

An overview of vertical handover decision strategies in heterogeneous wireless networks

Meriem Kassar*, Brigitte Kervella, Guy Pujolle

University Pierre & Marie Curie – Paris 6, UMR 7606, LIP6, 104, Avenue du Président Kennedy, 75016 Paris, France

Received 5 June 2007; received in revised form 14 January 2008; accepted 15 January 2008

Available online 31 January 2008

Abstract

In the next generation of wireless networks, mobile users can move between heterogeneous networks, using terminals with multiple access interfaces and non-real-time or real-time services. The most important issue in such environment is the *Always Best Connected* (ABC) concept allowing the best connectivity to applications anywhere at anytime. To answer ABC requirement, various vertical handover decision strategies have been proposed in the literature recently, using advanced tools and proven concepts. In this paper, we give an overview of the most interesting and recent strategies. We classify it into five categories for which we present their main characteristics. We also compare each one with the others in order to introduce our vertical handover decision approach.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Vertical handover; Heterogeneous wireless networks; Handover decision; Strategy; Network selection

1. Introduction

The deployment of various wireless technologies (2G, 3G, WLAN, WMAN, etc.) in combination with the evolution of Mobile Terminals (MTs) with multiple network interfaces and the development of IP-based applications (non-real-time or real-time) has allowed the user to have access to IP services anywhere at anytime from any network. One revolutionary step, driven by this universal wireless access, is the fourth Generation (4G) of wireless communications [1]. This next-generation of wireless systems represents a heterogeneous environment with different access networks technologies that differ in bandwidth, latency or cost. In this kind of environment, mobility management is the essential issue that supports the roaming of users from one system to another. Handover management, one of the mobility management components, controls the

change of the MT's point of attachment during active communication [2].

Handover management issues include mobility scenarios, metrics, decision algorithms and procedures. Mobility scenarios can be classified into horizontal (between different cells of the same network) and vertical (between different types of networks). In homogeneous networks, horizontal handovers are typically required when the serving access router becomes unavailable due to MT's movement. In heterogeneous networks, the need for vertical handovers can be initiated for convenience rather than connectivity reasons (e.g., according to user choice for a particular service). Two of the major challenges in vertical handover management are seamlessness and automation aspects in network switching. These particular requirements can refer to the *Always Best Connected* concept, of being connected in the best possible way in an environment of multiple access technologies, according to policies (expressed by rules based on parameters such as network conditions or user preferences) [3]. For that, a handover management technique must choose the appropriate time to initiate the handover and the most suitable access

* Corresponding author. Tel.: +33 144278843.

E-mail addresses: Meriem.Kassar@lip6.fr (M. Kassar), Brigitte.Kervella@lip6.fr (B. Kervella), Guy.Pujolle@lip6.fr (G. Pujolle).

network for a specific service among those available, and must maintain service continuity.

This paper presents the vertical handover management and focuses on the handover decision problem. It is interesting to situate the decision phase in the global process and to prove its contributions in the optimization of vertical handover performance. The vertical handover decision process answers when and where to hand over in a heterogeneous environment. The first choice can minimize for instance the signaling overhead and avoid unnecessary handovers. The second choice can satisfy network and user requirements. This process needs decision factors: decision criteria, policies, algorithms, control scheme, etc. Decision criteria include user preferences, network conditions, application requirements and terminal capabilities. These have to be evaluated and compared to detect and to trigger a vertical handover. For that, we explore many methodologies used: policy-enabled scheme, Fuzzy Logic and Neural Networks concepts, advanced algorithms such as multiple attribute decision making, context-aware concept, etc.

In the context of future wireless networks, many analysis, studies and tutorials were proposed in the literature: mobility management solutions [2], vertical handover design in 4G context [1], handover in hybrid mobile data networks [4], etc. No one was proposed including the different existing strategies in the vertical handover decision problem. So, our paper analyzes the most interesting and recent ones in the literature. We show how advanced tools as well as proven concepts can be used to solve such a problem and thus answering ABC requirement. We classify the strategies into five main categories: function-based, user-centric, multiple attribute decision, Fuzzy Logic and Neural Networks based, and context-aware strategies.

The paper is organized as follows. Section 2 presents the handover management process in a heterogeneous environment. Section 3 describes the vertical handover decision

and its characteristics taken into account in our study. Section 4 analyzes the different existing strategies and compares each one with the others. Section 5 proposes our vertical handover decision strategy. Finally, Section 6 concludes our work and gives some perspectives.

2. Handover management in heterogeneous wireless networks

Handover management is the key aspect in the development of solutions supporting mobility scenarios. It is the process by which MT maintains its connection active while moving from one point of attachment (base station or access router) to another. In this section, we describe the handover process features and we provide the motivation for analyzing the vertical handover decision problem in a heterogeneous environment. Fig. 1 shows the handover management concept features: mobility scenarios, handover process, types, control, and performance. The highlighted grey box represents the features closely related to the handover decision issue. It is explained in Section 2.2 and described in Sections 3 and 4.

2.1. Handover management process

Many works describe the handover process in three phases [5,6]:

- *Handover Information Gathering*: used to collect all the information required to identify the need for handover and can subsequently initiate it. It can be called also *handover initiation* phase or *system discovery*.
- *Handover Decision*: used to determine whether and how to perform the handover by selecting the most suitable access network (taking into account some criteria such as user preferences) and by giving instructions to the execution phase. It is also called *network* or *system selection*.

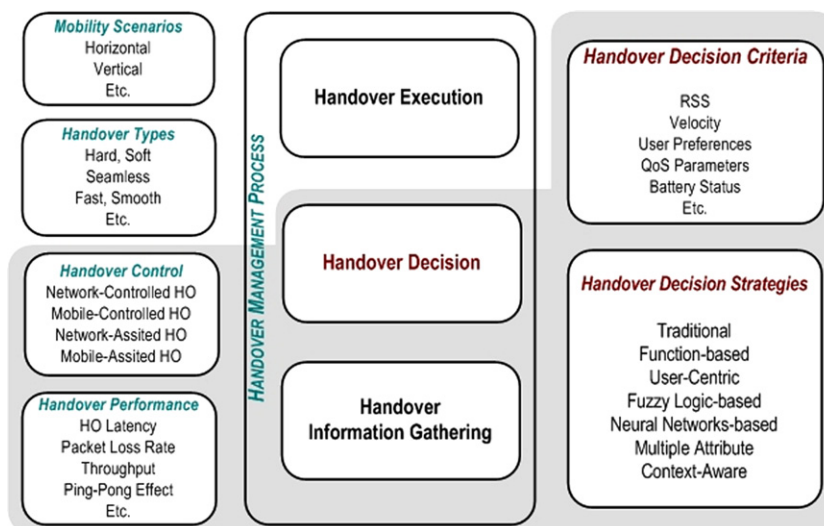


Fig. 1. Handover management concept.

- *Handover Execution*: used to change channels conforming to the details resolved during the decision phase.

The handover procedure can be characterized in various types. On one hand, the handover can be *hard* when the MT is connected to only one point of attachment at a time. It is referred to as a *break before make* handover. On the other hand, it can be *soft* when the MT is connected to two points of attachment for a while and it is referred to as *make before break* handover. For achieving seamlessness aspect in mobility scenarios, the handover has to be *seamless*. It means that the transition to the new network point of attachment is transparent to the user (no perceptible service degradation). So, it is the one that performs a *fast* handover (minimal handover latency) and a *smooth* handover (minimal packet loss).

