Context-Aware Decision Model for Vertical Handover in Heterogeneous Networks

Tansir Ahmed[†], Kyandoghere Kyamakya[*], and Markus Ludwig[†]
[†]BenQ Mobile, MD PBM TI 8, Haidenauplatz 1, 81667 Munich, Germany
[*]Institute of Communication Engineering, University of Hannover, Appelstrasse 9A, 30167 Hanover, Germany

***Abstract*–Present day mobile networks are evolving towards heterogeneous overlaying infrastructure. Traditional handover decisions that are mainly based on signal strength are not sufficient to provide ubiquitous and seamless mobility across heterogeneous networks. Intelligent handover decision is needed so that users can select the best option available from diverse networks and services as per their requirements. The decision mechanism is also needed to realize multimode functionality in current and future multimode mobile devices that will enable user applications to use multiple radio interfaces simultaneously and to switch between active interfaces based on application requirements and interface capabilities. This paper introduces a context-aware vertical handover decision model suitable for heterogeneous networks environment based on the *Analytic Hierarchy Process (AHP)* method. It illustrates the whole mechanism starting from grouping and matching of relevant contexts in the AHP model to the application management based on the handover decision.**

1. Introduction

Present day wireless communications networks and devices are experiencing a paradigm shift. Rapid emergence of diverse access technologies, e.g. WLAN, Bluetooth, 3GPP cellular networks (GSM, GPRS, UMTS), DVB-H (Digital Video Broadcasting-Handheld), etc would result in evolution of wireless networks towards heterogeneous all-IP infrastructure. In this heterogeneous *overlaying* infrastructure, users should be given the freedom to roam globally among multitude points of attachment of different access networks (*vertical handover*) as per their service requirements.

Conventional single interface mobile terminals are also evolving into multimode terminals. Currently, these multimode terminals do not possess true multimode functionality. They are limited to use only one radio interface at a time. But in the given heterogeneous scenario, these terminals should have true multimode functionality that would enable user applications to use multiple radio interfaces simultaneously and to switch between active interfaces based on application requirements and interface capabilities (*vertical handover*).

Traditional *horizontal handover (HO)* decision mechanisms that mainly depend on signal strength for decision making are unable to realize ubiquitous and seamless mobility across heterogeneous networks as well as true multimode functionality in mobile devices. In the given circumstances, intelligent HO decision mechanism is needed that would take into account user preferences, application requirements, device and interface capabilities, and other information or intelligence residing on the terminal side as well as on the network side, collectively known as *context information*.

The rest of the paper is organized as follows. Section 2 highlights the related work. Section 3 illustrates a context-aware vertical HO decision model based on the *Analytic Hierarchy Process (AHP)* [1] method. Finally, section 4 concludes the paper.

2. Related Work

References [2] and [3] propose a *fuzzy* based multiple-criteria decision making process to perform access network selection and vertical HO. The HO decision is, however, confined only to the cost constraints and application priorities specified by users. References [4] and [5] describe intelligent HO procedures especially for hybrid networks but the number of parameters considered for the HO decision is only the type of the radio access technology and the signal strength. In [6], different HO policies for heterogeneous networks are used considering as HO decision parameters mainly the type of air interface and the available bandwidth at the *access router (AR)*. Reference [7] presents an analytical vertical HO initiation model based on the criteria of *received signal strength (RSS)* and distance. Reference [8] also uses the AHP method in their decision making process. However, [8] and all other work mentioned earlier lack a detailed grouping of relevant context information, an elaborate decision model and algorithm that would consider a wide variety of the most important context information for selecting a suitable network in heterogeneous environment, precise calculation methods for mapping relevant contexts in the chosen model and for decision making, user interactions in the process, and lastly, the application management based on the HO decision.

3. Vertical Handover Decision Model

The AHP [1] method is a well-known and proven mathematical process to identify the most suitable choice among multiple alternatives based on multiple objectives. The task of our context-aware decision model is to select the most suitable interface for a given application among multiple options that would satisfy some primary objectives based on the values of some context parameters. In this regard, the AHP model perfectly fits in our context-aware decision making process. We have considered *mobile-initiated and controlled* vertical HO for the decision model.

