribution of Delta-goodput stochastic process, sampled at time instant k +1 (i.e. GP [k+1]), given the value of Delta-goodput stochastic process at time instant k (i.e. GP [k] = g*). It is expressed by the following formula:*

$$p_{cond}(g) = P(\Delta GP[k+1] = g|\Delta GP[k] = g^*) \text{ where } g \in \Delta GP \tag{5.3}$$

From the Wrong Decision Probability can be calculated as follows:

$$WDP = \begin{cases} \int_0^{\Delta GP_{max}} p_{cond}(g)\, dg & g^* > 0 \\ \int_{\Delta GP_{min}}^0 p_{cond}(g)\, dg & g^* < 0 \end{cases} \tag{5.4}$$

where ΔGP_{max} and ΔGP_{min} represent the maximum and minimum values for Delta-goodput process, respectively. By assuming that GP_i and GP_j have (i) a maximum value (i.e. $maxGP_{i,j}$, dependent on the particular radio technology i and j, respectively), and (ii) a minimum value equal to zero (i.e. $minGP_{i,j*} = 0$), being the network capacity always non-negative, or at least equal to 0 when no connection is available, then GP_{max} and GP_{min} was calculated as:

$$\begin{cases} \Delta GP_{max} = max\ GP_i - min\ GP_j \\ \Delta GP_{min} = min\ GP_i - max\ GP_j \end{cases} \Leftrightarrow \begin{cases} \Delta GP_{max} = max\ GP_i - min\ GP_j \\ \Delta GP_{min} = min\ GP_i - max\ GP_j \end{cases} \tag{5.5}$$

From [45] we recall that in order to prevent low accuracy in the assessment of ΔGP samples, in the rest of the chapter we assume that a first order exponential smoothing approximation is applied to the sequence $\Delta GP[k]$, as follows:

$$\langle \Delta GP[k+1] \rangle = \alpha \langle \Delta GP[k+1] \rangle - (1-\alpha)\Delta GP[k] \tag{5.6}$$

where the parameter α is in the range $[0,\ 1)$. Finally, after introducing the Delta-goodput stochastic process, the conditional probability p_{cond}, and the WDP, we describe how our proposed VHO algorithm works, in order to (i) maximize throughput performance, and (ii) limit the ping-pong effect. The algorithm considers a probability threshold P_{TH}, which defines the following metric for a correct vertical handover decision, such as:

$$\begin{cases} WDP < P_{TH} \Rightarrow & \text{a VHO occurs} \\ WDP > P_{TH} \Rightarrow & \text{a VHO does not occur} \end{cases} \tag{5.7}$$

As can be deduced by Eq. 5.7 the probability threshold P_{TH} represents the minimum value for WDP below which a vertical handover is executed, whenever a sign transition is detected. In the following sections, we briefly shall describe the main tasks acted by our VHO algorithm. For further details, we recall the previous work in [45] and in [3].

5.4 WDP-Based Vertical Handover algorithm

The initial value $GP[0]$ defines which network acts as Serving Network (SN) (i.e. if $GP[0] < 0$, then network j is the SN, otherwise network j). The SN is chosen on the basis of a goodput comparison (i.e., the network with higher goodput acts as Serving Network). The detection of a positive sign transition occurs when the predicted ΔGP sequence at time instant $k+1$ is α-times higher than the actual ΔGP sampled at the same time instant, such as

$$\langle \Delta GP[k+1] > 0 \rangle \Rightarrow \langle \Delta GP[k+1] \rangle > \alpha \Delta GP[k+1] \tag{5.8}$$

or equivalently, a negative sign transition occurs when the predicted ΔGP sequence at time instant $k+1$ is α-times lower than the actual ΔGP sampled at time instant $k+1$,

$$\langle \Delta GP[k+1] < 0 \rangle \Rightarrow \langle \Delta GP[k+1] \rangle < \alpha \Delta GP[k+1] \tag{5.9}$$

When a sign transition negative, or positive occurs, a VHO process initiated. The VHO execution is due only if the estimated WDP is below the chosen threshold P_{TH}; otherwise, the WDP will be recomputed after Δt seconds, till either a sign transition should occur, or the WDP would be lower than P_{TH}.

5.5 Conditional Probability approaches

In this Section we shall introduce three different probability distributions for WDP, which have been assumed in our work. As expressed in 1, the Delta-goodput is a discrete time stochastic process, which approximates the real continuous time process, where consecutive time instants are sampled each Δt time interval. As a consequence, we can reasonably assume that if Δt is very large (i.e., $\Delta t \to \infty$), there is no correlation between consecutive goodput samples. Conversely, if t is too short (i.e., $\Delta t \to 0$), there should be no difference between two consecutive goodput samples. By leveraging this consideration, we can approximate the exact estimation of the p_{cond} is out of the scope of this work the conditional probability to a Dirac delta function, as follows:

$$p_{cond}(\Delta t) = \begin{cases} 0, & \Delta t \Rightarrow \infty \\ \delta_0 & \Delta t \Rightarrow 0 \end{cases} \tag{5.10}$$

In the rest of the chapter a set of inter-medium cases (i.e., for $0 < \Delta t < \infty$) has been assumed. The analysis of the behavior of the VHO approach has been addressed with respect to three different conditional distributions, such as:

1. Uniform distribution, where no dependence between the Delta-goodput process sampled at the time instant k (i.e. $DeltaGP[k]$), and the sample at next time instant $k + 1$ (i.e. $\Delta GP[k + 1]$), is assumed;

2. Exponential distribution, where an exponential distribution has been assumed for the dependence between $\Delta GP[k]$, and $\Delta GP[k + 1]$;

3. Linear distribution, where a linear dependence between $\Delta GP[k]$, and $\Delta GP[k+1]$, has been assumed, as illustrated in [45].

5.1 depicts the graphical meaning of the WDPs, by considering the areas under the three distributions for positive values (i.e. white, grey, and black areas are for uniform, exponential, and linear distributions, respectively). The following Subsections A, B, and C describe the WDP calculus for the uniform, exponential, and linear approaches, respectively.

5.5.1 Uniform Distribution

For the uniform distribution (defined in [45] and reported here for completeness), the conditional probability p_{cond}^{unif} was expressed as:

$$p_{cond}^{(unif)} = \begin{cases} \frac{1}{\Delta GP_{max} + \Delta GP_{min}} & , -\Delta GP_{min} < \Delta GP < \Delta GP_{max} \\ 0 & , \text{otherwise} \end{cases} \tag{5.11}$$

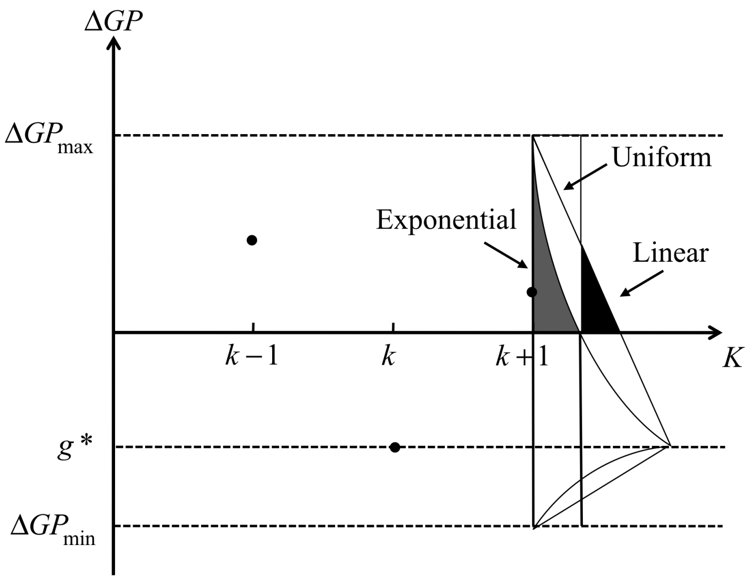

Figure 5.1: Wrong Decision Probability for different p_{cond} distributions. White, grey and black areas depict the WDPs for uniform, exponential, and linear distributions, respectively.

from which the WDP was directly calculated from 5.4 as:

$$WDP = \begin{cases} \frac{\Delta GP_{max}}{\Delta GP_{max}+\Delta GP_{min}} & , g^* > 0 \\ \\ \frac{-\Delta GP_{min}}{\Delta GP_{max}+\Delta GP_{min}} & , g^* < 0 \end{cases} \tag{5.12}$$

5.12 depends on the values of ΔGP_{max} and ΔGP_{min}, and by their ratio. Such dependence is related to the unbalancing degree of the WDP distribution.

5.5.2 Exponential Distribution

For the exponential distribution, the conditional probability is expressed as follows:

$$p_{cond}^{(exp)} = \begin{cases} y_1(g) & , \Delta GP_{max} \leq g \geq g^* \\ y_2(g) & , \Delta GP_{min} \geq g \leq g^* \\ 0 & , \text{otherwise} \end{cases} \tag{5.13}$$

where y_1 and y_2 are two exponential functions, both joining at the common point g*, such as:

$$\begin{cases} y_1(g) = c_1' e^{-g+g^*} + c_1'' \\ y_2(g) = c_2' e^{(g-g^*)} + c_2'' \end{cases} \tag{5.14}$$

5.13 depict both positive and negative values of exponential Delta-goodput distribution, whose maximum value is reached at $g = g*$. Let us assume that $y_1(g) = 0$ for $g = \Delta GP_{max}$, while $y_2(g) = 0$ for $g = \Delta GP_{min}$, such as:

$$\begin{cases} y_1(g = \Delta GP_{max}) = c_1' e^{(-\Delta GP_{max}+g^*)} + c_1'' = 0 \\ y_2(g = \Delta GP_{min}) = c_2' e^{(\Delta GP_{min}-g^*)} + c_2'' = 0 \end{cases} \tag{5.15}$$

From 5.15 we obtain the constraints c_1'' , and c_2'' , respectively

$$\begin{cases} c_1'' = -c_1' e^{(-\Delta GP_{max}+g^*)} \\ c_2'' = -c_2' e^{(\Delta GP_{min}-g^*)} \end{cases} \tag{5.16}$$

which, when replaced in 5.14, it becomes:

$$\begin{cases} y_1(g) = c_1'[e^{g^*}(e^{-g} - e^{-\Delta GP_{max}}] \\ y_1(g) = c_2'[e^{-g^*}(e^{g} - e^{-\Delta GP_{min}}] \end{cases} \tag{5.17}$$

Moreover, as defined in 5.13 we have y1(g*) = y2(g*), and 5.14 can be written as

$$c_1'[1 - e^{-\Delta GP_{max}} + g^*] = c_2'[1 - e^{\Delta GP_{min}} - g^*] \tag{5.18}$$

from $which c_1' = ac_2'$, where

$$a = \frac{1 - e^{\Delta GP_{min} - g^*}}{1 - e^{-\Delta GP_{max}} + g^*} \tag{5.19}$$

