Design and Evaluation of a Handover Decision Strategy for 4th Generation Mobile Networks

Wenhui Zhang, Juergen Jaehnert*, Klaus Dolzer

University of Stuttgart, Institute of Communication Networks and Computer Engineering
*University of Stuttgart, Super Computing Centre
E-mail: {zhang, dolzer}@ikr.uni-stuttgart.de, *jaehnert@rus.uni-stuttgart.de

***Abstract*– Fourth generation mobile communication systems are characterized by heterogeneous access networks and IP-based transport technologies. Different access technologies give users great flexibility in choosing services, which can be different in QoS support, business models, and service providers. Realization of seamless handovers to the best network section while considering QoS and AAAC (Authentication, Authorization, Accounting and Charging) calls not only for seamless handover protocols, but also intelligent handover decision strategies. The contribution of this paper is the design of a handover decision strategy in mobile All-IP networks to support seamless handover scenarios. Methods have been proposed to obtain QoS and AAAC information from candidate networks with minimum signalling overhead. New handover algorithms for wireless local area networks (WLANs) are also presented and evaluated with simulation results.**

I. Introduction

It is widely believed that IP will be the final means to integrate access networks of all technologies, wireless or wired in fourth generation (4G) mobile networks. The migration from traditional circuit switched networks towards a packet based wireless heterogeneous IP networks provides users great flexibility in choosing services, and also provokes a big pressure in both the network and mobile terminal design.

Such a 4G network architecture is currently being developed by the IST project Mobility and Differentiated Services in a Future IP Network (Moby Dick) [1]. This architecture includes the elements as shown in Fig. 1. Mobile end-systems are equipped with interfaces of W-CDMA (UMTS-TDD), WLAN (802.11b), and fixed networks (Ethernet). Access routers provide interfaces between the wireless and the wired core-network, and are enriched with enhanced IP capabilities. Network management servers are in the fixed network used for mobility management, QoS, security and paging issues, such as AAAC servers, QoS brokers, and Paging agents. The whole architecture is based on IPv6 exploiting all IPv6 specific support for IP based mobility management.

Fast Handovers, as described in [3] with some enhancements and context transfer techniques are used to provide seamless handovers. To provide QoS, the DiffServ model is adopted because of its high scalability and reduced signalling overhead. The association of DiffServ principles with the use

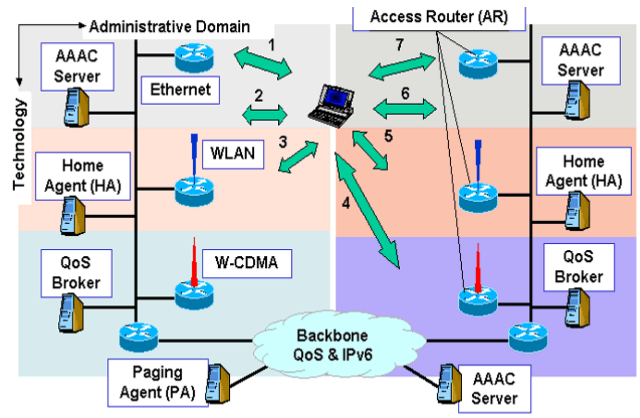

Fig. 1: Moby Dick network architecture

of QoS Brokers controlling a QoS domain provides a large-scale support for QoS. The AAAC support is based on the IRTF AAA Architecture [4] and enriched with auditing, metering and charging mechanisms.

A user profile stipulating various classes of services provided in different administrative domains and corresponding service prices are stored in the user's home network. Real-time services with quality comparable to traditional cellular networks should be generally accessible regardless of the technology and the access network and uninterrupted during a handover.

Such a heterogeneous system poses challenges in the handover design. These challenges come from the integration of the seamless handover with QoS support in an IP network while considering AAAC, and becomes more complicated if a potential large number of candidate networks are considered. This paper aims to design a handover decision strategy, that does not only provide users with the optimal network with minimum influence on the application, but also efficiently uses network resources.