In accordance with the AHP method, at first, we have to define some *primary objectives* for our decision model taking into account the preferences likely to be the most interesting to users (e.g. cost, interface priority based on coverage, etc) and *3GPP* defined *Quality of Service (QoS)* parameters [9]. We have chosen the following *six* objectives:

1) Consider interface priority.
2) Minimize cost.
3) Maximize mean throughput.
4) Minimize delay.
5) Minimize jitter.
6) Minimize Bit Error Rate (BER)/Frame Error Rate (FER).

3.1. Context Model

Context information may be classified based on their frequency of changes and based on their placement. In the former case, it is either *static* or *dynamic*, and in the latter case, it can be hosted either on the terminal side or on the network side. The contexts that do not change very often are static context information, whereas those that change quite frequently and may loose accuracy over time are dynamic context information. The context model chosen for the decision model is shown in Table 1.

TABLE 1
CONTEXT MODEL FOR DECISION ALGORITHM

Context Type	Terminal Side	Network Side
Static	Device capability, services, user preferences	Provider's profile
Dynamic	Running application type, reachable access points	Current QoS parameters

On the terminal side, **device capability** includes display size, resolution, battery life, RAM, processor speed, and multimode capability. All **services** offered by a terminal are classified into *three* **service types**, namely conversational/real-time services, interactive services, and streaming services. Each of the three service categories has its own **QoS requirements** (service precedence, delay, mean throughput, peak throughput, and reliability). **User preferences** are grouped as **interface preferences** for multimode terminal and **service preferences** (precedence of service types, billing constraints, QoS preferences). **Running application types** defines the service profile of currently running applications. **Reachable access points (APs)** identifies currently available networks and addresses of the APs.

On the network side, **Service provider's profiles** consist of provider's identity, policies, charging models, roaming agreement models (on-demand, settled before, or mixed), etc. **Current QoS parameters** define the current status of the available network QoS parameters.

3.2. Architecture of Context-Aware Decision Model

The architecture of the context-aware decision model is shown in Fig. 1. In this model, a user defines his preferences in some categories that should meet both application requirements and device capabilities. Capabilities of available networks are discovered and compared with the defined preferences by employing some intelligent algorithm, and finally, the most suitable network corresponding to the preferences is selected. Preferences for some parameters (e.g. cost) are best expressed as *discrete* values and for some other (e.g. mean throughput), *continuous* values. Discrete preferences are represented by *scores* (integers) and continuous preferences by upper and lower *limits*. For the ease of users, only discrete preferences are taken from the user side and mapped into continuous preferences at the backend. The decision model is based on types of services and all applications supported by the target mobile device are classified into the three service types mentioned earlier. Note that the AHP method is used only in our final stage (stage 4) of calculations.