Regardless of handover types, the *handover process control* [7] or the handover decision mechanism can be located in a network entity or in the MT itself. The handover decision usually involves some sort of measurements and information about when and where to perform handover and obtained from one entity or both. So, in *Network-Controlled HandOver (NCHO)*, the network entity has the primary control over the handover. In *Mobile-Controlled HandOver (MCHO)*, the MT must take its own measurement and make the handover decision on its own. When information and measurements from the MT are used by the network to decide, it is referred to as *Mobile-Assisted HandOver (MAHO)* like in GSM. When the network collects information that can be used by the MT in a handover decision, it is a *Network-Assisted HandOver (NAHO)*.

Among handover management solutions given at [2], one of the most popular scheme is Mobile IP [8], an IP layer mobility management protocol. This protocol is in charge of redirecting packets sent by a CN (Correspondent Node) to the MT or MN (Mobile Node) to its current location. It introduces mobility agents: a HA (Home Agent) and a FA (Foreign Agent). In Mobile IP, the handover procedure is carried out by these principles steps [2]:

- MN detects whether it has moved to a new access network by receiving or sending messages from or to mobility agents. This step is known as *agent discovery*.
- MN obtains a new temporary address, CoA (Care-of-Address) when it enters a new access network. MN registers the new CoA with its Home Agent, which sets up a new tunnel up to the end point of the new CoA and removes the tunnel to the old CoA. This step is known as *registration*.
- Once the new tunnel is set up, the HA tunnels packets destined to the MN using the MN's new CoA. This step is known as *routing and tunneling*.

In each type of wireless access network, we can find most of these handover process features. The availability

of different wireless access networks makes the handover decision problem an important issue to study in order to improve the vertical handover process.

2.2. Motivation for vertical handover decision issue

In a homogeneous environment, handover decision is considered as handover initiation step. It represents the decision of initiating or not the handover, and over which cell to hand over in cellular networks for instance. Traditionally, it is pointed out that the need for horizontal handover arises when the Signal Strength (SS) of the serving Base Station (BS) deteriorates below a certain threshold value.

In a heterogeneous environment, users can move between different access networks. They will benefit from different network characteristics (coverage, bandwidth, latency, power consumption, cost, etc.) that can not be compared directly. The handover process shown in Fig. 1 becomes more complex in such an environment compared to the homogeneous one.

Therefore, the more challenging problem is the handover decision and resolving it can influence the handover performance. It is referred to vertical handover decision which needs more criteria (not only SS) compared to horizontal handover. In our paper, we mean by *handover initiation* the decision for the appropriate moment to initiate the handover and by *network selection* the decision for the most suitable access network. The first choice can minimize for instance the signaling overhead, avoid unnecessary handovers and predict disconnections. The second choice must satisfy network and user requirements by deciding over which access network to connect at any point of time when many access networks are available for a specific service.

Thus, in each handover decision making process, we can face the following questions: *How does vertical handover decision process works? What are the handover decision criteria used? How are gathered the needed criteria? What are the handover decision policies applied? Who (terminal or network) is taking the decision? What are the handover performance optimizations that can be made?*

In our vision, the main issues are: combining decision criteria, comparing them and answering user needs anytime and anywhere. So, we must identify the decision factors (criteria, policy, strategies, etc.) and we must use them in order to optimize handover performance (e.g., throughput, handover delay).

In the following sections, we define the vertical handover decision problem, its characteristics and the different strategies proposed in the literature to resolve it.

3. Vertical handover decision

In this section, we give the main characteristics needed in a vertical handover decision strategy.

Beside the handover classification, given in the previous section, it could be interesting to introduce a more general classification according to initiation reasons for a vertical handover decision [9,10]:

- *Imperative or forced handover*: triggered by physical events regarding network interfaces availability.
- *Alternative or user handover*: triggered by user policies and preferences.

Although, rules have to be fixed for how and when to trigger the handover. These rules design a handover decision policy and use policy parameters, called metrics or decision criteria (monetary cost, QoS, power requirements, etc.). Thus, a vertical handover decision strategy should decide when to trigger the handover procedure, select and switch seamlessly to the most optimal access network from those available.

3.1. Handover decision criteria

Handover criteria are the qualities that are measured to give an indication of whether or not a handover is needed. We can regroup different criteria as follows:

- *Network-related*: coverage, bandwidth, latency, link quality (RSS (Received Signal Strength), CIR (Carrier-to-Interferences Ratio), SIR (Signal-to-Interferences Ratio), BER (Bit Error Rate), etc.), monetary cost, security level, etc.
- *Terminal-related*: velocity, battery power, location information, etc.
- *User-related*: user profile and preferences.
- *Service-related*: service capabilities, QoS, etc.

These criteria can be classified into static and dynamic depending on the frequency and causes of changes. Typically static criteria are user profile and the cost of the different access networks, whereas the MT's velocity and RSS are typically dynamic criteria.

3.2. Handover decision policy

Handover decision criteria help to determine which access network should be chosen and the handover decision policy represents the influence of the network on when and where the handover occurs. The traditional handover decision policy is based only on RSS [4,11]:

- *RSS*: choosing the new Base Station (BS) if $RSS_{new} > RSS_{old}$.
- *RSS with Threshold T* : choosing the new BS if $RSS_{new} > RSS_{old}$ and $RSS_{old} < T$.
- *RSS with Hysteresis H* : choosing the new BS if $RSS_{new} > RSS_{old} + H$.
- *RSS, Hysteresis and Threshold*: choosing the new BS if $RSS_{new} > RSS_{old} + H$ and $RSS_{old} < T$.

- *Algorithm with Dwell timer*: when starting the Dwell timer the condition is true. Handover is performed if the condition continues to be true until the timer expires.

In heterogeneous networks, vertical handover decision policy must evaluate additional criteria such as monetary cost, offered services, network conditions, terminal capabilities and user preferences. More criteria are needed not only for the decision of the appropriate time to perform the handover but also for user choice and intervention (user preferences among different access technologies) [1]. It is obvious to mention that the combination of all these criteria and the dynamicity of some of them will increase significantly the complexity of the vertical handover decision process. In general, this complex problem can be seen as: a user-centric (user satisfaction) problem, a context-aware (user, network and terminal context) problem, a network selection algorithm, or even a multi-criteria algorithm. Based on these different views of the handover decision problem, we establish a classification of the existing strategies, described in the next section.

4. Vertical handover decision strategies

In this section, we introduce a panel of the most well designed vertical handover decision strategies proposed in the literature. We distinguish five categories: functions, user-centric, Fuzzy Logic and Neural Network-based, multi-criteria, and context-aware strategies.

4.1. Decision function-based strategies (DF)

Vertical handover decision cost function is a measurement of the benefit obtained by handing over to a particular network. It is evaluated for each network n that covers the service area of a user. It is a sum of weighted functions of specific parameters. The general form of the cost function f_n of wireless network n is [1]:

$$f_n = \sum_s \sum_i w_{s,i} \cdot p_{s,i}^{n,i} \quad (1)$$

$p_{s,i}^{n,i}$: the cost in the i th parameter to carry out service s on network n ; $w_{s,i}$: the weight (importance) assigned to using the i th parameter to perform services (with $\sum_i w_i = 1$).