The remainder of this paper is organised as follows: in Chapter II, handovers are classified, followed by the strategy as introduced in Chapter III. New handover algorithms for WLANs are introduced in Chapter IV.

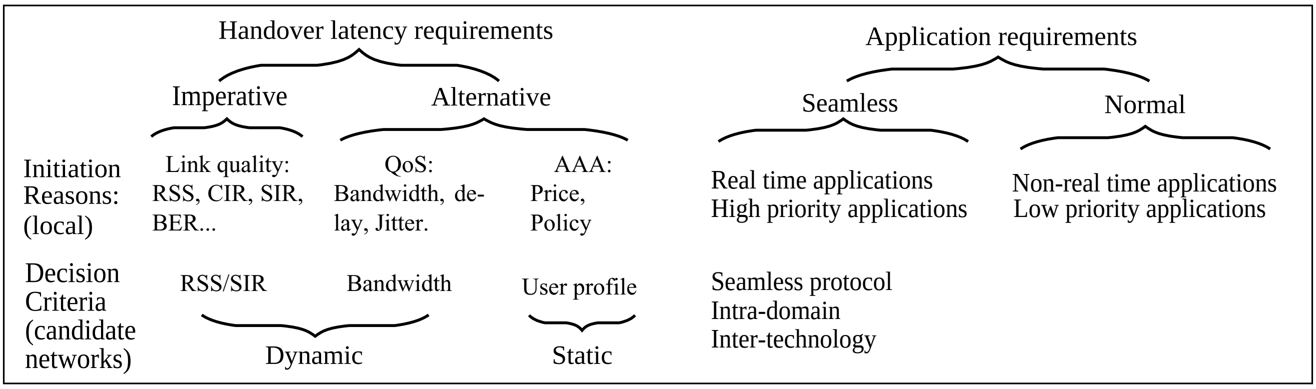

Fig. 2: Handover Classification

II. HANDOVER CLASSIFICATION

A. Mobile-controlled versus Mobile-assisted

Mobile-assisted handovers are used in 2G and 3G wireless networks [5], where mobile terminals send the measurement report to the network, and the network makes the handover decision. Handovers are usually carried out within an administrative domain, considering only one technology, and work efficiently in circuit switched networks. However, in IP-based heterogeneous networks, mobile-assisted handovers have some disadvantages.

Mobile-assisted handovers have disadvantages in case of inter-domain handovers. In principle, a security association (SA) is required between the communication partners in IP networks. This SA can be taken for granted within an administrative domain, but to establish it between different domains is not a trivial issue. Even if security associations between domains are established, the service profiles of a mobile user indicating the subscribed services can be different in different administrative domains. E.g. a user can have a certain service of premium class in domain A, but it does not entitle him to have the same service in domain B. When a handover is required between domains, the AAA information is usually not valid in the new domain and has to be refreshed from the home network of the user, which will complicate the overall signalling design and increase the handover latency. The technical difficulties also arise from the increased complexity in network entities if numerous handover decision criteria with divergent user preferences have to be considered. Mobile-assisted handovers also have operational difficulties. In order to find an optimal network for a user, user preferences have to be considered, such as service prices in different domains and personal preferences. User preferences need to be sent to the network by signalling over the air, which increases the signalling overhead in the air, and also induces latency. Moreover, a user may be unwilling to disclose such information to the network, and potentially does not trust the current network to find a cheaper network from its competitors. In addition, the competition between service providers may prevent a service provider from giving the business away to its competitors.

Regarding the disadvantages of mobile-assisted handovers, in this paper, we only consider mobile-controlled handovers, i.e. the mobile conducts the initiation and control of a handover. This strategy is more flexible and reduces the overall complexity in the network. However, mobile-controlled handovers require networks to disclose some network capability information to mobile terminals in order to choose optimal networks. Information such as bandwidth and capabilities to support certain services is transparent to mobile users in 2G and 3G networks. Whether network providers are willing to give users these information in 4G networks and how to properly protect these information are beyond the scope of this paper.