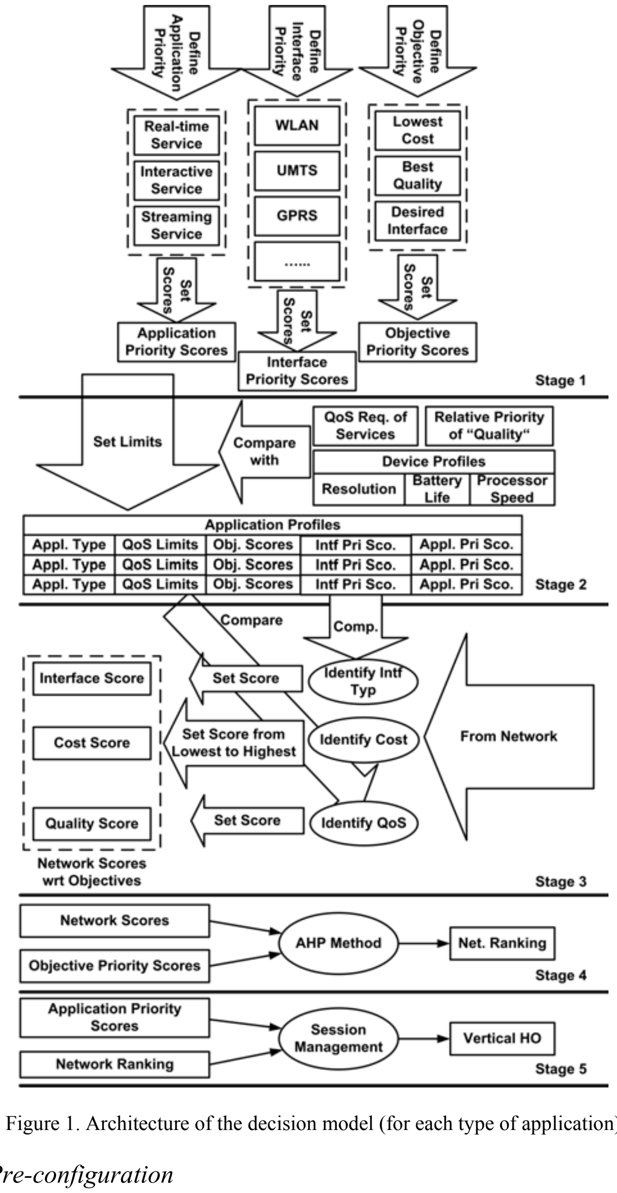

Figure 1. Architecture of the decision model (for each type of application)

Pre-configuration

Stage 1: Taking user inputs

A user needs to define *three* sets of relative priorities for each of the three types of services among (i) the primary objectives (*objective priorities*) (ii) available interfaces in a device (*interface priorities*) and (iii) the three types of services (*application priorities*). These preference inputs should be taken in as much user-perceivable and user-friendly way as possible. In our model, available options, in each case, are labeled with suitable *literals*. Each of the literals has a priority score calculated and assigned at the background. The user only needs to arrange the literals in a descending order starting with the one with the highest priority. Based on the arrangement of the literals priority scores between 1-9 are assigned, where 1 denotes the most preferred one and 9 denotes the least preferred one. Priority scores are equal-spaced integers whose space-gap is defined by (1), where N_p denotes the number of parameters, L_u and L_l denote the highest and lowest possible scores i.e. 9 and 1, respectively, and G denotes the numeric space-gap between two subsequent scores, which is rounded off to the nearest integer.

$$G = \frac{L_u - L_l}{N_p} \qquad (1)$$

As an example, among the primary objectives mentioned earlier objective 1 is labeled as "Desired Interface", objective 2 as "Lowest Cost", and objectives 3 to 6, in a group, as "Best Quality". Here, (1) results in $G = 3$ while using $L_u = 9$, $L_l = 1$, and $N_p = 3$. If a user arranges the literals as in the order "Lowest Cost", "Best Quality", and "Desired Interface" objective 2, objectives 3-6, and objective 1 have scores of 1, 4, and 7, respectively. Since "Best Quality" is the group of four parameters objectives 3-6 have the same score, 4. The process is illustrated in Fig. 2. Similar measures are taken in case of interface and application priorities. For the former, N_p equals to the number of interfaces in the terminal and for the latter, types of services. It is worth mentioning that each of the sets of literals should include a "Default" option.

Stage 2: Mapping limit values from discrete preferences

Considering the fact that the behavior of some context information especially network QoS parameters is very dynamic, it makes sense to express QoS preferences from users as continuous values in order to provide better flexibility while comparing them with network QoS parameters. At this stage, suitable limit values (upper and lower) for the four QoS parameters related directly to objectives 3-6 (mean throughput, delay, jitter, and BER/FER) are mapped from the discrete preferences of a user and other context information for each of the three types of services. While fixing the limit values it is important to note that higher values are not always better for all the four QoS parameters. It is always preferable to have values as high as possible for mean throughput, whereas as low as possible for delay, jitter, and BER/FER.

In case of mean throughput, the lower limit is always a fixed value, i.e. *minimum requirement*. This value is based on the analysis of contexts like QoS requirements for applications and device capabilities. The upper limit varies in accordance with the objective priority scores of the QoS based objectives (objectives 3-6) set earlier. For example, if the QoS based objectives have the highest priority (priority score equals 1) the upper limit is set at the highest possible value, on the contrary, if they have the lowest priority (priority score equals 7) it is set much nearer to the lower limit. The limit values for the other three QoS parameters are fixed likewise, except that the upper limit is always a fixed value in this case, i.e. *maximum tolerance* and the lower limit varies according to the objective priority scores.