The first policy-enabled handover strategy was proposed in 1999 in [12], which introduced the cost function to select the best available network in the decision making. The parameters used are bandwidth B_n that network n can offer, power consumption P_n of using the network device for n and monetary cost C_n of n . The cost of using a network n at a certain time, with $N(i)$ as the normalization function of parameter i is defined as:

$$f_n = w_b \cdot N(1/B_n) + w_p \cdot N(P_n) + w_c \cdot N(C_n) \quad (2)$$

The network that is consistently calculated to have the lowest cost is chosen as the target network. Therefore, this cost function-based policy model estimates dynamic network conditions and includes a stability period (a waiting period

before handovers) to ensure that a handover is worthwhile for each mobile.

The proposed policy-enabled handover system allows users to express policies on what is the best network and when to hand over. The system operating environment is a Mobile IP infrastructure (see Section 2.1) in which all the handover decisions and operations are done at the MT. In handover operation, the packets sent by the CN to the MN go through its HA. The HA routes the packets either to the multicast CoA (a group of WLANs) or the unicast CoA of the MN. When MN is in WLANs, a reverse tunneling is used where packets are routed to the HA first then to the CN. To achieve flexibility, the system separates the decision making scheme from the handover mechanism (routing table manipulation and sending location updates). To achieve seamlessness, the system considers user involvement (for policy specification) with minimal user interaction (for automation). To improve system stability in the handover mechanism, load balancing solution is proposed avoiding the handover synchronization problem (simultaneous decision by many mobiles). For that, the authors implemented a performance agent that collects the information on the current bandwidth usage at base stations, and periodically announces this information to its coverage. Since, all data traffic goes through base stations, they have the most accurate information on current bandwidth usage and the available bandwidth in the network. They solve the problem through a randomized stability period.

A number of papers use similar cost functions in the handover decision process. Chen et al. [13] proposes an adaptive scheme based on handover decision process described in [12]. The authors use the utility function (higher utility = target network), to evaluate the reachable wireless networks discovered (bandwidth and movement speed as factors) and to quantify the QoS (Quality of Service) provided by the wireless network on the MT. They introduce two adaptive handover decision methods adjusting the stability period defined in [12], according to the network resources and the running applications on the MT: one measures several utility ratios ($U_{\text{target}}/U_{\text{current}}$) and one relies on the ratio of two measured utility ratios. These two methods avoid unnecessary handovers and make the process faster comparing to the non-adaptive stability period method [12]. In the proposed handover scheme, the handover decision method is preceded by a system discovery method. The latter is based on an adaptive interface activating method that adjusts the interface activating interval relying on the distance between the MT and the base station. For that, an ideal coverage concept (i.e., the real coverage in a wireless overlay network) is introduced in which MT's position information and a Location Service Server can assist MT in deciding when to activate its interfaces. Thus, the system discovery method can balance the power consumption and system discovery time.

In [14], an optimized cost function is used to evaluate the target network (based on QoS factor) establishing a tradeoff between user satisfaction (gains in QoS) and net-

work efficiency. It is applied on two vertical handover policies, one for all the user's active sessions collectively (handed over the same target network) and one for each of the user's active sessions independently (with prioritization). Elimination constraints (RSS and channel availability) and a prioritized multi-network scheme were introduced to, respectively, reduce the delay and improve the throughput for an MT with multiple active sessions. The proposed vertical handover decision is based on a policy-based networking architecture (i.e., IETF framework).

All the described decision strategies were evaluated on two types of networks: WLAN and WWAN such as GSM [12] or GPRS [13,14].

Compared to the strategy in [14,12] and [13] do not evaluate user satisfaction, which is the main purpose of user-centric strategies presented in the following section, with more advanced functions.

4.2. User-centric strategies (UC)

Among the different criteria that a vertical handover decision takes into account, user preferences, in terms of cost and QoS, is the most interesting policy parameter for a user-centric strategy. In [15], a model is proposed based on a handover decision evaluated, from the user point of view, as the most convenient handover to his specific needs (cost and QoS). The authors propose two handover decision policies (fixing a threshold value) between GPRS and WiFi networks: (1) the MT will never abandon GPRS connection without connection blackouts and (2) the algorithm searches for just WiFi access points with connection blackouts. The first one will satisfy that user who is willing to pay for having its connections as granted as possible. The second one will satisfy the user from the connection cost point of view but will disappoint his expectation of QoS. Based on these policies, the simulation results have shown by varying the handover decision policy, the performance of some applications running on the user terminal (FTP, HTTP, and Telnet) improves whereas others becomes worse. In order to find the optimum handover decision policy maximizing this performance, they define a cost function as follows:

$$C = T_{\text{WiFi}} \cdot c_{\text{WiFi}}(h) + T_{\text{GPRS}} \cdot c_{\text{GPRS}}(h) \quad (3)$$

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.

It proves that the willingness to pay expressed by the user can be satisfied when adopting a suitable handover decision policy.

The decision scheme is integrated into a network selection process module. This module is in charge, on one hand, of retrieving periodically data from the network monitoring

module (i.e., actual network availability conditions) and, on the other hand, of getting user preferences specifies through the user profile management module. The authors implemented the proposed user-centric model integrating a Mobile IP-like distributed mobility protocol to support the roaming of MNs in the WiFi and GPRS domain.

In [16], Ormond et al. analyze only user satisfaction by using a utility function for non-real-time data services (FTP file transfer). The network selection decision algorithm is based on consumer surplus value (difference between the monetary value of the data transferred and the actual price charged) with predicting the transfer completion time. Thus, if the price that the user pays for the transfer is less than the value they were willing to pay it is interesting for the user to save money. In order to choose the appropriate utility function, the decision metrics are user's risk attitudes: neutral (user prefers equally paying less to experiencing less delay), seeking (user prefers alternative of less delay to assured money saving) and adverse (user prefers to be certain of paying less). The used simulation model has a topology of a wired network (hosts the sink for all application data) connected to two WLAN APs with a multihomed terminal integrating the CS-based network selection strategy. The test results have shown that the value of employing an appropriate utility function for a given user preference is highly dependent of the file size.

The described user-centric functions propose handover decision policies and criteria mainly for user satisfaction and non-real-time applications. Deciding for the most appropriate network that answers user satisfaction and network efficiency, more criteria to retrieve from the different available networks and more advanced techniques have to be considered.

4.3. Multiple attribute decision strategies (MAD)

The handover decision problem deals with making selection among limited number of candidate networks from various service providers and technologies with respect to different criteria. This is a typical MADM (Multiple Attribute Decision Making) problem. In the study of decision making, terms such as multiple objective, multiple attribute and multiple criteria are often used interchangeably. Distinctions can be made between the different concepts [17]: Multiple Criteria Decision Making (MCDM) is sometimes applied to decisions involving multiple objectives or multiple attributes, but generally when they both apply. Multiple Objective Decision Making (MODM) consists of a set of conflicting goals that cannot be achieved simultaneously. Multiple Attribute Decision Making (MADM) deals with the problem of choosing an alternative from a set of alternatives which are characterized in terms of their attributes. The most popular classical MADM methods are:

- (1) *SAW* (Simple Additive Weighting): the overall score of a candidate network is determined by the weighted sum of all the attribute values.