As the integral of the probability distribution WDP must be unitary, such as

$$\int_{\Delta GP_{min}}^{g^*} y_2(g)\, dg + \int_{g^*}^{\Delta GP_{min}} y_1(g)\, dg \tag{5.20}$$

we obtain,

$$\int_{\Delta GP_{min}}^{g^*} c_2'[e^{(g-g^*)} - e^{(\Delta GP_{min} - g^*)}]\, dg + \int_{g^*}^{\Delta GP_{min}} c_1'[e^{(-g+g^*)} - e^{(-\Delta GP_{max} + g^*)}]\, dg = 1 \tag{5.21}$$

Equation 5.21 becomes:

$$c_2'[1 - e^{(\Delta GP_{min} - g^*)}(1 + g^* + \Delta GP_{min})] + c_1'[-1 + e^{(-\Delta GP_{max} + g^*)}(1 + g^* - \Delta GP_{max})] = 1 \tag{5.22}$$

where

$$c_1' = \frac{a}{B + aA}, c_2' = \frac{1}{B + aA} \tag{5.23}$$

with

$$\begin{aligned} A &= 1 - e^{(-\Delta GP_{min} - g^*)} - e^{(\Delta GP_{min} - g^*)}(g^* - \Delta GP_{min}) \\ B &= -1 + e^{(\Delta GP_{max} - g^*)} - e^{(\Delta GP_{max} + g^*)}(-g^* + \Delta GP_{max}) \end{aligned} \tag{5.24}$$

The probabilities p1 and p2 for the components of WDP, defined as:

$$WDP = \begin{cases} p_1(g) & , g^* \leq 0 \\ p_2(g) & , g^* \geq 0 \end{cases} \tag{5.25}$$

are respectively,

$$p_1(g) = \int_0^{\Delta GP_{max}} y_1(g)\,dg = \\ c_1' \Delta GP_{max} - c_1' e^{\Delta GP_{max}} \Delta GP_{max} = \\ c_1' \Delta GP_{max}(1 - e^{-\Delta GP_{max}}) \tag{5.26}$$

and

$$p_2(g) = \int_{\Delta GP_{max}}^{0} y_2(g)\,dg = \\ - c_2' \Delta GP_{min} + c_2' e^{\Delta GP_{min}} \Delta GP_{min} = \\ c_2' \Delta GP_{min}(-1 + e^{-\Delta GP_{min}}) \tag{5.27}$$

Finally, we can get the WDP for the exponential distribution as follows:

$$WDP = \begin{cases} c_1' \Delta GP_{max}(1 - e^{-\Delta GP_{max}}) & , g^* \leq 0 \\ c_2' \Delta GP_{min}(-1 + e^{-\Delta GP_{min}}) & , g^* \geq 0 \end{cases} \tag{5.28}$$

In the exponential case the WDP has two drawbacks that should be considered, and for which a solution is proposed, such as:

- The WDP calculus is more complex with respect to the uniform and linear approaches. Proposed solution: the calculus can be performed for constant time, and a pre-calculate table can be used to have data already available;
- The value of WDP can be extremely low (i.e. null approximation, due to the decreasing exponential component), and then any implementation should be able to deal with a so low accuracy. Proposed solution: usage of a normalization of the GP values.

5.5.3 Linear Distribution

The WDP linear approach has been previously given in [45]. It is important to remark that this distribution differs with respect to:

- The uniform approach, because it is depending on the g* value;
- The exponential approach, because it does not suffer of its drawbacks except for the null approximation for some peculiar cases, (i.e. when g* is very close to GPmax).

We reminded the WDP:

$$WDP = \begin{cases} -\frac{\Delta GP_{max}^2}{(\Delta GP_{max}-\Delta GP_{min})(g^*-\Delta GP_{max})} & , g^* < 0 \\ \frac{\Delta GP_{min}^2}{(\Delta GP_{max}-\Delta GP_{min})(g^*-\Delta GP_{min})} & , g^* > 0 \end{cases} \tag{5.29}$$

5.6 Numerical Results

This section deals with the main simulation results which validate which proposed WDP distribution better works to (i) limit the ping-pong effect, and (ii) maximize the throughput.
The three different WDP distributions have been modeled in a MATLAB simulation environment. We refer to a network setup as depicted in [45], with two different available networks, representing a medium-range (i.e., cell radius around 500 m), and small-range (i.e., cell radius around 100 m) transmission coverage network, respectively.

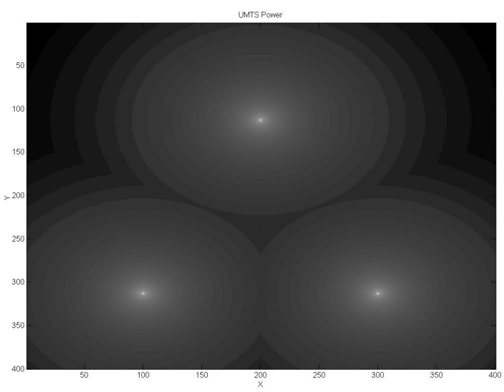

Figure 5.2: Power Network 1

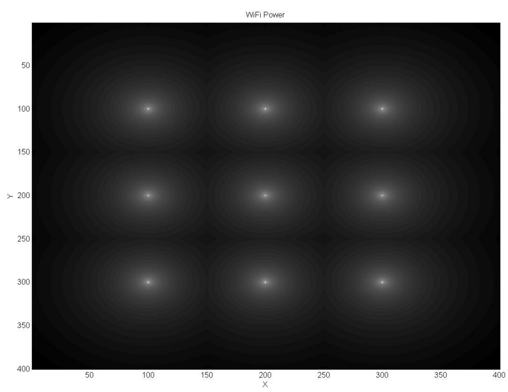

Figure 5.3: Power Network 2

Simulations parameters are reported in Table 5.6 and power profiles in Fig. 5.2 and in Fig. 5.3.

Parameter	Network 1	Network 2
Capacity [Bits/s]	10^6	10^6
Cell radius [m]	600	120
Sensibility [dB]	100	100
Maximum Tx Power [dBm]	43	20
Receiver gain [dB]	2	2
Transmission gain [dB]	2	2
Carrier Frequency [Hz]	2400	2400

Table 5.1: WDP simulation parameters

This network scenario depicts a real use case, where Wi-Fi can act as the small-range network, while UMTS as the medium-range network.
To prove and compare the effectiveness of the three proposed WDP distributions, we evaluated the following simulation results:

- *Cumulative Received Bits (CRBs)* i.e., the amount of bits received by the mobile terminal during the overall simulation period. The CRBs are a parameter to be maximized, in order to identify the most appropriate WDP distribution;

- *Number of Vertical Handovers (NVHOs)* i.e., the amount of vertical handover occurrences, executed by the mobile terminal during the overall simulation period. The NVHOs represent a parameter to be minimized, in order to identify the best WDP distribution.

All three probability distributions are compared against a theoretical optimum, i.e. an ideal algorithm that is able to know exactly the values GP[k], and GP[k+1]. The optimum approach can always provide a correct handover decision.
Notice that the CRBs and NVHOs are very sensitive to the simulated network scenario [45], although we can assume that the trend is almost the same in every situation. Moreover, simulation results are strongly dependent by lower layer aspects (i.e., MAC behavior or physical level impairments, like shadowing).

Fig. 5.4 shows the CRBs versus the probability threshold (i.e., PTH), which values have been evaluated via simulations. We can notice that:

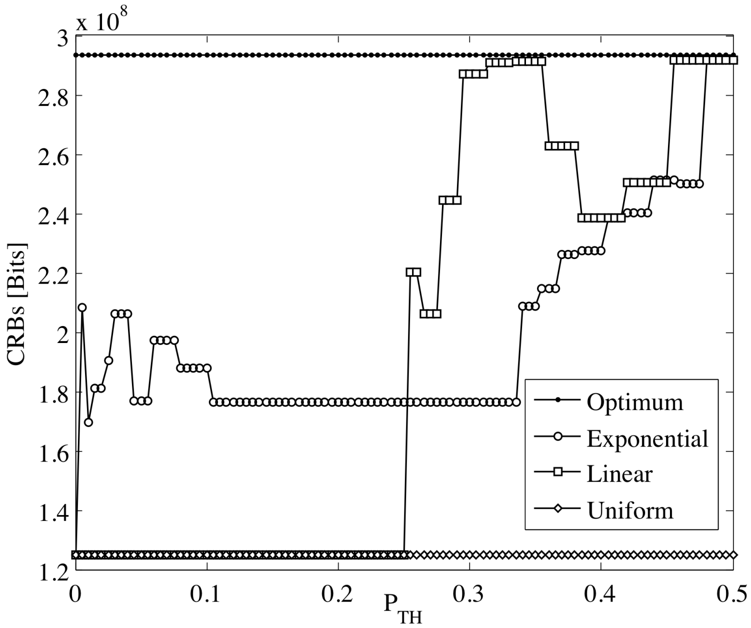

Figure 5.4: CRBs for different WDP distributions, and optimum case

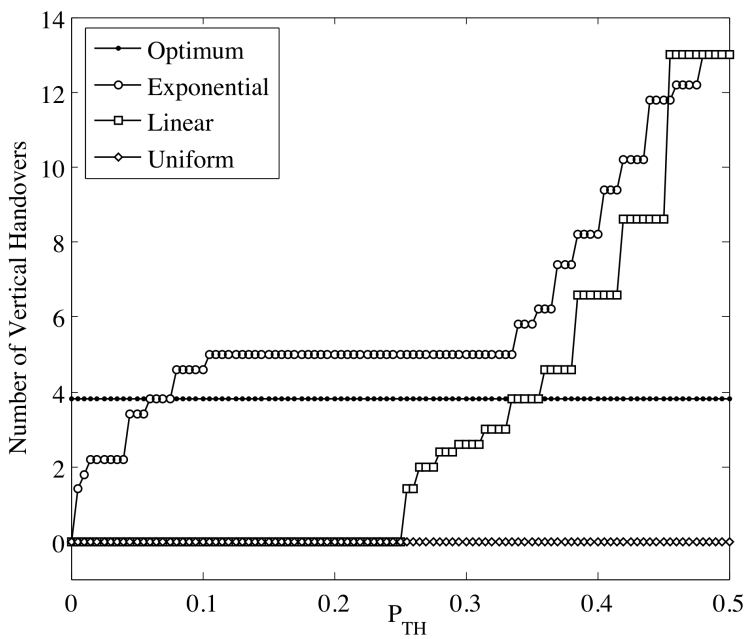

Figure 5.5: Number of VHOs for different WDP distributions, and optimum case.