B. Handover Initiation and Decision Criteria

Handovers in 4G mobile networks are not only carried out in order to maintain a connection, but also to provide users better services and to meet individual requirements. A mobile terminal should make a handover decision based on user preferences automatically, and also allow a user to manually intercept the handover decision if desired.

In principle, handovers can be categorised as imperative and alternative handovers according to initiation reasons. Handovers due to low link quality are imperative, because both the handover decision and execution have to be done fast in order to keep on-going connections. Primarily, the received signal strength (RSS) measured from the access point and neighbouring access points are used for handover decisions, other criteria are also used, such as carrier to interference ratio (CIR), signal-to-interference ratio (SIR), bit error rate (BER), etc. [10]. Handovers, which are used to provide a user with better performance or to meet a particular preference, can be considered as alternative handovers, as shown in Fig. 2. These handovers can tolerate longer handover latency, and can be sub-divided into QoS related handovers, and AAA related handovers. E.g., a user might require more bandwidth to speed up a data transfer, or need a cheaper network to reduce the service cost.

Based on the duration of the validity of the information, these criteria can be further sub-divided into static and dynamic information. E.g., AAA information, such as the network domain, user service profile in a network and business model can be considered as static, which can be buffered or acquired before the handover. However, QoS information, such as the SNR, BER and the available bandwidth are very dynamic, and have to be updated continually, which implies hard timely constraints on signalling.

In order to make a handover decision that meets the respective handover requirements, certain information from candidate networks have to be retrieved. There are many criteria that can be used for handover decisions. Usually, it makes sense to combine one or more criteria for handover decisions. Therefore, in this paper, we consider the RSS/SIR, the available bandwidth, and the user profile for handover decisions. Moreover, application requirements play an important role in handover decisions, e.g., handovers for real time or high priority services need to be made seamless, which can only be supported within a domain or between different technologies [9].

III. Handover decision Strategy

In the following, we introduce a new handover decision strategy, whose key point is to obtain the candidate networks information.

A. *Pre-selection for Handover Measurements*

In 2G and 3G networks, handover measurements are controlled by the network. In UMTS system, prior to handover decisions, neighbouring cell parameters, such as frequencies and scrambling codes, which are needed for the quick determination of neighbouring cells, are sent to the mobile terminal by UTRAN/RNC [7]. Without prior knowledge, a mobile terminal has to scan channels in different frequencies to find an corresponding access point as foreseen in 802.11b WLAN. Tests have shown that scanning 13 channels in 802.11b WLAN takes up over 400ms, which accounts for most of the layer two (L2) handover latency. In principle, a mobile terminal can query the current access point to provide such information to assist handover measurements. However, promptly providing such information upon user requests sounds not practical if the information for a large number of access points is needed. This invokes the need to notify mobile terminals the potential service points' parameters prior to the handover process. The proper selection of networks for handover measurements from a potentially large number of neighbouring cells can reduce the amount of handover measurements and the decision complexity, which in turn lowers the battery drain.

We suggest the following pre-selection scenario: After the authentication of a user, the local AAA server sends the user a neighbouring cell list including physical parameters, administrative domains, cell identifiers, IP addresses and IP prefixes. With these parameters, the user can easily measure the neighbouring cells upon a handover request, and configure care-of-addresses (CoAs) for Mobile IP [2] if needed. The list can be sent as a value-added location based service provided by the UMTS system [8] to assist seamless handovers. If the size of the list is too large for cellular systems, it can be divided into small packets and transmitted after the registration process within a certain period of time. This is based on the assumption that a mobile user might not request a handover right after the registration. Alternatively, only the list of neighbouring cells in the immediate adjacent of a user is sent upon registration based on the location information, and it is refreshed when the user moves to a new location.

The advantage of this scenario is that it saves the uplink signalling from handover measurements. In the downlink, clearly, the more handovers requested by the users, the higher benefit in saving the bandwidth this scenario has. Mobile users can also easily authenticate cells during the measurement process and filter out the cells, which do not belong to their service providers. This is very important in finding access points, which are operating in the free frequency bands.