- (2) *TOPSIS* (Technique for Order Preference by Similarity to Ideal Solution): the chosen candidate network is the one which is the closest to ideal solution and the farthest from the worst case solution.
- (3) *AHP* (Analytic Hierarchy Process): decomposes the network selection problem into several sub-problems and assigns a weight value for each sub-problem.
- (4) *GRA* (Grey Relational Analysis) is then used to rank the candidate networks and selects the one with the highest ranking.

A comparison of three of these models was established in [6] with attributes (bandwidth, delay, jitter, and BER). It showed that SAW and TOPSIS provide similar performance to the traffic classes used. GRA provides a slightly higher bandwidth and lower delay for interactive and background traffic classes. AHP was used to determine the weights for the three models requiring information about the relative importance of each attribute. Results also showed that all four algorithms depend on the importance weights assigned to the parameters.

A network selection mechanism [18] has been proposed combining AHP (to achieve weighting of QoS parameters based on user preferences and service application) and GRA (to rank the network alternatives) techniques in order to find a tradeoff between user preferences, service application and network conditions. The mechanism is divided into three logical function blocks: "collecting data" which collects user preferences and network conditions, "processing data" which processes user-based data by AHP and normalizes network-based data by GRA, and "making decision block" which finalizes the process of balancing user preference, service application and network condition. The results revealed that it can work efficiently for an UMTS/WLAN system and also reduce the complexity of implementation significantly.

While vertical handover decision with multiple attribute is a complex problem, AHP seems to be the most popular method to decompose it into a hierarchy of simpler and more manageable subproblems [18]. These subproblems can be decision factors or weights according to their relative dominances to the problem. Decision factors can be solution alternatives and AHP selects the solution alternative with the largest synthesized weight. Briefly, AHP method is a three step process [19]:

- (1) Decomposes the decision problem into different levels of the hierarchy (identification of the decision criteria).
- (2) Compares each factor to all the other factors within the same level through pairwise comparison matrix (such as comparing objectives at the first level and networks with the respect of each objective at the second level).
- (3) Calculates the sum of products of weights obtained from the different levels, and selecting the solution with the highest sum.

This method is considered as a well-known and proven mathematical process. Otherwise, such a classical method remains insufficient to handle a decision problem with imprecision in decision criteria. More advanced methods are needed or combined with classical ones to get more efficient decision strategies. It is the scope of the following sections.

4.4. Fuzzy logic and neural networks based strategies (FL/NN)

Fuzzy Logic (FL) and Neural Networks (NN) concepts are applied to choose when and over which network to hand over among different available access networks. These are combined with the multiple criteria or attribute concept in order to develop advanced decision algorithms for both non-real-time and real-time applications. It is pointed out that classical MADM methods can not efficiently handle a decision problem with imprecise data that decision criteria could contain. For that, the use of FL is not only to deal with imprecise information but also to combine and evaluate multiple criteria simultaneously. Hence, FL concept provides a robust mathematical framework in which vertical handover decision can be formulated as a Fuzzy MADM.

In [4,20], a NN based vertical handover algorithm is proposed to satisfy user bandwidth requirements. It detects the RSS drop and makes handover decision. It is a three-layer back propagation NN used for pattern recognition. In a case of WLAN–GPRS handover based on a Mobile IP architecture, RSS samples from the AP are the input of the system. The output is a binary signal: 0 means that the MT should continue communicating with the AP and 1 means that the MT should make the handover and communicate with the BS. The authors have shown that the NN architecture performs better than traditional handover decision algorithms (RSS-based or hysteresis-based) in terms of handover delay and number of unnecessary handovers. However, such architecture requires prior knowledge of the radio environment and needs much configuration before deployment. In [20], the system architecture integrates geolocation capabilities into WLAN networks in order to exchange location data between IP devices and for MT to get information that it is approaching WLAN AP. NN-based strategy performs handover decision algorithm for choosing only the appropriate time to handover (based on RSS). Whereas, FL-based strategy performs handover decision algorithm for choosing the appropriate time and the most suitable access network according to user preferences.

In their works [5,21], Chan et al. proposed a solution incorporating FL in which terrestrial (GPRS and UMTS) and satellite mobile networks operate alongside each other. The handover decision algorithm aims at selecting

a segment or a network for a particular service that can satisfy objectives based on some criteria (such as low cost, good RSS, optimum bandwidth, low network latency, high reliability and long life battery) and taking into account the preferred access network. It is defined as an MODM algorithm, which requires inputs from the system (link quality, network characteristics and user profile) and the user (user preferences, application type, etc.) (see Fig. 2). The input from the user is used to determine the weighting given to each of the criteria used from the system such as the importance of the cost compared to the importance of the received QoS. The segment selection has two stages:

- (1) The *fuzzification* and *weighting* procedures:
 - The fuzzification evaluates and compares the available segments. Data from the system are converted into fuzzy sets in which each comparative criteria (such as cost criterion) can be represented by any value between 0 and 1 depending on a membership function. These representative values (known as membership values) for the fuzzy sets are obtained by mapping the measurements for a particular parameter onto a membership function.
 - The weighting evaluates the importance of each criteria based on instructions received from the network provider and the user. It uses AHP method (presented in Section 4.3) influenced by user preferences (cost, quality and application used) as criteria.
- (2) The *decision making*: application of the weightings to each criterion according to the defined objectives in a decision function. The chosen segment is the segment with the highest membership values of the decision function.

The first stage can be done before the handover initiation. The second stage uses the model proposed by Yager, as described by [21], which can generally be implemented in any multi-criteria system. In [5], the proposed system uses Mobile IP infrastructure. System procedures are defined as registration, location management and handover management functionalities. Handover management procedure retrieves the necessary information on the active segment (link measurements, user profile and QoS information), tries to negotiate the degradation or improvement in the QoS offered to the user (i.e., handover decision), and finally requests IP connectivity for Mobile IP registration (i.e., handover execution).