- Exponential and linear approaches well approximate the optimum trend for values of threshold near the maximum, i.e. PTH = 0.5;
- Uniform approach is not strongly depending on the probability threshold, and gives very low values of CRBs. It shows that the values in k and k+1 are not totally uncorrelated, and it is possible to exploit the correlation using a probabilistic mechanism;
- Linear distribution has a rapid increase over the value PTH= 0.25. For PTH¡ 0.25 no VHO occur, and it is justified because the WDP is in the range [0.25, 0.5], as expected from the peculiar distribution;
- Exponential distribution has a rapid grow, but is slower than the linear distribution in approaching the optimum trend, due to the possibility of getting wrong decisions.

Fig. 5.5 shows the NVHOs for the three distributions, from which we evince that:

- Uniform approach does not make any vertical handover, because the WDP is always higher than the probability threshold;
- Exponential and linear approaches give almost the same results for high values of PTH;
- The number of VHOs for the optimum approach is generally lower than results obtained with exponential and linear approaches, but it gives higher cumulative received bits.

5.7 Final Remarks

A probabilistic analysis for goodput performance has been designed and evaluated, in order to optimize vertical handover procedures. Three approximations for the Wrong Decision Probability have been introduced and compared, in terms of simulation results (i.e. throughput maximization, and limitation of number of executed vertical handovers).
The linear and exponential approximations with different values of probability

thresholds achieve a good balance between throughput, and the number of vertical handovers. Linear and exponential approaches represent good solutions because they do not relay, as the optimum approach does, on the exact forecasting of the Delta-goodput process at the next time instant. To summarize, none of the approximations reaches the optimum trend, but both the uniform and the exponential approaches are close to the optimum for high value of the threshold parameters.
In the present thesis we have not considered the validation of the conditional probability distribution in a dynamic environment, where the sampling time interval for the goodput changes dynamically. Such testing may require an industrial implementation.
In the next chapter we will consider the fact that not only a single parameter is considered and that they can be affected by uncertainty.

Chapter 6

Multi-criteria decision process for vertical handover under statistical uncertainty

6.1 Introduction

Continuity of service is a key factor in next generation networks especially concerning the provision of video on demand and streaming of multimedia content. It should be supported by a plethora of different procedures including registrations, accounting, scheduling and mobility management. Among such procedures, in an heterogeneous network environment, the vertical handover should be considered, i.e. choosing the "best" network at any given time [23] considered possible different networks with different technologies.

Among the different decision criteria that can be applied to vertical handover [27] the class of Multiple attribute decision strategies (MAD) is of great interest because they give a well know mathematical background applied in different fields ranging from planning of electrical generation to project management.

Different kind of MAD can be applied to the problem of network selection, i.e. the fundamental part of vertical handover where the "best" network is chosen, such as SAW (Simple Additive Weighting), TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), AHP (Analytic Hierarchy Process) and

GRA (Grey Relational Analysis). A comparison of the previous techniques is given in [44]. where SAW and TOPSIS gives almost the same performance while GRA provides better for specific service type, i.e. interactive and background.
TOPSIS is analyzed more deeply in [8]. where the technique is evaluated considering both different services and QoS requirements.
A fundamental problem of MAD technique, hence also of TOPSIS, is called "rank abnormalities" [46]: a change of ranking can occur when one or alternatives appears or disappears from the set of valid ones. The problem of rank abnormalities that affect TOPSIS in the context of link selection is analyzed by the authors of [6] where an iterative solution approach is given.
In all the previous work the authors do not address the possible uncertainly in the input parameters that can happen due to: (i) error in the measurement process, (ii) statistical fluctuations of the input parameters. A way to handle uncertainty is through the use of fuzzy logic as the approach given in [51] and [7] In the previous work data are fuzzificated and linguistic variables are used.
The aim of this chapter is to propose a new TOPSIS method where the uncertainty is modeled through the use of probability distribution and statistic properties (decision under risk) are exploited analyze the uncertainty in the data using a probabilistic approach where some properties like measurement error are modelled and can be exploited but such approach do not permit to accurately model the variations of the input nor the measurement error.
Section 6.2 details the functional architecture (presented in [37]) that is used in the present thesis and that can support both the ordinary TOPSIS and the TOPSIS using uncertaintly modeled through the use of probability distributions (described in Section 6.3). Section 6.4 describes the numerical results.

6.2 Functional architecture

Fig. 6.1, mutuated by the one presented in [37] shows a functional architecture for the decision process that can handle uncertain parameters. The overall architecture can be decomposed into three main blocks: (i) *Handover Information Discovery* ,(ii) *Handover Decision* and (iii) *Handover Execution* each corresponding to the main phases of the vertical handover procedure.

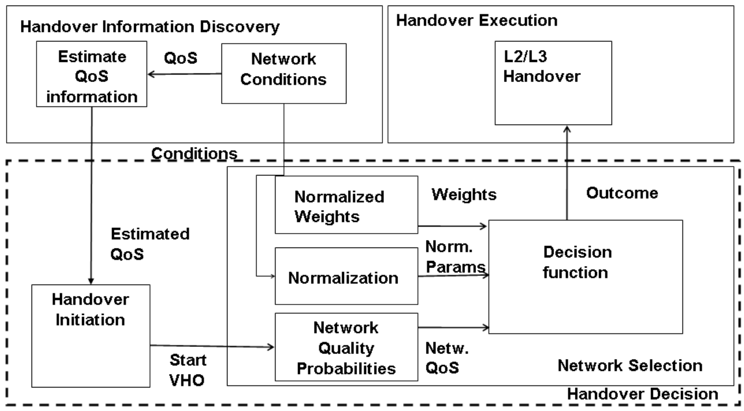

Figure 6.1: Functional Architecture

Handover Information Discovery maintains the databases concerning network conditions including both static as well as dynamic information. It is also able to assess probabilistic information concerning network quality of service.
Handover Decision it can be divided into an handover initiation and a network selection sub blocks. The former, basing on the information provided by the Handover Decision block decides if the QoS parameters have an acceptable level, if not it triggers the network selection. The former decide the "best" network basing on specified weights and a normalization procedure. It is important to hi light that a non-compensatory phase before a compensatory one may be used. The compensatory phase may include a soft or hard approach as defined in the next section.
Handover Execution includes necessary procedures for attachment to the candidate network like IP address provision and network registry. It is assumed that the procedure is always successful, if not a new decision process may be started excluding or not the unsuccessful network.
The decision algorithms that will be presented in the following sections may be implemented in the *Decision Function* block. They will be based on the input provided by the *Weights*, *Normalization* and *Network Quality Probabilities*. The

algorithms will not esplicitelly consider the interactions with *L2/L3 Handover* functional block.

6.3 Decision process

If a parameter is below a certain parameter-specific threshold service impairments may appear, e.g. interruptions in voice communication. Such situation is, generally called *out of service*.
Defined the parameter specific threshold as τ_j $j = 1, ..., M$ where M is the number of parameters the global out of service probability for a network i may be defined:

Definition 4. *Out of service* $P_{out,i}$ *the probability for a network* i *that at least one of its parameter is below a threshold* τ_j *.*

$$P_{out,i} = 1 - \prod_{j=1}^{M}(1 - P_{i,j}(x_j \leq \tau_j)) \tag{6.1}$$

where $P_{i,j}(x_j \leq \alpha)$ *indicates the probability that the parameter* j *of the network* i *is below the value of* α*.*

The next subsections define two algorithm to take into account the out of service probability. The ordinary TOPSIS is described in Appendix A where the different steps are analyzed, a graphical representation is depicted in 6.2 where the ideal and worst solution for two parameters are included.

6.3.1 TOPSIS with out of service probability

Out of service may be included in the selection matrix of TOPSIS as defined in 6.2.

$$S = \begin{pmatrix} m_{1,1} & ... & m_{1,M} & 1 - p_{out_1} \\ ... & ... & ... & ... \\ m_{N,1} & ... & m_{1,1} & 1 - p_{out_N} \end{pmatrix} \tag{6.2}$$

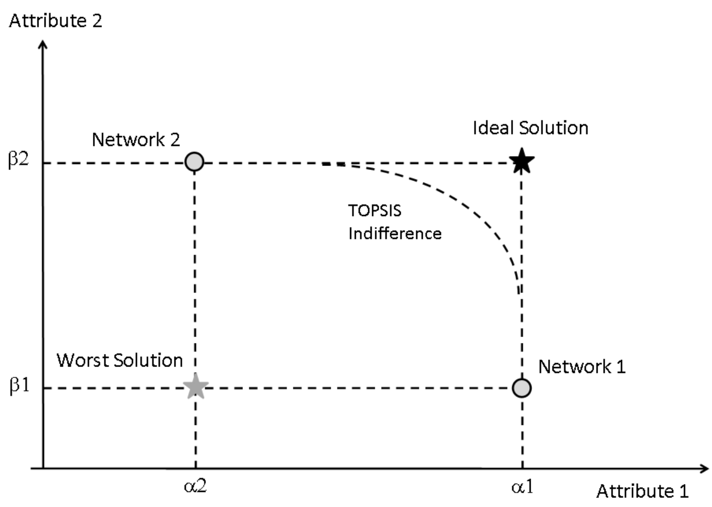

Figure 6.2: Ideal and Worst networks definition through two parameters (attributes) with a qualitative indifference curve.

In such case the $M+1$ parameter is the out of service and it shall be weighted appropriately in fact the importance of out of service parameter depends by the number of parameters but can be adjusted when the weights (w) are determined. The definition of the ideal and worst solutions can be shown in Fig. 6.2 where two networks are identified by two attributes. The ideal network is composed by the best performance parameters among networks while the worst network presents the minimum values for performance parameters among networks. To better illustrate the concept give an example.