Here, an example is given showing how the pre-selection is accomplished based on handover requirements and the neighbouring cell list, which results in a significant reduction in the amount of the candidate networks to be measured. Suppose a mobile user has service contracts with three service providers: A, B and C, each providing the user with certain kinds of services from WCDMA and WLAN cells with different prices. The service profile and security association between the service providers are shown in Table 1, which is available in the home network and also stored in the mobile terminal.

Provider	Services	Price	Security
A	voice, video	low	no
B	voice, video	medium	C
C	voice	high	B

Table 1: Service profile and security association

Type	Services	Price	Seamless
Voice	A, B, C	not considered	B2, C1, C2
Video	A, B	A, BL1, BL2	BL1, BL2

Table 2: Handover pre-selection

Assume that in a certain area, each provider has two WCDMA cells and two WLAN cells denoted as AC1, AC2, BC1, BC2, CC1, CC2, AL1, AL2, BL1, BL2, CL1, and CL2. If the mobile user is moving out from the cell BC1 with an ongoing voice call, an imperative seamless handover from BC1 to other cells has to be made. Usually it does not make sense to handover to a WLAN cell to continue a connection, so in the final selection, only B2 and C1 and C2 are suitable. If the user has a streaming video application in BC1 and is looking for a cheaper network, a seamless handover is also needed. Considering the allowed service and the price, only A, BL1 and BL2 are applicable, furthermore, cells from A are eliminated because seamless handovers are not possible. The selection results are listed in Table 2.

B. Candidate Network Capability Discovery

Usually, measurements of neighbouring cells provide a mobile terminal with cell identifiers and the signal strength, on which L2 handovers are based. In an IP-based network, network prefixes of Access Routers (ARs) are needed for a mobile terminal to configure the CoA. Using standard methods from Mobile IP [2] to configure the CoA can easily take several seconds. With the registration scenario mentioned above, a mobile terminal can compare the cell identifier with the neighbouring cell list obtained upon registration and get the IP prefix. The CoA can be acquired by using stateless auto configuration avoiding Duplicate Address Detection, as mentioned in [9]. An alternative way to speed up CoA configuration is to embed IP addresses and prefixes in L2 broadcast. In 802.11b WLANs, the typical beacon interval is around 100ms, if 32 bytes AR IPv6 address and prefix are sent with each beacon in a 2Mbps date rate network, the overhead is only about 0.2%. In WCDMA, each cell has a 30kbps broadcast channel which neighbouring cells can listen to. If 32 bytes IPv6 address and prefix are broadcast every measurement interval 200ms [7], the corresponding bit rate is only 1.28kpbs.

As mentioned previously, static information from candidate networks can be acquired prior to the handover process. However, dynamic candidate network information, such as QoS capabilities in a new network has to be obtained in real time. IETF Seamoby working group has provided a Candidate Access Router Discovery (CARD) protocol [6], where a mobile terminal can perform CARD direct with a new access router if it can simultaneously connect to the new and old access router. However, performing CARD directly with the new access router requires a mobile terminal to configure a new CoA and make registration in the new network, which takes a lot of time. Moreover, if there are two new candidate networks of the same technology, the mobile can only connect to them in sequence, and the information retrieved from the first network will be obsolete for the handover decision.

Retrieving capability information from candidate networks via the current AR has scalability problems, because ARs, AAA and QoS servers would be very busy handling CARD requests, and the load in air also suffers from the required signalling overhead. In addition, the security association between ARs in different domains is required in order to perform CARD, which is extremely difficult if it is needed between each pair of ARs in different domains. An easy solution for the security association is to establish security associations only between AAA servers belonging to one domain, and the CARD is transferred between AAA servers. However, this approach will complicate the CARD signalling.