It is possible to have FL and NN concepts in the same vertical handover decision strategy. In [22], the algorithm consists of a Modified Elman Neural Network (MENN) for the number of users predicted after the handover (considered as an input of the adaptive multi-criteria decision) and a Fuzzy Inference System (FIS) that makes the analysis of relevant criteria and does the final decision according to the inputs and the IF-THEN rules (IF *condition*

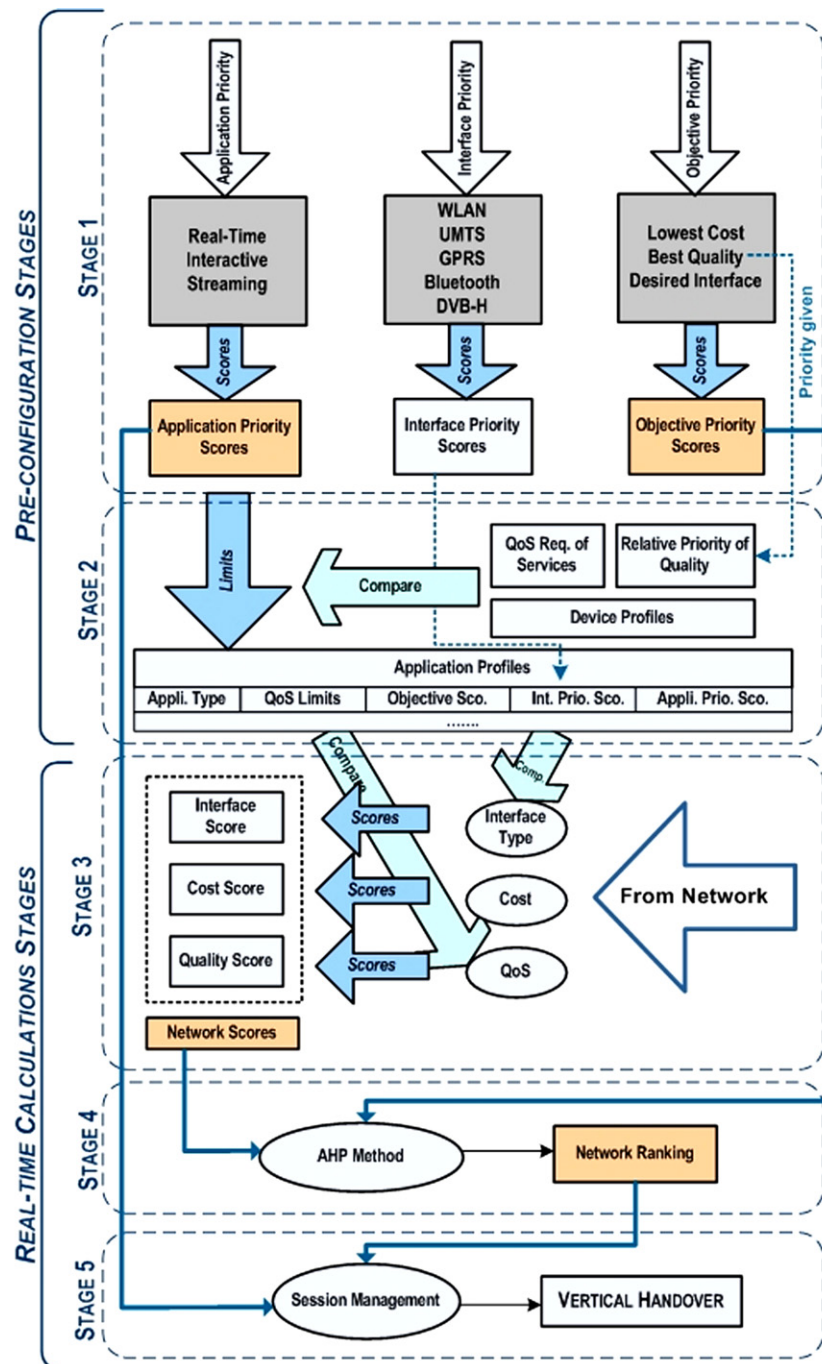


Fig. 2. FL based-segment selection strategy [21].

THEN action). The FIS represents the same system of the fuzzification stage in [5]. RSS measurement indicates the current radio link quality, and acts as a pretreatment that helps to decide whether to trigger the decision process. The inputs of the fuzzy multiple criteria inference system are the traffic bandwidth, the velocity of MT and the number of users predicted. The authors proved by simulations, on a UMTS–WLAN scenario, that the adaptive multi-criteria vertical handover decision algorithm can do accurately handover decision and performs better in guaranteeing the QoS of the after-handover communica-

tions comparing to the conventional algorithm (based on RSS criterion).

4.5. Context-aware strategies (CA)

The context-aware handover concept is based on the knowledge of the context information of the mobile terminal and the networks in order to take intelligent and better decisions [23]. Thus, a context-aware decision strategy manages these information and evaluate context changes to get decisions on whether the handover is necessary and

on the best target access network. Context information, relevant for the handover decision algorithm, are mentioned in Section 3.1 as handover decision criteria. It is related to the terminal or device (its capabilities, location, etc.), the user (its preferences), the network (QoS, coverage, etc.) and the service (QoS requirements, service type such as real-time, interactive or streaming etc.). Two context-aware decision solutions, [24] and [25], are based on AHP method defined in Section 4.3. This method is chosen to identify the most suitable choice (interface for a given application) among multiple alternatives that would satisfy some primary objectives based on the values of some context parameters.

In [24], the authors present a framework with an analytical context categorization and a detailed handover decision algorithm. It consists of two main components: the context repository and the adaptability manager. The context repository gathers, manages, and evaluates context information from different parts of the network. The adaptability manager decides about adaptation to context changes and handover execution. It is in charge of the vertical handover decision process. It is a rule-based process deciding when to invoke the handover operation (by evaluating terminal's location changes) and to which network (by evaluating QoS of the current and alternative networks). The QoS based network selection is based on AHP method and applies user perceived QoS. It has to satisfy multiple objectives: *maximizing user device preferences and application bandwidth*; and *minimizing jitter, delay, and loss and bandwidth fluctuations*. The proposed solution provides smart decision mechanisms with combining context gathering and handover decision processing. It is characterized as a mobile-assisted solution, since useful measurements are gathered from different parts of the system and the terminal.

Prototype experiments have used different types of access network (Ethernet, WLAN and GPRS) and a streaming application and have considered different scenarios (decision rules). It has shown that smart decision mechanisms are necessary for smooth adaptations to a variety of context changes.

Otherwise, [24] has a drawback: context information gathering is performed at a single point, i.e., the context repository. Beside the fact that it can form a single point of failure, it requires frequent communication between the terminal and the network, resulting in increased overhead on the radio link.

Another approach in [23], similar to the solution described in [24], proposes an advanced context management in which context information is distributed at more than one context repository (a user profile repository, a location information server and a network traffic monitor). It describes a flexible software-like deployment scheme which minimizes the handover decision time. This is achieved by software agents that are used for the preparation of the collected context data and by the algorithm needed for the handover at the context collection point. The software module that includes this information is

downloaded at the decision point (e.g., at the terminal) in advance and invoked at handover time.

The authors have implemented a prototype reproducing an overlapping UMTS–WLAN environment and using different test scenarios (one scenario without context-aware handover service and others with it). Context information concerns user's context information (location, speed and trajectory) and QoS required by the application. Through prototype evaluation, this approach proves that handovers are more efficient when context information are considered. Moreover, it seems to be flexible in a way that it ensures the possibility to use different protocols in exchanging different types of context information and to use different context-aware decision algorithms on mobile terminals. Specific decision algorithms can be used, context matrix evaluation as a simple linear calculation or a rule-based logic algorithm.

In [25], Ahmed et al. have developed and analyzed an intelligent handover decision algorithm (based also on AHP) including the session transfer (i.e., application management) which is a lack in [24]. They have considered a mobile-initiated and controlled solution. The context-aware decision algorithm is processed for each service type currently running on the device. Primary objectives were defined in terms of *Lowest Cost*, *Preferred Interface* and *Best Quality* (maximizing throughput, minimizing delay, jitter and BER). It has five stages (see Fig. 3): two stages of pre-configuration and three stages of real-time calculations.