Example 6.3.1 Two Networks, one parameter
If only one parameter and two network are considered S can be written as:

$$S = \begin{pmatrix} m_{1,1} & 1 - p_{out,1} \\ m_{2,1} & 1 - p_{out,2} \end{pmatrix}$$

if $m_{1,1} \geq m_{2,1} \wedge 1 - p_{out,2} \geq 1 - p_{out,1}$ S can be rewritten as:

$$S = \begin{pmatrix} \frac{m_{1,1}}{\sqrt{m_{1,1}^2 + m_{2,1}^2}} & \frac{1-p_{out,1}}{\sqrt{(1-p_{out,1})^2+(1-p_{out,2})^2}} \\ \frac{m_{2,1}}{\sqrt{m_{1,1}^2 + m_{2,1}^2}} & \frac{1-p_{out,2}}{\sqrt{(1-p_{out,1})^2+(1-p_{out,2})^2}} \end{pmatrix}$$

then the distances are:

$$\begin{cases} d_1^+ = \sqrt{\left(\frac{p_{out,2}-p_{out,1}}{\sqrt{(1-p_{out,1})^2+(1-p_{out,2})^2}}\right)^2} \\ d_1^- = \sqrt{\left(\frac{m_{1,1}-m_{2,1}}{\sqrt{m_{1,1}^2+m_{2,1}^2}}\right)^2} \\ d_2^+ = \sqrt{\left(\frac{m_{2,1}-m_{1,1}}{\sqrt{m_{1,1}^2+m_{2,1}^2}}\right)^2} \\ d_2^- = \sqrt{\left(\frac{p_{out,1}-p_{out,2}}{\sqrt{(1-p_{out,1})^2+(1-p_{out,2})^2}}\right)^2} \end{cases}$$

then the vicinity coefficient are:

$$\begin{cases} cc_1 = \dfrac{\sqrt{\left(\frac{m_{1,1}-m_{2,1}}{\sqrt{m_{1,1}^2+m_{2,1}^2}}\right)^2}}{\sqrt{\left(\frac{m_{1,1}-m_{2,1}}{\sqrt{m_{1,1}^2+m_{2,1}^2}}\right)^2}+\sqrt{\left(\frac{p_{out,2}-p_{out,1}}{\sqrt{(1-p_{out,1})^2+(1-p_{out,2})^2}}\right)^2}} \\ cc_2 = \dfrac{\sqrt{\left(\frac{p_{out,1}-p_{out,2}}{\sqrt{(1-p_{out,1})^2+(1-p_{out,2})^2}}\right)^2}}{\sqrt{\left(\frac{p_{out,1}-p_{out,2}}{\sqrt{(1-p_{out,1})^2+(1-p_{out,2})^2}}\right)^2}+\sqrt{\left(\frac{m_{2,1}-m_{2,1}}{\sqrt{m_{1,1}^2+m_{2,1}^2}}\right)^2}} \end{cases}$$

assuming not equals to zero all the distance the condition for the choice of network 1 vs network 2 can be written as: $d_2^+ d_1^- \geq d_1^+ d_2^-$ or explicitelly as:

$$\left|\frac{m_{2,1}-m_{1,2}}{\sqrt{m_{1,1}^2+m_{2,1}^2}}\right| \geq \left|\frac{p_{out,2}-p_{out,1}}{\sqrt{\left(1-p_{out,2}^2\right)^2+\left(1-p_{out,1}^2\right)^2}}\right| \tag{6.3}$$

6.3.2 TOPSIS under uncertainty: a soft approach

As already said each parameter may be affect by uncertainly therefore in this paragraph we illustrate an algorithm that can exploit the statistical behavior of the parameters. Since we base our work on TOPSIS the next subsection describe the modification necessary for each step of TOPSIS, namely: (i) *parameters evaluation*, (ii) *ideal and non-ideal solution* and (iii) *network selection*. The three phases are summarized in 6.3.

The next sections detail each phase hilighting the possible different choice.

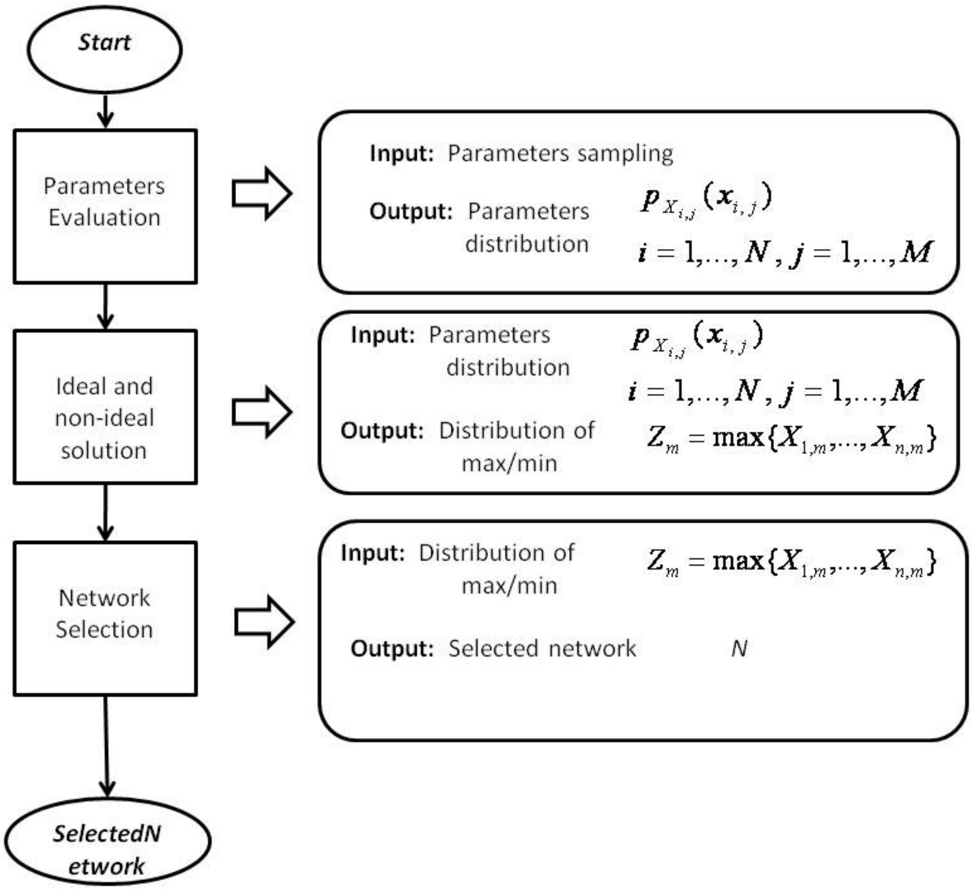

Figure 6.3: Soft TOPSIS algorithm

6.3.2.1 Parameters evaluation

Each parameter may assume values depending by a statistical description, in the following we assume that the random variable that can be associated to the j-th parameter on the i-th network is $X_{i,j}$ with its probability density function $p_{X_{(i,j)}}(x_j)$. An analogous to the selection matrix is therefore the probabilistic selection matrix S composed of n rows and m columns where each row is a network and each column is a parameter. The generic $S(i,j)$ element represents the probability distribution of the $i-th$ network for the $j-th$ parameter (represented by the random variable $X_{i,j}$

$$S = \begin{pmatrix} p_{X_{1,1}}(x_1) & ... & p_{X_{1,M}}(x_M) \\ ... & ... & ... \\ p_{X_{N,1}}(x_1) & ... & p_{X_{1,M}}(x_M) \end{pmatrix} \tag{6.4}$$

We had supposed that each probability distribution is non zero in a limited interval, such assumption represents the fact that the distribution can be a measure of a physical quantities like available bandwidth. Therefore it is possible to assume:

$$p_{X_{i,j}}(x_j) \neq 0,\ \ x_j \in [a_{i,j} b_{i,j}] \tag{6.5}$$

with $a_{i,j}$ and $b_{i,j}$ limited.

Moreover we assume that S is already a normalized matrix, i.e. $a_{i,j},\ b_{i,j} \in [0, 1]$.

6.3.2.2 Ideal and non-ideal solution

The ideal network was defined as the maximum among all the possible networks. More specifically if $Y_{\max,j}$ is the random variable associate to $j - th$ parameter and assuming all the parameters statistically independent:

$$\begin{aligned} &Y_{\max,j} = \max\{X_{1,j}, ..., X_{N,j}\} \Rightarrow \\ &P(Y_{\max,j}) = P(X_{1,j})...P(X_{N,j}) \Rightarrow \\ &p_{Y_{\max,j}}(x_j) = \sum_{i=1}^{N} p_{X_{i,j}}(x_j) \prod_{l=1, l\neq i}^{N} P(X_{l,j}) \end{aligned} \tag{6.6}$$

therefore the ideal network can be described as a multivariate random variable $Y_{\max}$

$$p_{Y_{\max}}(x_1, ..., x_M) = \{p_{\max,1}, ..., p_{\max,M}\} \tag{6.7}$$

Analogously for the non-ideal network:

$$\begin{aligned} &Y_{\min,j} = \min\{X_{1,j}, ..., X_{N,j}\} \Rightarrow \\ &P(Y_{\min,j}) = 1 - (1 - P(X_{1,j}))...(1 - P(X_{N,j})) \Rightarrow \\ &p_{Y_{\min,j}}(x_j) = \sum_{i=1}^{N} p_{X_{i,j}}(x_j) \prod_{l=1,l\neq i}^{N} (1 - P(X_{l,j})) \end{aligned} \tag{6.8}$$

therefore the ideal network can be described as a multivariate random variable $Y_{\min}$

$$p_{Y_{\min}}(x_1, ..., x_M) = \{p_{\min,1}, ..., p_{\min,M}\} \tag{6.9}$$

6.3.2.3 Network selection

The network selection can be decomposed in two phases: the first is the distance calculus and the second is the selection. Concerning the distance calculus three types of distance are considered: (i) the expected value, (ii) the Kullback-Leiber "divergence" and (iii) a modified version of the Hellinger distance.
Concerning the "*expected value distance*" the expected value of the maximum and minimum distribution is calculated and the distance to such value is used.
For the maximum we can write:

$$\begin{aligned} &d^2_{\max,i,j} = E(X_{i,j}) - E(Y_{\max,j}) \Rightarrow \\ &d_{\max,i} = \sqrt{\sum_{j=1} d^2_{\max,i,j}} \end{aligned} \tag{6.10}$$

Analogously concerning the minimum we can write:

$$\begin{aligned} &d^2_{\min,i,j} = E(X_{i,j}) - E(Y_{\min,j}) \Rightarrow \\ &d_{\min,i} = \sqrt{\sum_{j=1} d^2_{\min,i,j}} \end{aligned} \tag{6.11}$$

The selection coefficient can be written as:

$$cc_i = \frac{\sqrt{\sum_{j=1}^{M}(E(X_{i,j}) - E(Y_{\min,j}))^2}}{\sqrt{\sum_{j=1}^{M}(E(X_{i,j}) - E(Y_{\min,j}))^2} + \sqrt{\sum_{j=1}^{M}(E(X_{i,j}) - E(Y_{\max,j}))^2}} \tag{6.12}$$

We can note that the coefficient is only a function of the expected value therefore it is not possible to discern between two networks with different statistical behavior if only one parameter is considered. If two network and only one parameter is considered the following theorem applies.

Theorem 6.3.1
If two network are considered with only one parameter the distance between one network and the maximum is equal to the distance between the minimum and the other network

$$(E(X_{1,1}) - E(X_{min,1}))^2 = (E(X_{2,1}) - E(X_{max,1}))^2 \tag{6.13}$$

Dim ref. to Appendix A for a detailed demonstration.