Even if the security association is established between ARs in different domains, the difference in the user profile between domains can hinder the CARD process, which is outlined in Fig. 3. Suppose a mobile terminal (MT) connecting to the old AR (oAR), which the MT is about to leave, needs the capability information from the new AR (nAR1) in the same domain,

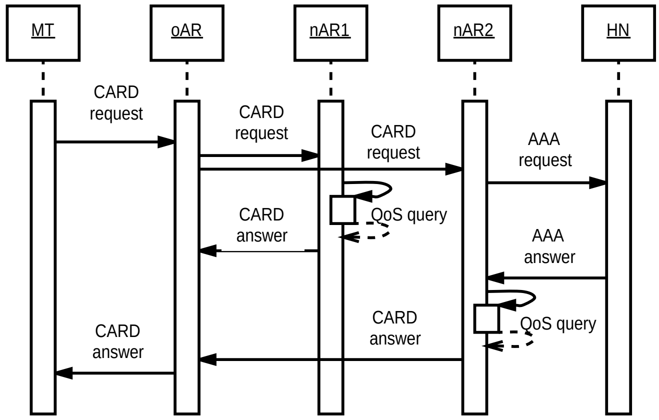

Fig. 3: CARD signaling

and an additional new AR (nAR2) in a different domain. Within a domain, the oAR can transfer the user profile in the CARD request to nAR1, and nAR1 needs only to check the QoS capability. But the user profile in the oAR is not valid in nAR2, and the new user profile has to be retrieved from the user's home network (HA), which will delay the whole CARD process.

We assume that the air interface is the bottleneck of the wireless network and suggest use L2 broadcast to assist QoS capability retrieving from candidate networks. The available capacity of a cell is embedded in L2 broadcast, from which, a mobile user can calculate if the cell has enough capability to support a required application. E.g. in UMTS, admission control can be based on throughput [7]. A new requesting user is admitted if the calculated load does not exceed a pre-defined threshold. The load difference due a new request can be calculated by (1):

$$\Delta L = \frac{1}{1 + \frac{W}{v \cdot E_b / N_0 \cdot R}} \tag{1}$$

where W is the chip rate, R is the bit rate of the application, E_b / N_0 is the assumed energy per user bit divided by noise spectral density, and v is the assumed voice activity. For a certain application, a user can deduce the load increase in a cell and calculate if he can be admitted based on the load of the new cell. In principle, if the network and the mobile terminal use the same algorithm and the mobile terminal is provided with the accurate information, they should have the same result for admission control.

C. Candidate Network Selection

Based on the acquired information, candidate networks are checked against the criteria, if there are more than one networks which satisfy the handover requirements, a selection has to be made to choose the best one. All the handover decision criteria from the candidate networks can be weighted, or fuzzy logic can be adopted to combine the different criteria to find the best network [11].

IV. HANDOVERS IN WLAN

A short latency in the handover decision is beneficial in handovers from a cell in the cellular network to a WLAN cell and also the other way round, in order to take advantage of the large bandwidth and the expected low price in WLAN, and also to avoid dropping a connection. The handover decision in WLAN is based on the received signal strength (RSS) of beacons from the access points [10], which necessitate a thorough understanding of the characteristics of the RSS of WLAN beacons. Measurements of the RSS of beacons show that the RSS has dramatic variations from the average value. Based on the theoretical analysis and corroborated by the measurement data, the RSS variation from average power can be modelled as correlated Gamma random variables [12].

A new handover decision algorithm is designed based on Linear Regression, which is a statistical method often used to make predictions about a signal value from sample data. When the RSS from the current access point drops below a threshold, the linear regression is started to calculate the trend of the RSS for both the current and new access points for a minimum time *T*. The handover is made when the predicted RSS of the new access point is greater than that of the old access point over a hysteresis margin.

Simulations of simple direct movements with constant speed have been carried out for the gentle path loss environment and the movement around a corner. The handover decision using the regression algorithm are compared with decisions using the traditional hysteresis algorithm [5]. To avoid inaccurate decision using the traditional hysteresis algorithm, moving average from previous sample data received within time *T* and a first-order low-pass filer with the weighting factor *b* are used to remove the rapid variation in the RSS. A handover is successful if it is only executed once during a movement, and only handovers with success probability greater than 99% are considered. The handover decision latency relative to the theoretical handover time in the gentle path loss environment is plotted in Fig. 4, which shows that the regression algorithm requires less decision time than the hysteresis algorithm using moving average and first-order low-pass filter. However, for handovers around a corner, the hysteresis algorithm with first-order low-pass filter is preferred, because it has low latency compared with the moving average and the regression algorithm.