The stages of *pre-configuration* are performed as follows:

- (1) *Taking user inputs*: defining the relative priorities among the primary objectives, the available interfaces and three types of services (real-time, interactive and streaming) with fixing priority scores between 1 (the most preferred one) and 9 (the least preferred one).
- (2) *Mapping limit values from discrete preferences*: expressing user QoS preferences as limits (upper and lower) in order to provide better flexibility while comparing them with network QoS parameters (very dynamic). These limit values, which are related directly to the priority given to the objectives of *Best Quality* (i.e., BER, delay, jitter, and throughput), are mapped for each of the three service types. It is based on QoS requirements of specific service type and device capabilities. At the end of the pre-configuration phase, three sets of data (scores and limits) for the three service types, are grouped together and stored as application profiles (see array in Fig. 3). Thanks to this array, these values are applied to compare and score networks, in the next stage, based QoS parameters of all the available networks. The stages of *real-time calculations* are performed for a particular type of running application as follows:
- (3) *Assigning scores to available networks*: comparing the capabilities of the reachable networks (interface, cost and QoS) with the pre-configured user preferences (scores and limits based on the primary objectives).

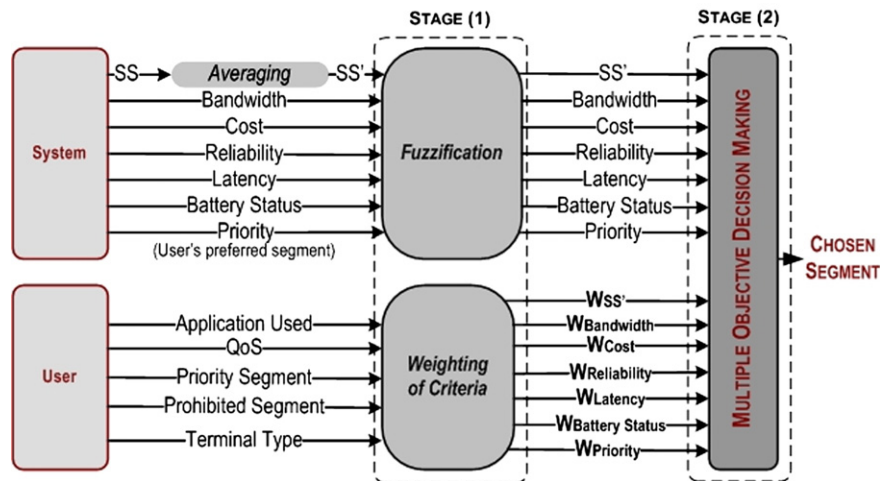


Fig. 3. Context-aware decision algorithm [25].

- (4) *Calculating network ranking* based on AHP method [19] through an objective pairwise comparison matrix (at first level) and network pairwise comparison matrix (at second level).
- (5) *Managing the session*: employing a session transfer scheduling algorithm in order to switch applications to the selected network.

It is simple (suitable for practical multimode mobile devices) by considering basic mathematical calculations and flexible or versatile (easily configurable by users). For simulation, the authors implemented a module with technology dependent individual and standard mobility features. This module receives notifications from the decision module and accordingly shifts applications to the alternative interface. It either generates a new IP address if the alternative interface does not have a configured one or retrieves the IP address already configured for the interface and then associates the address to applications [25]. The authors have considered a multimode mobile terminal with 2 WLAN interfaces and using an interactive service (web browsing application). The decision algorithm worked precisely and intelligently in accordance with the primary objectives. It was applied on only one decision rule (i.e., when more than one access network is available). It was evaluated for one performance parameter: time delays (experienced at the decision making and the session transfer phases). The total average handover delay is about 50–65 ms. It proves that the decision algorithm would work perfectly for more delay sensitive applications such as voice conversation or real-time video.

4.6. Synthesis

Our comparative study shows different issues to vertical handover decision problem: good network performance, user satisfaction, flexibility, efficiency, and multi-criteria solution. In a vision of an heterogeneous environment, we see that traditional handover decision strategies (RSS-

based), presented in Section 3.2, are not sufficient to make a vertical handover decision. They do not take into account the current context or user preferences. So, vertical handover decision strategy involves complex considerations and tradeoffs. It needs to be flexible and efficient considering the useful criteria and reasonable policies or rules applicable to both user's professional and personal communications.

In Table 1, we summarize the given strategies compared on different characteristics such as multi-criteria choice, efficiency, or service types supported.

The *multi-criteria* solution seems to be an essential aspect in vertical handover decision. For instance, NN-based strategy, compared to the other strategies, is based on only one parameter (i.e., RSS measurements) and on one type of handover policy (i.e., keeping WLAN connection when it is available).

User consideration is also a very interesting characteristic in vertical handover decision. It can include user intervention (user preferences), user interaction (with automation or not) or user satisfaction. Considering this aspect, User-Centric (UC) and Context-Aware (CA) strategies are the most relevant ones.

We compare also the different strategies regarding few common characteristics: efficiency, flexibility and complexity. It means by *efficiency* the possibility of obtaining a precise decision with good performance; and by *flexibility* the separation of the handover decision mechanism from the whole handover management process and its adaptation with additional parameters or functionalities. The implementation of cost or utility functions (DF) seems to be more flexible for the use of vertical handover policies but less efficient on this aspect for real-time applications. The use of FL and MAD algorithms allows analyzing complex problems. It answers well to vertical handover decision problem by giving the best and accurate solution with regrouping all the decision factors. CA strategies try to ensure a high flexibility as important as a high efficiency facing an heterogeneous environment.

Table 1
Comparison between vertical handover decision strategies

Vertical handover decision strategy	Traditional (RSS-based)	DF	UC	MAD	FL/NN	CA
Multi-criteria	No	Yes	Yes	Yes	Yes (FL) No (NN)	Yes
User consideration	No	Low	High	Medium	Medium	High
Efficiency	Low	Medium	Medium	High	High	High
Flexibility	Low	High	High	High	Medium	High
Implementation complexity	Low	Low	Low	Medium	High	Medium
Service type supported	Non-real-time	Non-real-time and Real-time	Non-real-time	Non-real-time and real-time	Non-real-time and real-time	Non-real-time and real-time

DF, decision function; UC, user-centric; MAD, multiple attribute decision; FL/NN, Fuzzy Logic/Neural Networks, CA, context-aware.

Concerning the *implementation complexity*, we focus on the fact that the decision mechanism can be complex in itself. NN-based strategy seems to be the more complex one due to its complicated topology and, hence, it is difficult to implement in practice. However, FL and NN based strategies together consider only a few context parameters and can be more complex to be suitable for practical multimode mobile terminal with limited resources. Otherwise, some CA strategies, compared to FL/NN strategies, apply classical MADM methods that use simpler calculations. One of the advantages of MAD, FL and CA strategies is combining and evaluating multiple decision criteria simultaneously. But, some decision criteria or contextual information can be imprecise or unavailable. FL is the tool involved in decision strategies to cope with imprecision. It is proved by the use of Fuzzy MADM formulation in some FL strategies in combination with MAD ones.

Regarding *service types*, non-real-time applications (file transferring, web browsing, etc.) are supported by all the strategies. Only traditional and user-centric (UC) strategies do not support real-time applications (streaming, video or voice conferencing, etc.).