As already said the expected value do not take into account the statistical behavior therefore a statistical "distance" can be used. We evaluate the *Kullback-Leiber divergence*, in such case the distances can be computed as:

$$\begin{cases} d_{i,j}^{+} = \int_0^1 p_{X_{i,j}}(x_j) \log\left(\frac{p_{X_{i,j}}(x_j)}{p_{X_{max,j}}(x_j)}\right) \\ d_{i,j}^{-} = \int_0^1 p_{X_{i,j}}(x_j) \log\left(\frac{p_{X_{i,j}}(x_j)}{p_{X_{min,j}}(x_j)}\right) \end{cases} \tag{6.14}$$

Under the assumption of parameter independences and using eq. 6.14 it is possible to write the total distance as:

$$\begin{cases} d_i^{+} = \sum_{j=1}^{M} d_{i,j}^{+} \\ d_i^{-} = \sum_{j=1}^{M} d_{i,j}^{-} \end{cases} \tag{6.15}$$

The two previous distances do not permit to weight more the lowest values than the highest therefore we use a modified version of the *Hellinger distance*:

$$
\begin{aligned}
d^+_{i,j} &= \sqrt{1 - \int_0^1 \sqrt{p_{X_{i,j}} p_{\max,j}}\, dx_j} \\
d^-_{i,j} &= \sqrt{1 - \int_0^1 \sqrt{(1-x_j)^k p_{X_{i,j}} p_{\max,j}}\, dx_j}
\end{aligned}
\tag{6.16}
$$

where $(1-x_j)^k$ permits to weight more the lowest values than the highest. Using the distance already defined three strategies for network selection are possible:

- *maximum from the minimum* the network with the highest distance from the minimum is chosen, i.e. $\max_i d^-_i$;
- *minimum from the maximum* the network with the lowest distance from the maximum is chosen, i.e. $\min_i d^+_i$;
- *vicinity coefficient* the network with the highest vicinity coefficient is chosen, i.e. $\max_i \frac{d^-_i}{d^+_i + d^-_i}$;

It is important to hilight that generally the three strategies do not give always the same solution.

The soft approach complexity resides in the following operations: (i) calculus of the parameters distribution and (ii) determination of the maximum/minimum distributions and distance calculus. Concerning the determination of the maximum/minimum it is possible to see that such operations have a complexity of O(NMn) where N is the number of network, M is the number of parameters and n is the number of beans that are used in the implementation to approximate the continuous distribution.

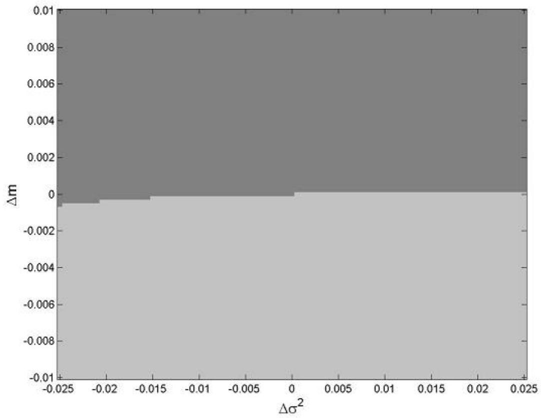

Figure 6.4: Decision for hard approach light grey=network one, otherwise network two

6.4 Numerical Results

6.4.1 Gaussian Parameter use case: two networks, one parameter

In the first scenario we examine the results in terms of outage probability and parameter performances when two networks are considered, each described by only one parameter assumed Gaussian. The first network is therefore described by a Gaussian $N(0.5, 0.075)$ while the second network is described by a Gaussian with varying parameters $N(\alpha, \beta)$ with $\alpha \in [0.49, 0.51]$ and $\beta \in [0.05, 0.1]$.
The decision using the hard approach can be seen in Fig. 6.4, it is mainly based on expected value except for low values of out-of-service probability.

The decision can be unbalanced toward the out-of-service probability using the modified Hellinger distance and different values of the parameter k. More specifically we can note that using an high value of the out-of-service is more considered.

To better understand the concept it is possible to see the values of vicinity coefficient (reported in Fig. 6.5) when the two networks have the same value of expected value. The vicinity coefficient becomes dependent by k and the out-of-service (represented by the unbalance of standard deviation) becomes sig-

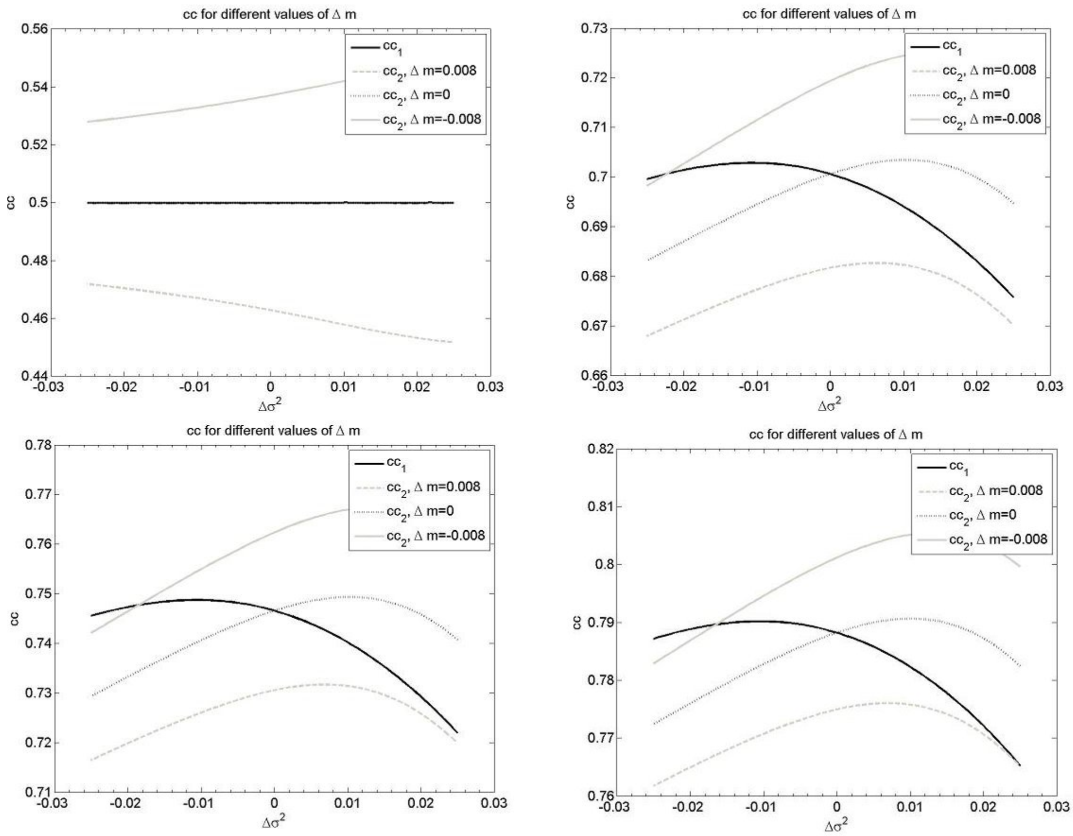

Figure 6.5: Vicinity coefficient for different k values, k=0,1,2,5

nificative.

Another way to visualize the concept is thorough the decision frontier. Fig. 6.6 shows the decision frontier for different values of k parameter using the modified Hellinger distance, it is possible to note that:

- *k=0* gives a decision based only on the maximum parameter value. The algorithm is not capable to discern between network with different second order statistics;
- *increasing k* modifies the decision frontier considering more the differences in terms of out-of-service probability. The chosen network is not always the one with the lowest out-of-service probability nor it is the one with the

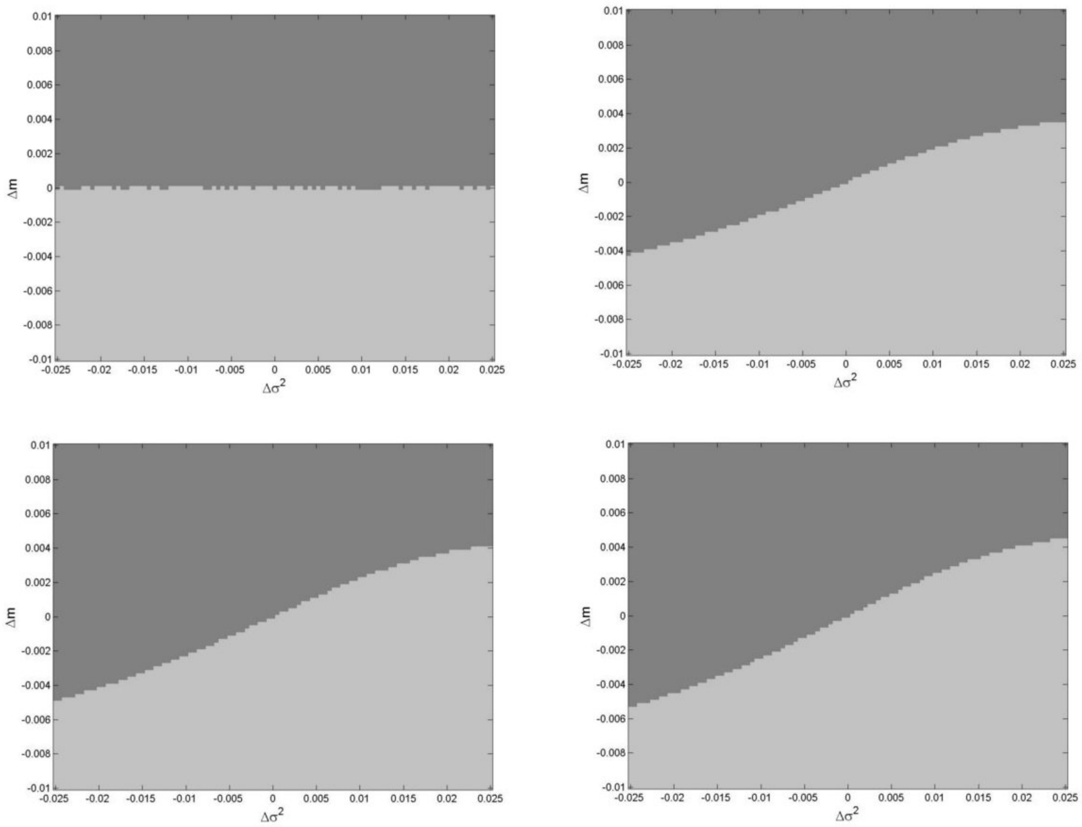

Figure 6.6: Decision for different k values, k=0,1,2,5 light grey=network one, otherwise network two

highest parameter value.

The performance of the soft (k=2) and hard approaches in terms of efficiency concerning out-of-service can be seen in 6.7 where the two approaches are compared with the choice using minimum out-of-service probability.

The soft approach permits to increase the out-of-service efficiency (from around 0.6 of hard approach to over 0.7 of the soft one) at the expense of parameters efficiency (as can be seen in 6.8 the parameter efficiency goes from around 1 of the hard approach to over 0.9 of the soft approach).
The results provided in Fig. 6.7 and Fig. 6.8 permits to assert that the soft approach (i.e. modified Hellinger distance) gives an adeguate trade-off between the outage probability and parameter efficiency.