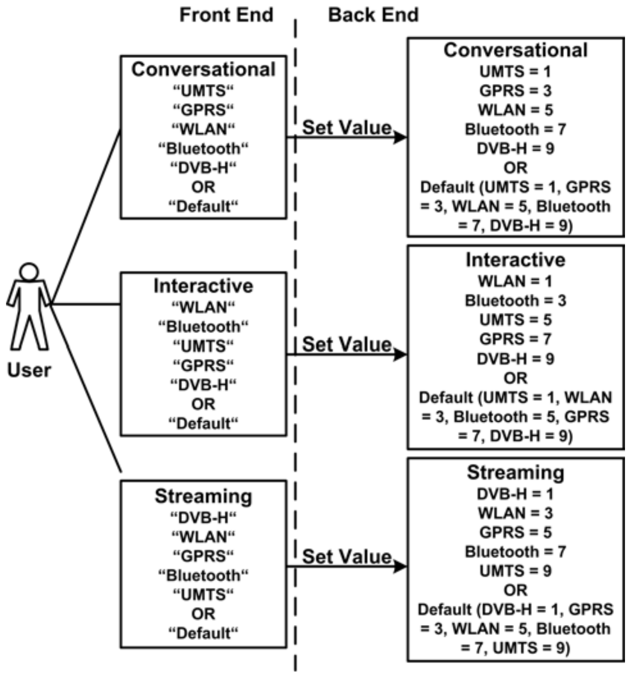

Figure 2. Taking user inputs on interface priority

At the end, we have three sets of *preconfigured* data (scores and limits) for the three service types. They are grouped together and stored as *application profiles* (see Fig. 1) where individual service type is identified by *application type*. All the running applications inside a mobile device would have predefined *application type*. Thus, any application, during runtime, could be paired with individual set of *application profiles* based on its *application type*.

Real-time calculations

The following stages perform real-time calculations for a particular type of running application.

Stage 3: Assigning scores to available networks

At this stage, capabilities of reachable networks (including the current network, if any) are compared with preconfigured user preferences based on six primary objectives and suitable scores are assigned to each of the networks. A multimode mobile device would always monitor (layer-2 or layer-3 monitoring, or both) each of its interfaces for reachable networks. It is assumed that shared contexts of networks like current QoS parameters and cost would always be advertised by available networks, or the terminal may utilize layer-2 or layer-3 probing. Context information of each available network would be stored in the terminal in *reachable network profiles* (see Fig. 1). The terminal may also involve some intelligent method to measure, specially, network QoS parameters from available layer-2 or layer-3 signal as such contexts may not be advertised explicitly by the networks, but this scenario is out of scope for this paper.

Assignment of scores to available networks based on discrete preferences like interface priority and cost constraint is straightforward. Depending on the type of interface an appropriate interface priority score, defined by the user in stage 1, is assigned to each available network. In case of cost objective, all the available networks are compared with each other and assigned with appropriate equal-spaced scores between 1-9 based on (1) in a descending order, where the cheapest network has a score of 1. If a particular network does not advertise the cost information it is assigned with a score of 9 (costliest network) as a default value.

$$\begin{aligned} S_i &= \left(1-\frac{n_i-l_i}{u_i-l_i}\right)\times 10 && ; l_i < n_i < u_i \\ &= 1 && ; n_i \geq u_i \\ &= 9 && ; n_i \leq l_i \end{aligned} \tag{2}$$

$$\begin{aligned} S_i &= \left(\frac{n_i-l_i}{u_i-l_i}\right)\times 10 && ; l_i < n_i < u_i \\ &= 1 && ; n_i \leq l_i \\ &= 9 && ; n_i \geq u_i \end{aligned} \tag{3}$$

In case of continuous preferences, QoS parameters of all available networks are compared with the individual parameter limit values defined in stage 2. If u_i and l_i denote the upper and lower limits of a particular continuous preference and n_i denotes the value offered by a network for that particular parameter the network score, S_i, based on the preference is calculated using (2) and (3). Eq. (2) is used for continuous preferences like mean throughput, where the target value is preferred to be as high as possible. On the contrary, (3) is used for continuous preferences like delay, jitter, and BER/FER, where the target value is preferred to be as low as possible. If there is any missing parameter i.e. not advertised by a particular network its default value is used. Values of l_i and u_i are the default values for (2) and (3), respectively.

Stage 4: Calculating network ranking based on AHP method

At this stage, *ranking* of the available networks are performed based on the objective priority scores and network scores assigned at stage 1 and 3, respectively. The calculations use the AHP method [1], which is a three step process.