Some other characteristics such as stability and handover performance (in terms of throughput, handover delay and packet loss) do not appear in Table 1. Only DF strategies evaluate the *stability* aspect which, on one hand, ensures that a handover is worthwhile for each mobile and, on the other hand, copes with handover synchronization when simultaneous decisions are taken by many terminals. Answering the stability issue maintains load balancing and avoids unnecessary handovers. *Handover performance* depends strongly on the type of applications used (real-time or non-real-time). Beside the comparative features mentioned previously, we add interesting ones that some strategies have in common: Mobile IP functionalities support in [5,12,15,20,23] and their evaluation between two kinds of access networks: WLAN and cellular networks such as GSM, GPRS, and UMTS [12,15,18,22,24].

According to our comparison of vertical handover decision strategies in Table 1, CA strategies seem to have the best performance answering well to the majority of the given characteristics. This category is followed by MAD

and FL strategies. Thus, it is obvious to see that FL and CA strategies are better enhanced in a way that both can be combined with MADM methods. From one side, FL concept provides a robust mathematical framework [5,21]. To the other side, context-aware concept manages context information and evaluates context changes to select appropriate adaptation methods such as the vertical handover [24,25]. Moreover, the use of this concept is well proven in more than one research area.

5. Our proposal

Based on our synthesis on vertical handover decision strategies given previously, we build our strategy while taking advantage of the most interesting solutions and particularly the best aspect of each one.

First of all, a multi-criteria solution is needed. Contextual information (from terminal and network sides) can be used as multiple criteria useful enough to avoid wrong handover decisions, then bad performance. So, we consider a context-aware vertical handover decision. It should be conscious of the possibilities offered by each access network, it should know MT's movements and it should take into account QoS requirements for the demanding service.

Otherwise, some of the contextual information presents uncertainty (incomplete or unavailable). Advanced decision algorithms are needed to deal with this kind of information. It is needed also to compare and evaluate these information which are different from one access network to another. So, we consider a FL system which is flexible, capable of operating with imprecise data and can be used to model nonlinear functions with arbitrary complexity.

Finally, to cope with the complexity aspect, we need vertical handover policies expressing rules that contribute to shape and help the whole handover decision process.

This combination of a context-aware approach using policies can provide an efficient and flexible vertical handover decision solution. We give more flexibility in a way that the whole handover process is completely controlled by the mobile (Mobile-Controlled HandOver, MCHO). It reduces more the overall complexity in the network, the signaling overhead and the handover latency than a

Mobile-Assisted Handover (MAHO). Most conducted experiments and publications in vertical handovers [5,10,25,23], even regarding policies, promote a MCHO decision model in which the MT is responsible for making decisions and to put all the intelligence at the MT. Therefore, we opt for a MCHO solution with respect to transfer of handover decision criteria and more precisely regrouping context information. Thus, MT conducts the initiation (at the decision phase) and the control of the handover (at the execution phase). Otherwise, MCHO does not exclude the assistance from the network in a way that it needs information, such as capabilities or bandwidth, to choose the most optimal network among those available. Moreover, this proves the distribution of computation between MTs compared to a centralized approach (a Network-Controlled HandOver, NCHO).

Our vertical handover decision approach described in Fig. 4. The main phases of the handover management process, shown in this figure, have been introduced in Section 2.1: Handover Information Gathering, Handover Decision, and Handover Execution.

Handover Information Gathering collects all the contextual information required to apply handover policies stored in *Handover Policies Repository*, through monitoring and measurements. It is responsible to keep the repository up to date. This repository stores a set of policies expressing decision rules that govern the choices in the whole process. These policies can be user (based on user preferences), corporate (such as enterprise specifications) or network (based on network operator rules) policies. They are applied on scenarios as follows: when a

mobile user is moving out of the serving access network and will enter another access network shortly; or when a mobile user is connected to a particular network, but chooses to be handed over to another available access network for its future service needs. This is how the handover policies repository applies decision rules based on connectivity, networks availability, user, and corporate preferences.

It is pointed out, in Section 2, that *Handover Decision* decides if a handover is needed at this point of time or not (the appropriate moment to initiate the handover), provided by the *Handover Initiation* module and it chooses the best target access network, provided by the *Network Selection* module.

The handover initiation stage is performed by a *Fuzzy Logic System* (FLS) with fuzzification/defuzzification mechanisms as described in [5]. Deciding for the correct time to initiate a vertical handover can reduce the subsequent handovers, limit the signaling messages and predict disconnections. It uses contextual information such as velocity, coverage or location.

Two steps have to be performed before the network selection process itself: the *criteria scoring* in which the importance of each decision criteria is evaluated according to user preferences, and the *network scoring* in which the available networks are evaluated and compared for each handover decision criteria. At the next step, the Analytic Hierarchy Process (AHP) method, as a MODM, is employed as described in [19]. It is chosen due to its ability to vary its weighting between each objective regarding user preferences.

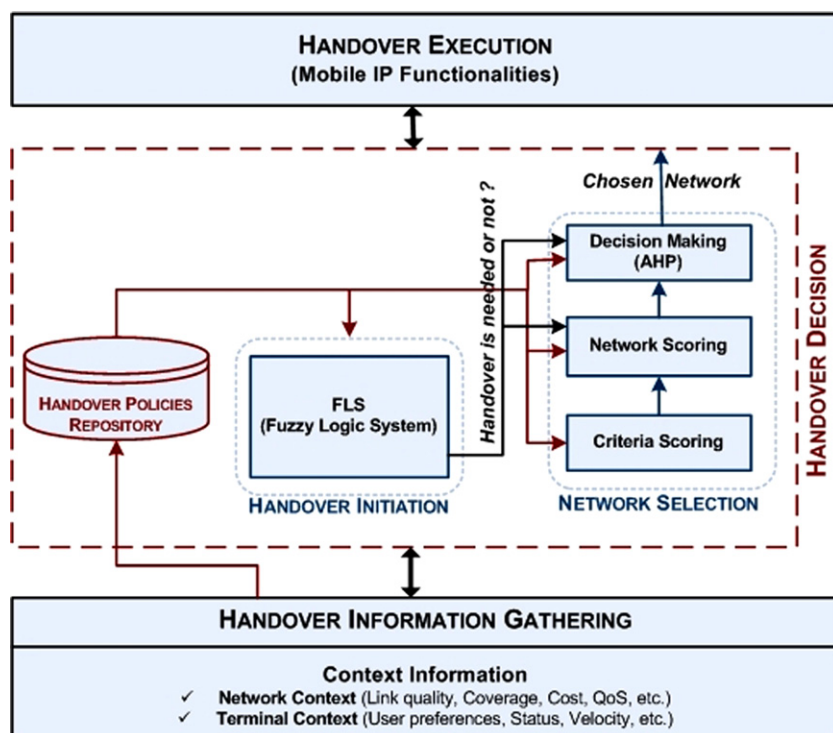


Fig. 4. Our vertical handover decision scheme.

Handover Execution establishes the IP connectivity through the chosen access network using Mobile IP functionalities.

Our vertical handover decision scheme is detailed in a handover management solution proposed in [26].