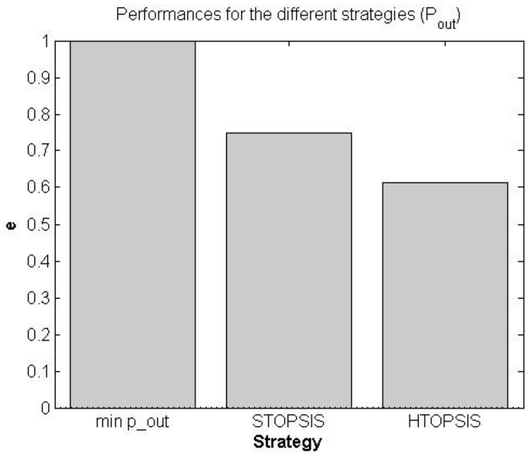

Figure 6.7: Efficiency in terms of out-of-service probability

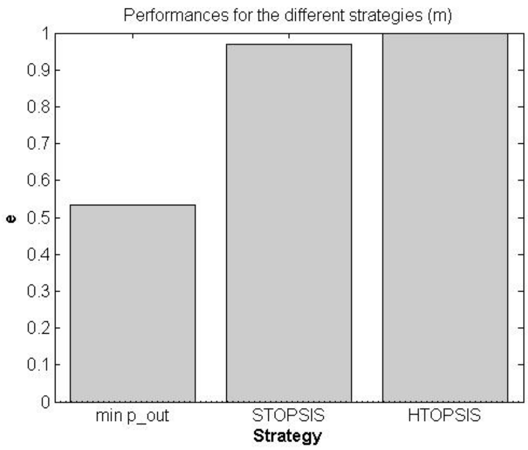

Figure 6.8: Efficiency in terms of parameter value

The value of trade-off can be adjusted by using different values of the parameter k (as can be seen by Fig. 6.9. An increase of k permits to achive a slightly decrease of out-of-service probability.

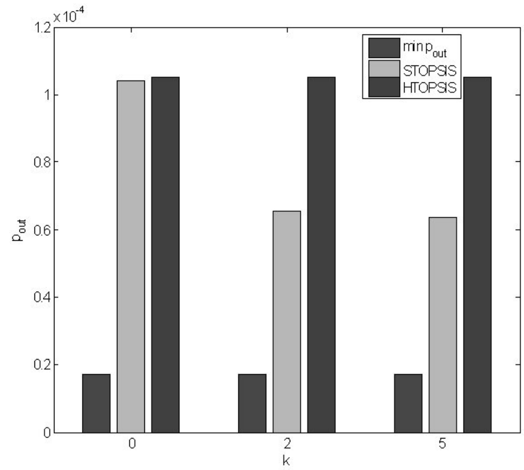

Figure 6.9: Impact of k on out-of-service-probability

6.4.2 Gaussian Parameter case study: multiple networks, multiple parameters

6.4.2.1 Parameters description and efficiency results

In the present section we analyze the algorithm behviour when multiple networks and multiple parameters are present assuming all Gaussian, i.e. the probability density function for a network i are $N_{1,i}(\alpha_{i,1}, \beta_{i,1}), ..., N_{1,m}(\alpha_{i,m}, \beta_{i,m})$ for $i = 1, ..., N$ with α and β randomly chosen in the same range of the single parameter case study, i.e. randomly chosen in the interval between 0.49 and 0.1, 0.05 and 0.1 the mean and variance respectively.
It should be noticed that using Gaussian distribution with very small variance can well approximate a pdf defined within the interval $[0, 1]$. Additionally using such values of the distributions, i.e. very near mean and variance, permits to hilight the differences in the choice by the algorithm reducing the natural imbalance between parameters.
The outage threshold is fixed for all the considered parameters considered at a value of 0.2. Generally the values can be differently for each parameters depending by the values and services.
In Fig 6.10 we can note that the soft approach using the maximum to the mini-

mum Hellinger modified distance gives the best efficiency in terms of out-of-service probability because it weights low values differently than higher.
Additionally we can note that the hard approach outperforms the two soft approaches (using Hellinger and Kullback-Leiber that gives the same results) because it weights equally the out-of-service and the other two parameters.

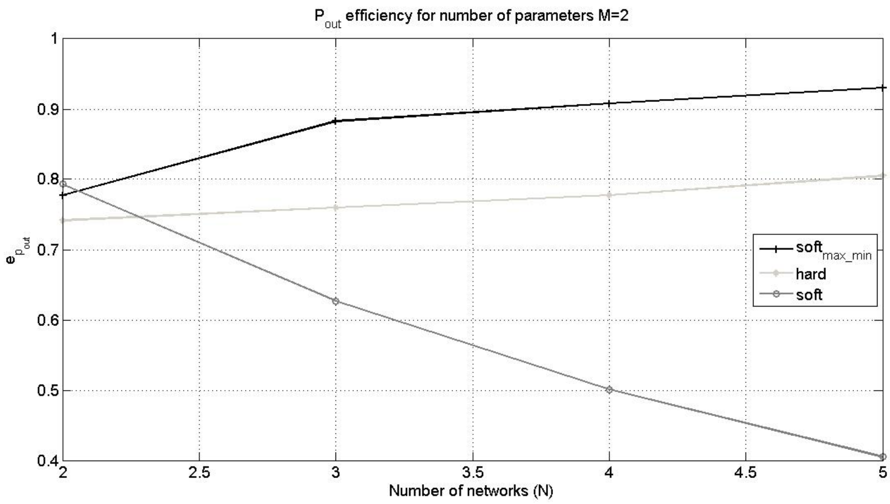

Figure 6.10: P_{out} efficiency for two parameters

We can not compare different parameters directly using a single index therefore we use the vicinity coefficient of the hard approach as the comparison parameter because it gives a Pareto efficient solution.

The results can be seen in Fig. 6.11 where the hard approach exhibit the maximum efficiency, i.e. one. The maximum from minimum distance efficiency decrease increasing the number of networks, in fact it prefers the solution with the minimum out-of-service probability. We can note that the soft approach increases its vicinity coefficient efficiency at the expense of out-of-service, anyway it is low (around 0.7 when compared with the 0.85 using the maximum from minimum approach).
An analogous behavoiur can be seen with five networks respect to the number of parameters as in Fig. 6.12.

When the number of parameters grows the hard approach becomes the best because the network have almost the same parameters and they are differentiate

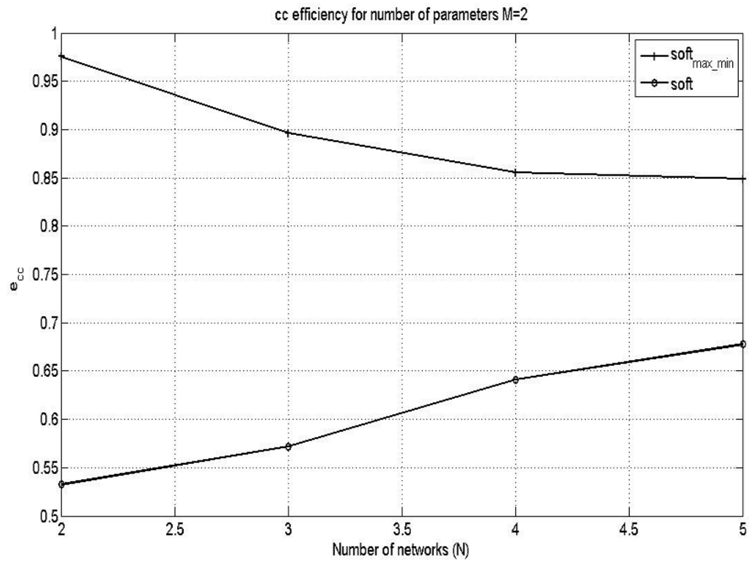

Figure 6.11: *cc* efficiency for two parameters

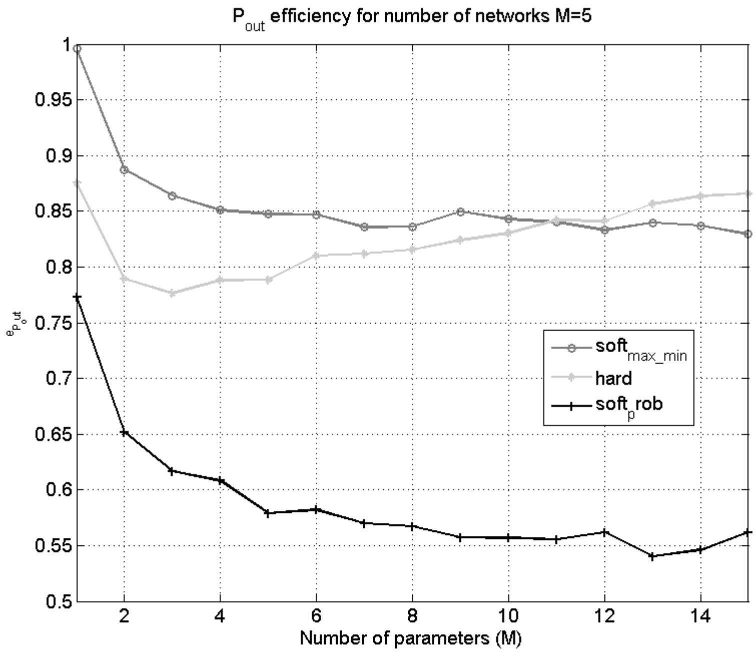

Figure 6.12: P_{out} efficiency for five parameters

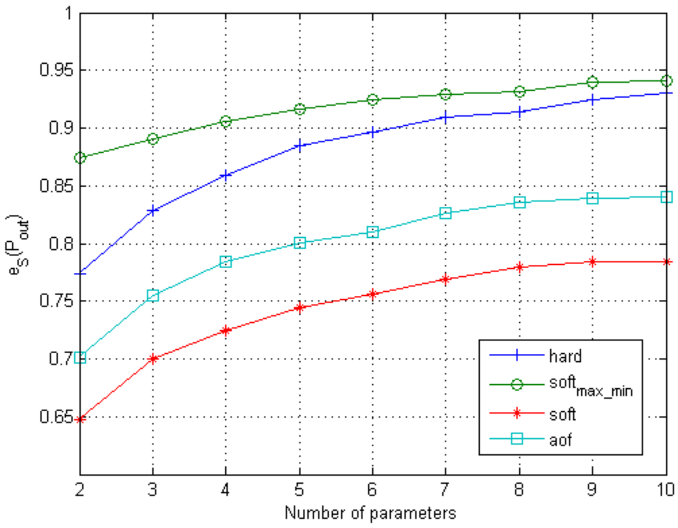

Figure 6.13: P_{out} efficiency for three different approaches, namely *soft*,$soft_{min\_max}$ and *hard*

only by the out-of-service probability.