Step 1: At first, the relative scores among the objective priority scores set by the user at stage 1 are calculated. Relative scores are scaled linearly between 1-9 [1]. Relative scores between any two particular scores are calculated using (4), (5), and (6), where RS_{ab} is the relative score between parameters *a* and *b*, and S_a and S_b are their respective scores.

$$\frac{1}{RS_{ab}} = \left(1-\frac{S_b}{S_a}\right)\times 10 \qquad ; S_a > S_b \tag{4}$$

$$RS_{ab} = \left(1-\frac{S_a}{S_b}\right)\times 10 \qquad ; S_a < S_b \tag{5}$$

$$RS_{ab} = 1 \qquad ; S_a = S_b \tag{6}$$

With the calculated relative scores the priorities (i.e. weights) for the six objectives in terms of the overall goal i.e. selecting a suitable network are calculated using *pairwise comparison matrix* [1] for objectives. It consists of the relative scores calculated in the previous step. The dimension of the pairwise comparison matrix *A* for the objectives, as shown below, is flexible and dependent on the number of chosen objectives (6 × 6, in our case):

$$A = \begin{bmatrix} 1 & RS_{12} & RS_{13} & RS_{14} & RS_{15} & RS_{16} \\ \frac{1}{RS_{12}} & 1 & RS_{23} & RS_{24} & RS_{25} & RS_{26} \\ \frac{1}{RS_{13}} & \frac{1}{RS_{23}} & 1 & RS_{34} & RS_{35} & RS_{36} \\ \frac{1}{RS_{14}} & \frac{1}{RS_{24}} & \frac{1}{RS_{34}} & 1 & RS_{45} & RS_{46} \\ \frac{1}{RS_{15}} & \frac{1}{RS_{25}} & \frac{1}{RS_{35}} & \frac{1}{RS_{45}} & 1 & RS_{56} \\ \frac{1}{RS_{16}} & \frac{1}{RS_{26}} & \frac{1}{RS_{36}} & \frac{1}{RS_{46}} & \frac{1}{RS_{56}} & 1 \end{bmatrix} \tag{7}$$

Matrix *A* is then normalized by dividing each element by individual sum of column. The normalized matrix A_{norm} is shown in (8).

$$A_{norm} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & b_{36} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} & b_{46} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} & b_{56} \\ b_{61} & b_{62} & b_{63} & b_{64} & b_{65} & b_{66} \end{bmatrix} \tag{8}$$

At the end, the average values of each row for objective *i* are calculated to give the priorities for each objective (p_1, p_2, p_3, p_4, p_5, p_6) with respect to the overall goal using (9).

$$p_i = \frac{b_{i1}+b_{i2}+b_{i3}+b_{i4}+b_{i5}+b_{i6}}{6} \tag{9}$$

Step 2: The relative scores among the scores of the available networks assigned at stage 3 in terms of individual objective is calculated using (4), (5), and (6). Then the network conformances (i.e. weights), c_{ij}, for *i* number of available networks in terms of each of *j* number objectives are calculated in similar fashion as described in step 1. For example, two available networks, WLAN and GPRS, score 1 and 8, respectively, in terms of cost. Their pairwise comparison matrix with respect to cost is shown below, where (4) is used for calculating the relative score RS_{12} between the two networks in terms of cost.

$$\begin{array}{c} \\ Net1 \\ Net2 \end{array} \begin{array}{c} \begin{array}{cc} Net1 & Net2 \end{array} \\ \begin{bmatrix} 1 & RS_{12} \\ \dfrac{1}{RS_{12}} & 1 \end{bmatrix} \end{array} \quad (10)$$

$$\begin{array}{c} \\ WLAN \\ GPRS \end{array} \begin{array}{c} \begin{array}{cc} WLAN & GPRS \end{array} \\ \begin{bmatrix} 1 & 8.75 \\ \dfrac{1}{8.75} & 1 \end{bmatrix} \end{array} \quad (11)$$

Normalizing (11) we get,

$$\begin{array}{c} \\ WLAN \\ GPRS \end{array} \begin{array}{c} \begin{array}{cc} WLAN & GPRS \end{array} \\ \begin{bmatrix} 0.9 & 0.9 \\ 0.1 & 0.1 \end{bmatrix} \end{array} \quad (12)$$

Now, the network conformances $c_{1,\ cost}$ and $c_{2,\ cost}$ for WLAN and GPRS, respectively, is calculated from (12) using (9). Thus, $c_{1,\ cost} = (0.9 + 0.9)/2 = 0.9$ and $c_{2,\ cost} = (0.1 + 0.1)/2 = 0.1$.