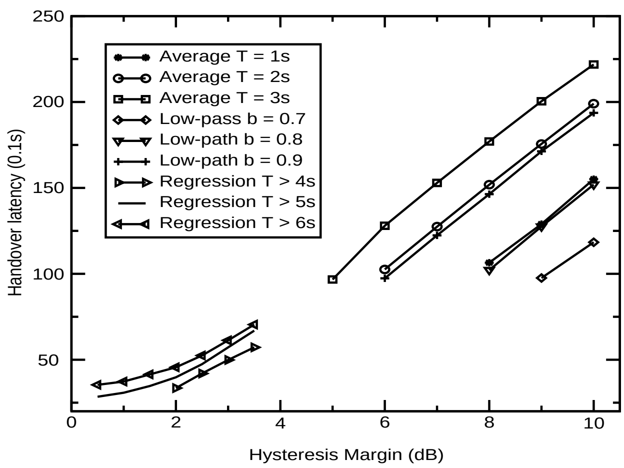

Fig. 4: Handover latency comparison for gentle path loss environment

The combination of the regression algorithm and the algorithm using the first-order low-pass filer with hysteresis margin can be used for WLAN handovers. A handover is carried out when either of the algorithms indicates that a handover is necessary, which can reduce handover latency in the gentle path loss environment and also avoid dropping a connection when a mobile user moves around a corner.

V. CONCLUSIONS

This paper outlines a handover strategy for 4G wireless networks. Handovers are classified as imperative and alternative handovers, and decision criteria are classified as static and dynamic information. Pre-selection of candidate networks can significantly reduce the number of networks to be measured. Acquiring candidate network capability information can be carried out with the help of L2 broadcast, which simplify the signalling. In the end, a new handover algorithm using the linear regression is presented in order to reduce the handover latency in WLAN.

ACKNOWLEDGEMENT

The work presented was partly funded by the EU project IST-2000-25394 “Moby Dick”.

REFERENCES

[1] Moby Dick: Mobility and Differentiated Services in a Future IP Network, http://www.ist-mobydick.org, Jan. 2003.

[2] D. Johnson, C. Perkins, J. Arkko, “Mobility Support in IPv6, IETF draft-ietf-mobileip-ipv6-19.txt,” work in progress, Oct. 2002.

[3] Rajeev Koodli, “Fast Handovers for Mobile IPv6,” IETF draft-ietf-mobileip-fast-mipv6-05.txt, work in progress, Sep. 2002.

[4] C. de Laat, et al., “Generic AAA Architecture,” IETF RFC 2903, Aug. 2000.

[5] Nishth D. et al., “Handoff in Cellular systems,” IEEE Personal Communications, Dec. 1998.

[6] CARD design Team,“Candidate Access Rotuer discovery,“ IETF draft-ietf-seamoby-card-protocol-00.txt, work in progress, Oct. 2002.

[7] Harri Holma, Antti Toskala, “WCDMA for UMTS,“ John Wiley&Sons, Ltd., 2001.

[8] 3GPP TS 25.305 V5.4.0, “Stage 2 functional specification of User Equipment (UE) positioning in UTRAN,” Mar. 2002.

[9] M.Liebsch, et al., “Solutions for Ipv6-based mobility in the EU project Moby Dick,” World Telecom Congress, Paris, Sep. 2002.

[10] Kaveh Pahlavan, et al., “Handoff in Hybrid Data Networks,“ IEEE Personal Communications, Apr. 2000.

[11] P. Chan, et. al. “Mobilty Management Incorporating Fuzzy Logic for a Heterogeneous IP Environment,“ IEEE Communication Magazine, Dec. 2001.

[12] Zhang W. “Modeling of WLAN Beacon Signal Strength Measured in an Indoor Environment,“ Internal report, IKR, University of Stuttgart, 2003.