6.4.2.2 Comparison with the single objective function approach

In the present paragraphs we compare the proposed TOPSIS with the Single Objective Function (called Aggregate Objective Function AOF in the following). The used parameters are described in the previous paragraph.
We use a different efficiency definition in order to hilight the differences between the approaches.

Fig. 6.13 shows the efficiency in terms of outage that is defined as:

$$e_{\mathcal{S}}(P_{out}) = \frac{\log P_{out}^{(\mathcal{S})}}{\log P_{out}^{(min)}} \tag{6.17}$$

where $P_{out}^{(min)}$ is the minimum outage probability and $\log P_{out}^{\mathcal{S}}$ is the outage probability of the strategy $\mathcal{S} = \{hard, soft, soft_{max\_min}, aof\}$.
The best approach is the soft one that considers the maximum distance from the minimum as a selection criterion. More specifically for low number of parameters (two) there is a difference of around 9% efficiency with the hard approach.

When the number of parameters grows, the hard approach and the $soft_{max\_min}$ tend to have almost the same performance since the outage probability tends to dominate the difference between the parameters, i.e. the networks are almost equivalent with respect to all the other parameters.
The soft approach gives low efficiency for the considered range of parameters since it is not capable to weight lower values appropriately (where the network is in outage), it grows with the number of networks because more parameters may be considered that have an effect on outage probability.
It can be noticed from Fig. 6.13 that the proposed approach ($soft_{max\_min}$) gives the best performances for a limited (i.e. less than ten) number of parameters.
It should be highlighted that generally the parameters that have a statistical behavior are QoS parameters, e.g. delay, packet drop, bandwidth,... therefore they may be considered few in numbers.
It is possible to show the sub-optimality of the soft approach through the similarity coefficient efficient in Fig. 6.14. Such efficiency is defined as:

$$e_{\mathcal{S}}(cc) = \frac{cc^{(S)}}{cc^{(H)}} \tag{6.18}$$

where cc_H is the similarity coefficient of the hard approach, $cc_{\mathcal{S}}$ is the similarity coefficient of considered strategy, i.e. $\mathcal{S} = \{hard, soft, soft_{max\_min}, aof\}$.
We remind that the similarity coefficient of the hard approach gives the solution with the best trade-off between parameters, while any other solution gives lower performance (in one or multiple parameters). Therefore, similarity coefficient of the hard approach may be considered the reference strategy for the determination of the efficiency in terms of parameters.
The efficiency of the $soft_{max_{min}}$ increases with the number of parameters because the network with the lowest outage may coincide with the network with the highest values in parameters.
The *soft* approach gives a low efficiency because its choice is not effective in the determination of the optimum, each parameter is considered separately.
When compared with the standard technique called SOF the soft approaches, i.e. *soft* and $soft_{max\_min}$, give almost the same performances as the SOF with respect to $e_{\mathcal{S}}(cc)$ while there is an increase of efficiency when compared against

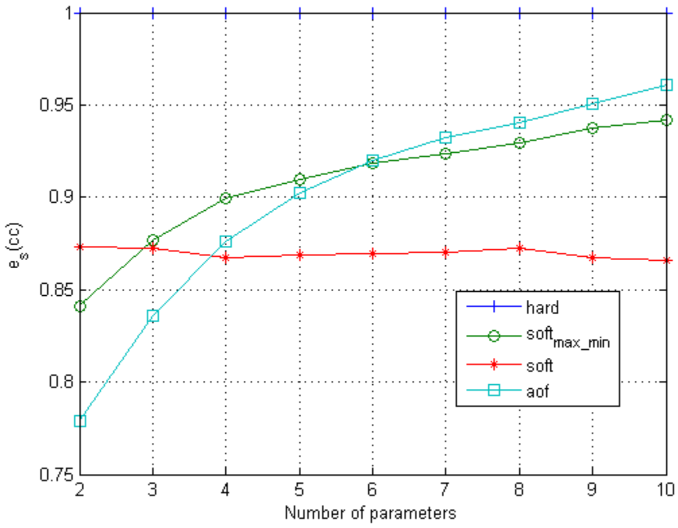

Figure 6.14: similarity coefficient efficiency for *hard*, $soft_{min\_max}$ and *soft*

??. Summarizing the $soft_{max\_min}$ permits to achieve a good trade-off between outage probability and parameter values when the number of parameters is low (less than ten) and the outage probability is more important than the absolute value of the different parameters.

6.5 Final Remarks

In this chapter we presented a novel approach to the vertical handover based on statistical information aimed at a trade-off between out-of-service and parameters value.

An hard and a soft approach have been defined basing on a standard multi-criteria decision making technique (Total order by similarity to ideal solution, TOPSIS). The soft approach using a statistical "distance", i.e. Hellinger distance, to achieve better results in terms of out-of-service probability when the distance employed is the distance from the minimum. A trade off in terms of parameters efficiency is hilighted.

The hard approach, even if not efficient as the soft, can be employed when the complexity is an issue that have great importance.

Gaussian distribution are employed to determine the performance, to permit a certain degree of freedom between the out-of-service and the value of the parameter.

Chapter 7

Conclusion

In this thesis innovative procedures for the decision process of vertical handover (VHO) have been introduced considering also possible architectures (including parameters and interactions) for their support.
We have started with the introducion of the VHO from the network operator point of view, i.e. VHO is considered a service planning problem: the network operator should choose which services are available on which network. We have formalized the problem as a linear optimization problem and have given three strategies depending by the capability of the network to operate on service flows with transcoding and depending by the trade-off between quality and number of users served that the operator wants to obtain.
We addressed the VHO problem from the mobile terminal side firstly using a single objective function for the specific problem of broadcast technologies and after extending the work considering the possibility to optimize network parameters such as when adaptive modulation and coding is possible.
The use of a single objective function may introduce an oscillation in the decision that is especially harmful in case of VHO (battery usage and service impairments may be introduced), such phenomenon is called *ping-pong* effect. We have introduced a statistical approach based on the wrong decision probability for dual mode terminals in order to limit the number of necessary handover, maximizing the received bits and maintaining a low computational complexity.
To overcome the limitation imposed by the dual mode approaches and to include statistical behaviour in the decision process we have introduced and modified an

algorithm based on a well defined multi-criteria decision process technique called Total Order by Preference to Similarity to Ideal Solution (TOPSIS).
The solution was based on the employment of statistical “distances”, namely Kullback-Leiber and Hellinger have been considered, and the usage of different decision criteria, e.g. distance from the minimim, distance from the maximum and vicinity coefficient.
Limited analytical properties have been demostrated in simplified cases, i.e. two networks and one parameters, while the performances for more networks and more parameters have been evaluated using numerical simulations.
The proposed solution outperforms existing solution in terms of out-of-service-probability with a decrease in parameters performance. There is also a limited increase in complexity in terms of number of operations and parameters that should be taken into account.

Appendix A

.1 Technique for Order Preference by Similarity to Ideal Solution

This appendix describes the TOPSIS technique applied to network selection:

1. *Define the selection matrix S*, S is composed by N rows that represent the possible networks and M colums that represent the parameters for each network.

$$S = \begin{pmatrix} m_{1,1} & m_{1,2} & ... & m_{1,M} \\ ... & ... & ... & ... \\ m_{N,1} & m_{N,2} & ... & m_{N,M} \end{pmatrix}$$

2. *Normalization of S.* A simple normalization technique

$$m'_{i,j} = \frac{m_{i,j}}{\sqrt{\sum_{j=1}^{N} m_{i,j}}}$$

3. *Weight*, each element of S' is multiplied by the appropriate weight. $(w_i \; i = 1, ..., N)$

$$m'_{i,j} = w_i * m'_{i,j}$$

4. *Ideal and non ideal solution*

$$A_j^+ = \max_i m_{i,j} \; j = 1, ..., M$$
$$A_j^- = \min_i m_{i,j} \; j = 1, ..., M$$

5. *Distance separation*

$$d_i^+ = \sqrt{\sum_{j=1} M(m_{i,j} - A_j^+)^2} \; i = 1, ..., N$$
$$d_i^- = \sqrt{\sum_{j=1} M(m_{i,j} - A_j^-)^2} \; i = 1, ..., N$$

6. *Preference*

$$cc_i = \frac{d_i^-}{/} d_i^- + d_i^+ i = 1, ..., N$$

7. *Network Selection* The chosen network is the one with the highest cc

$$n = \arg\max_i cc_i$$

.2 Properties of soft handover

n the present annex some properties that was defined in the thesis are demonstrated or analyzed in detail.

Theorem .2.1

If two network are considered with only one parameter the distance between one network and the maximum is equal to the distance between the minimum and the other network

$$d_1 - = d2^+$$
$$\text{or analogously} \tag{1}$$
$$(E(X_{1,1}) - E(X_{min,1}))^2 = (E(X_{2,1}) - E(X_{max,1}))^2$$

Dim: The $E(X_{max,1}))$ can be calculated as:

$$\begin{aligned} & E(X_{max,1}) = \int_0^1 x p_{X_{\max,1}}(x_1)\, dx_1 \Rightarrow \\ & = \int_0^1 x p_{X_{1,1}}(x_1) P_{X_{2,1}}(x_1) + p_{X_{2,1}}(x_1) P_{X_{1,1}}(x_1)\, dx_1 \end{aligned} \tag{2}$$

Using 2 we had calculated the expected value of the minimum as:

$$\begin{aligned} & E(X_{min,1}) = \int_0^1 x p_{X_{\min,1}}(x_1)\, dx_1 \Rightarrow \\ & = \int_0^1 x(p_{X_{1,1}}(x_1) + p_{X_{2,1}}(x_1) - p_{X_{1,1}}(x_1) P_{X_{2,1}}(x_1) - p_{X_{2,1}}(x_1) P_{X_{1,1}}(x_1))\, dx_1 = \Rightarrow \\ & = E(X_{1,1}) + E(X_{2,1}) - E(X_{max,1}) \end{aligned} \tag{3}$$

Then

$$\begin{aligned} & d_1^- = (E(X_{1,1}) - E((X_{min,1})))^2 = \\ & = (E(X_{1,1}) - E(X_{1,1}) + E(X_{2,1}) - E(X_{max,1}))^2 = \\ & = (E(X_{2,1}) - E(X_{max,1}))^2 = d_2^+ \end{aligned} \tag{4}$$

References

[1] 3GPP. 3gpp. www.3gpp.org, 2010. 17, 33

[2] OPEN MOBILE ALLIANCE. Architecture of the environment using the standard transcoding interface. Technical report, OMA, May 2007. 17