Step 3: The overall ranking of each available network is determined by calculating the sum of products of network conformances in terms of individual objective (obtained from step 2) and objective priorities for that particular objective (obtained from step 1). For i number of available networks and j number of objectives, the overall ranking R_i can be obtained from (13).

$$R_i = \sum_1^{ij} c_{ij}(p_j) \quad (13)$$

R_i is always in the range of 0-1. The network with the highest rank is finally selected.

Stage 5: Session management

At this final stage, an efficient session transfer scheduling algorithm is employed in order to switch applications to the selected network. The scheduling algorithm takes into account the application priority score set by the user at stage 1 and the rank of the selected network obtained from (13) at stage 4. For i number of running applications the overall score, O_i, is calculated using (14), where, R_d and R_i, respectively, are the ranks of the current and the selected network for the i^{th} application, and a'_i is the normalized value of its application priority score, a_i, given by (15).

$$O_i = a'_i(R_i - R_d) \quad (14)$$

$$a'_i = \left(1 - \frac{a_i}{10}\right) \quad (15)$$

The value of O_i is always between -0.9 to +0.9. For a given application, $O_i = 0$ or $O_i < 0$ means that the application is already using the optimum interface and it needs not to be switched to an alternative one. For all $O_i > 0$, applications are switched in accordance with their O_is in a descending order starting with the one with the highest O_i.

4. Conclusions

In this paper, a context-aware decision model based on the AHP method has been presented. The model that takes into account context information from both the terminal and network side should be suitable for vertical HO decision making process in heterogeneous networks environment. The model is fully flexible and dependent on the number of chosen objectives that will determine the dimension of the pairwise comparison matrix for objectives as well as the number of such matrices for networks in terms of each objective. The decision model uses basic mathematical calculations that could be particularly suitable for embedded hardware in a mobile device. It is a service type based algorithm which means that the whole process is executed once for each type of running application, not for every running application. Thus, even in the worst case the total number of execution of the whole process is restricted to only three times while applications of all three types are running. This is particularly useful in minimizing processing time, handover delay, and CPU and memory usage. An efficient algorithm for vertical HO initiation, i.e. the invocation of the decision module has been regarded out of scope for this paper.

In future research, we intend to develop an efficient algorithm for vertical HO initiation and to extend the algorithm further taking into account user location and movement and location and coverage information of the reachable APs.

References

[1] T. L. Saaty, "How to make a decision: The Analytic Hierarchy Process", *European Journal of Operational Research*, 1990, Vol. 48, pp. 9-26.

[2] B. Hongyan, H. Chen, J. Lingge, "Intelligent Signal Processing of Mobility Management for Heterogeneous Networks", *Proc. of the 2003 IEEE Int. Conference on Neural Networks and Signal Processing*, 14-17 Dec. 2003, Vol. 2, pp. 1578-1581.

[3] P. Chan *et al.*, "Mobility Management incorporating Fuzzy Logic for a Heterogeneous IP Environment", *IEEE Communications Magazine*, Dec. 2001, Vol. 39, Issue 12, pp. 42-51.

[4] M. Stemm, R. H. Katz, "Vertical Handoffs in Wireless Overlay Networks", *ACM Journal on Mobile Networks and Applications*, Vol. 3, Issue 4, pp. 335-350.

[5] K. Pahlavan, et al., "Handoff in Hybrid Mobile Data Networks", *IEEE Personal Communications*, April 2000, Vol. 7, Issue 2, pp. 34-47.

[6] H. J. Wang, R. H. Katz, J. Giese, "Policy-Enabled Handoffs across Heterogeneous Wireless Networks", *Proc. of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA '99)*, 25-26 Feb. 1999, pp. 51-60.

[7] H. Bing, C. He, L. Jiang, "Performance Analysis of Vertical Handover in a UMTS-WLAN Integrated Network", *Proc. of the 14th IEEE Int. Symposium on PIMRC 2003*, Sep. 2003, Vol. 1, pp. 187-191.

[8] S. Balasubramaniam, J. Indulska, "Handovers between Heterogeneous Networks in Pervasive Systems", *IEEE Proc. on Communication Technology (ICCT 2003)*, April 2003.

[9] "Services and Service Capabilities", *3GPP TS 22.105 v. 3.6.0*, 1999.