Our solution focuses on the handover decision aspect providing flexibility and efficiency thanks to advanced multiple objective and context-aware decision algorithms (FL and AHP) and policies expressing rules that shape the whole decision process. Thus, it should provide handover performance optimization and should prepare the handover execution stage. This last stage of the handover process uses Mobile IP functionalities (for maintaining IP connectivity) to ensure service continuity (see Section 2.1).

6. Conclusion

In this paper, we give an overview of the vertical handover decision process with a classification of the different existing vertical handover decision strategies. It has shown that advanced evaluation functions and optimized architectures are needed to perform better handover decision making for user satisfaction as well as for the efficient use of network resources. Each strategy gives answers to the different questions listed in Section 2.2. Thus, the goal of handover decision process consists of finding the appropriate time to perform the handover and the most optimum access network according to user demands, network resources and terminal capabilities. Regarding the entire handover management process, additional considerations are taken into account in most of the proposed strategies: Mobile IP-like infrastructure and the types of access network types such as WLAN and cellular networks. So, to build a handover management solution, some issues have to be considered in the choice of a vertical handover decision strategy: handover control (at the mobile terminal or network), information gathering, handover execution procedure, more available access networks and handover performance evaluation. In the work in progress, we are proposing a sophisticated handover management scheme focusing on the described handover decision strategy (for handover initiation and network selection) that can be applied on an interesting network infrastructure such as a loosely-coupled 3G-WLAN interworking based on Mobile IP functionalities. This work intends to answer vertical handover challenges as flexibility, efficiency, seamlessness, automation and performance optimization. The application of our vertical handover scheme can be extended to other interworking cases such as 3G-WMAN and WLAN-WMAN.

Acknowledgement

This research is fully funded by ITEA SUMO (Service Ubiquity in MOBILE and Wireless Realm) Project (ITEA 04042).

References

- [1] J. McNair, F. Zhu, Vertical handoffs in fourth-generation multinet-work environments, *IEEE Wireless Communications* 11 (3) (2004) 8–15.
- [2] I. Akyildiz, J. Xie, S. Mohanty, A survey of mobility management in next-generation all-IP-based wireless systems, *IEEE Wireless Communications* 11 (4) (2004) 16–28.
- [3] E. Gustafsson, A. Jonsson, Always best connected, *IEEE Wireless Communications* 10 (1) (2003) 49–55.
- [4] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J. Makela, R. Pichna, J. Vallström, Handoff in hybrid mobile data networks, *IEEE Personal Communications* 7 (2) (2000) 34–47.
- [5] P. Chan, R. Sheriff, Y. Hu, P. Conforto, C. Tocci, Mobility management incorporating fuzzy logic for a heterogeneous IP environment, *IEEE Communications Magazine* 39 (12) (2001) 42–51.
- [6] E. Stevens-Navarro, V. Wong, Comparison between vertical handoff decision algorithms for heterogeneous wireless networks, in: *Proceedings of IEEE Vehicular Technology Conference (VTC-Spring)*, vol. 2, 2006, pp. 947–951.
- [7] F. Zdarsky, J. Schmitt, Handover in mobile communication networks: who is in control anyway?, in: *Proceedings of the 30th Euromicro Conference*, 2004, pp. 205–212.
- [8] C.E. Perkins, Mobile IP, *IEEE Communications Magazine* 40 (2002) 66–82.
- [9] M. Bernaschi, F. Cacace, A. Pescapè, S. Za, analysis and experimentation over heterogeneous wireless networks, in: *First International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities*, 2005 (Tridentcom 2005), 2005, pp. 182–191.
- [10] W. Zhang, J. Jaehnert, K. Dolzer, Design and evaluation of a handover decision strategy for 4th generation mobile networks, in: *The 57th IEEE Semiannual Vehicular Technology Conference*, 2003 (VTC 2003-Spring), vol. 3, 2003, pp. 1969–1973.
- [11] G. Pollini, Trends in handover design, *IEEE Communications Magazine* 34 (3) (1996) 82–90.
- [12] H. Wang, R. Katz, J. Giese, Policy-enabled handoffs across heterogeneous wireless networks, *Second IEEE Workshop on Mobile Computing Systems and Applications*, 1999 (Proceedings WMCSA'99), 1999, pp. 51–60.
- [13] W. Chen, J. Liu, H. Huang, An adaptive scheme for vertical handoff in wireless overlay networks, in: *Proceedings on the 10th International Conference on Parallel and Distributed Systems*, 2004 (ICPADS 2004), 2004, pp. 541–548.
- [14] F. Zhu, J. McNair, Optimizations for vertical handoff decision algorithms, in: *IEEE Wireless Communications and Networking Conference*, 2004 (WCNC 2004), vol. 2, 2004, pp. 867–872.
- [15] A. Calvagna, G. Di Modica, A user-centric analysis of vertical handovers, in: *Proceedings of the Second ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, 2004, pp. 137–146.
- [16] O. Ormond, J. Murphy, G. Muntean, Utility-based intelligent network selection in beyond 3G systems, in: *IEEE International Conference on Communications (ICC 2006)*, vol. 4, 2006, pp. 1831–1836.
- [17] R. Ribeiro, Fuzzy multiple attribute decision making: a review and new preference elicitation techniques, *Fuzzy Sets and Systems* 78 (2) (1996) 155–181.
- [18] S. Quiyang, A. Jamalipour, A network selection mechanism for next generation networks, in: *International Conference of Communications*, 2005 (ICC 2005), vol. 2, 2005, pp. 1418–1422.
- [19] T. Saaty, How to make a decision: the analytic hierarchy process, *European Journal of Operational Research* 48 (1) (1990) 9–26.
- [20] J. Makela, M. Ylianttila, K. Pahlavan, Handoff decision in multi-service networks, in: *The 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2000 (PIMRC 2000), vol. 1, 2000, pp. 655–659.

- [21] P. Chan, Y. Hu, R. Sheriff, Implementation of fuzzy multiple objective decision making algorithm in a heterogeneous mobile environment, in: *IEEE Wireless Communications and Networking Conference, 2002 (WCNC 2002)*, vol. 1, 2002, pp. 332–336.
- [22] Q. Guo, J. Zhu, X. Xu, An adaptive multi-criteria vertical handoff decision algorithm for radio heterogeneous network, in: *IEEE International Conference on Communications, 2005 (ICC 2005)*, vol. 4, 2005, pp. 2769–2773.
- [23] Q. Wei, K. Farkas, C. Prehofer, P. Mendes, B. Plattner, Context-aware handover using active network technology, *Computer Networks* 50 (15) (2006) 2855–2872.
- [24] S. Balasubramaniam, J. Indulska, Vertical handover supporting pervasive computing in future wireless networks, *Computer Communications* 27 (8) (2004) 708–719.
- [25] T. Ahmed, K. Kyamakya, M. Ludwig, A context-aware vertical handover decision algorithm for multimode mobile terminals and its performance, in: *Proceedings of the IEEE/ACM Euro American Conference on Telematics and Information Systems (EATIS 2006)*, 2006, pp. 19–28.
- [26] M. Kassar, B. Kervella, G. Pujolle, Architecture of an intelligent intersystem handover management scheme, *Future Generation Communication and Networking* 1 (2007) 332–337.