[3] A.M.VEGNI. *PhD Thesis*. Roma3, 2010. 55, 59

[4] P.A.A. ASSUNCAO AND M. GHANBARI. A frequency-domain video transcoder for dynamic bit-rate reduction of mpeg-2 bit streams. *Circuits and Systems for Video Technology, IEEE Transactions on*, **8**[8]:953 –967, dec. 1998. 18

[5] S. BALASUBRAMANIAM AND J. INDULSKA. Vertical handover supporting pervasive computing in future wireless networks. *Computer Communication Journal, Special Issue on 4G/Future Wireless Networks*, **0**[27/8]:708–719, 2003. 54, 55

[6] FAROOQ BARI AND VICTOR LEUNG. Multi-attribute network selection by iterative topsis for heterogeneous wireless access. In *Consumer Communications and Networking Conference, 2007. CCNC 2007. 4th IEEE*, pages 808 –812, jan. 2007. 72

[7] FAROOQ BARI AND VICTOR LEUNG. Network selection with imprecise information in heterogeneous all-ip wireless systems. In *WICON '07: Proceedings of the 3rd international conference on Wireless internet*, pages 1–10, ICST, Brussels, Belgium, Belgium, 2007. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). 72

[8] Farooq Bari and Victor C.M. Leung. Automated network selection in a heterogeneous wireless network environment. *Network, IEEE*, **21**[1]:34 –40, jan. 2007. 72

[9] Orville G et al. Brim. *Personality and Decision Processes, Studies in the Social Psychology of Thinking*. Stanford, 1962. 10

[10] Andrea Calvagna and Giuseppe Di Modica. A user-centric analysis of vertical handovers. In *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, WMASH '04, pages 137–146, New York, NY, USA, 2004. ACM. 12

[11] P.M.L. Chan, Y.F. Hu, and R.E. Sheriff. Implementation of fuzzy multiple objective decision making algorithm in a heterogeneous mobile environment. In *Wireless Communications and Networking Conference, 2002. WCNC2002. 2002 IEEE*, **1**, pages 332 – 336, mar 2002. 13

[12] Mei-Juan Chen, Ming-Chung Chu, and Shen-Yi Lo. Motion vector composition algorithm for spatial scalability in compressed video. *Consumer Electronics, IEEE Transactions on*, **47**[3]:319 –325, aug. 2001. 18

[13] Y Chen. Zipf's laws in text modeling. *International Journal of General Systems*, **0**[15]:233–252, 1989. 24

[14] C. Chi, X. Cai, R. Hao, and F. Liu. Modeling and analysis of handover algorithms. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 4473 –4477, nov. 2007. 55

[15] J. Arkko D. Johnson, C. Perkins. Rfc3375: Mobility support in ipv6. Technical report, IETF, June 2004. 8

[16] ETSI. Digital video broadcasting (dvb); ip datacast over dvb-h: Architecture. Technical Report TR 102 469 v.1.1.1, ETSI, 2004. 16, 32

[17] ETSI. Digital video broadcasting (dvb) transmission system for handheld terminals (dvb-h). Technical Report EN 302304 v1.1.1., ETSI, 2004. 16, 31

[18] ETSI. European telecommunications standards institute (etsi). http://www.etsi.org, 2010. 16

[19] S. Fernandes and A. Karmouch. Vertical mobility management architectures in wireless networks: A comprehensive survey and future directions. *Communications Surveys Tutorials, IEEE*, **PP**[99]:1 –19, 2010. 47

[20] Kai-Tat Fung, Yui-Lam Chan, and Wan-Chi Siu. Low-complexity and high-quality frame-skipping transcoder for continuous presence multipoint video conferencing. *Multimedia, IEEE Transactions on*, **6**[1]:31 – 46, feb. 2004. 18

[21] G. Gardikis, A. Kourtis, and P. Constantinou. Dynamic bandwidth allocation in dvb-t networks providing ip services. *Broadcasting, IEEE Transactions on*, **49**[3]:314 – 318, sep. 2003. 18

[22] M. Ghanbari. Two-layer coding of video signals for vbr networks. *Selected Areas in Communications, IEEE Journal on*, **7**[5]:771 –781, jun. 1989. 17

[23] E. Gustafsson and A. Jonsson. Always best connected. *Wireless Communications, IEEE*, **10**[1]:49 – 55, feb. 2003. 71

[24] T. Inzerilli and A.M. Vegni. A reactive vertical handover approach for wifi-umts dual-mode terminals. In *Consumer Electronics, 2008. ISCE 2008. IEEE International Symposium on*, pages 1 –4, april 2008. 55

[25] T. Inzerilli, A.M. Vegni, A. Neri, and R. Cusani. A location-based vertical handover algorithm for limitation of the ping-pong effect. In *Networking and Communications, 2008. WIMOB '08. IEEE International Conference on Wireless and Mobile Computing,*, pages 385 –389, oct. 2008. 55

[26] ISO/IEC. Information technologygeneric coding of moving pictures and associated audio informationpart 2: Video. Technical report, ISO/IEC, 1995. 17

[27] Meriem Kassar, Brigitte Kervella, and Guy Pujolle. An overview of vertical handover decision strategies in heterogeneous wireless networks.

In *Comput. Commun.*, **31**, pages 2607–2620, Newton, MA, USA, 2008. Butterworth-Heinemann. 11, 71

[28] EunHae Kim, SuJung Yu, SungMin Yoon, and JooSeok Song. An efficient paging message scheme for vertical handoff decision in convergence of telecommunication and broadcasting. In *Broadband Convergence Networks, 2006. BcN 2006. The 1st International Workshop on*, pages 1 –5, april 2006. 33, 34

[29] Jae-Won Kim, Goo-Rak Kwon, Nam-Hyeong Kim, A. Morales, and Sung-Jae Ko. Efficient video transcoding technique for qos-based home gateway service. *Consumer Electronics, IEEE Transactions on*, **52**[1]:129 – 137, feb. 2006. 18

[30] Won-Ik Kim, Bong-Ju Lee, Jae-Su Song, Yeon-Seung Shin, and Yeong-Jin Kim. Ping-pong avoidance algorithm for vertical handover in wireless overlay networks. In *Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007 IEEE 66th*, pages 1509 –1512, oct. 2007. 55

[31] R. Koodli. Rfc5568: Mobile ipv6 fast handovers. Technical report, IETF, July 2009. 8

[32] Shyam Lal and D.K. Panwar. Coverage analysis of handoff algorithm with adaptive hysteresis margin. In *Information Technology, (ICIT 2007). 10th International Conference on*, pages 133 –138, dec. 2007. 56

[33] Hanjin Lee, Dongwook Kim, Byungchun Chung, and Hyunsoo Yoon. Adaptive hysteresis using mobility correlation for fast handover. *Communications Letters, IEEE*, **12**[2]:152 –154, feb. 2008. 55

[34] N. Nasser, A. Hasswa, and H. Hassanein. Handoffs in fourth generation heterogeneous networks. In *Communications Magazine, IEEE*, **44**, pages 96 –103, oct. 2006. 5

[35] V. Ollikainen and Chenyuan Peng. A handover approach to dvb-h services. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 629 –632, july 2006. 33, 34

[36] OMA. Open mobile alliance. www.openmobilealliance.org, 2010. 17

[37] Eng Hwee Ong and J.Y. Khan. On optimal network selection in a dynamic multi-rat environment. *Communications Letters, IEEE*, **14**[3]:217 –219, march 2010. 72

[38] C. Perkins. Rfc3344: Ip mobility support for ipv4. Technical report, IETF, August 2002. 8

[39] Daji Qiao, Sunghyun Choi, and K.G. Shin. Goodput analysis and link adaptation for ieee 802.11a wireless lans. *Mobile Computing, IEEE Transactions on*, **1**[4]:278 – 292, oct-dec 2002. 45

[40] Thomas L. Saaty. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation (Decision Making Series)*. Mcgraw-Hill, 1980. 13

[41] R. Shacham, H. Schulzrinne, S. Thakolsri, and W. Kellerer. Rfc 5631: Session initiation protocol (sip) session mobility. Technical report, IETF, October 2009. 9

[42] Bo Shen, I.K. Sethi, and B. Vasudev. Adaptive motion-vector resampling for compressed video downscaling. *Circuits and Systems for Video Technology, IEEE Transactions on*, **9**[6]:929 –936, sep. 1999. 18

[43] H. Soliman, Flarion, C. Castelluccia, K. El Malki, and L. Bellier. Rfc4140: Hierarchical mobile ipv6 mobility management (hmipv6). Technical report, IETF, August 2005. 8

[44] E. Stevens-Navarro and V.W.S. Wong. Comparison between vertical handoff decision algorithms for heterogeneous wireless networks. In *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, **2**, pages 947 –951, may 2006. 72

[45] G. Tamea, A.M. Vegni, T. Inzerilli, and R. Cusani. A probability based vertical handover approach to prevent ping-pong effect. In *Wireless Communication Systems, 2009. ISWCS 2009. 6th International Symposium on*, pages 181 –185, sept. 2009. 47, 55, 56, 57, 59, 61, 65, 66, 67

[46] E. Triantaphyllou. Multi-criteria decision making methods: A comparative study. *Applied Optimization*, **44**, 2000. 72

[47] N. Vulic, S.M.H. de Groot, and I.G.M.M. Niemegeers. Dvb-h - umts integration at radio access level. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 1250 –1254, april 2007. 33

[48] Shun-Ren Yang, Sheng-Ying Yan, and Hui-Nien Hung. Modeling umts power saving with bursty packet data traffic. *Mobile Computing, IEEE Transactions on*, **6**[12]:1398 –1409, dec. 2007. 37

[49] Sang-Jo Yoo, David Cypher, and Nada Golmie. Lms predictive link triggering for seamless handovers in heterogeneous wireless networks. In *Military Communications Conference, 2007. MILCOM 2007. IEEE*, pages 1 –7, oct. 2007. 34

[50] F. Zarai, I. Smaoui, L. Kamoun, and M.M. Banat. Reducing handoff blocking probability in fourth generation wireless networks. In *Electronics, Circuits, and Systems, 2009. ICECS 2009. 16th IEEE International Conference on*, pages 811 –815, dec. 2009. 56

[51] Wenhui Zhang. Handover decision using fuzzy madm in heterogeneous networks. In *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, **2**, pages 653 – 658 Vol.2, march 2004. 72

[52] Yan Zhang. Handoff probability in wireless networks over rayleigh fading channel: A cross-layer approach. In *Communications, 2007. ICC '07. IEEE International Conference on*, pages 5660 –5664, june 2007. 56

[53] Yan Zhang, Z.H. Shao, and M. Fujise. Handoff probability in wireless mobile networks. In *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, **5**, pages 3407 – 3411 Vol. 5, May 2005. 56

[54] Huamin Zhu and Kyung sup Kwak. Adaptive handoff using distance information. In *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, **2**, pages 932 –936, may 2006. 55