Scenarios and Vertical Handover Strategies in Heterogeneous Networks Environment

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*–In the future era of communication, realizing ubiquitous and seamless mobility in heterogeneous networks environment is of prime importance. Users will need to select the best option available from diverse networks and service scenarios. New strategies need to be introduced in order to handle complex scenarios that would enable support of context-awareness. This paper investigates different communication and vertical handover scenarios in heterogeneous networks environment and, with respect to those, analyzes possible strategies and mechanisms that would need to be taken into account for realizing ubiquitous mobility. It also identifies a context model for the terminal as well as for the network side.**

1. Introduction

Emergence of diverse access technologies, e.g. WLAN, Bluetooth, 3GPP cellular networks (GSM, GPRS, UMTS), DVB-H, etc, would result in heterogeneous *overlaying* infrastructure. Users should be given the freedom to roam globally among multitude points of attachment of different access networks as per their service requirements.

Mobility across heterogeneous networks would result in new challenges that traditional mechanisms like *horizontal handover (HO)* would fail to overcome. The new challenges would be context management, mobility management, QoS management, authentication and authorization management, and accounting management. A variety of current and future wired and wireless networking technologies could be transformed into seamless communication environments through deployment of *context-awareness* in different management processes. Therefore, at the very outset, it is needed to outline a detailed context model.

Additionally, it is worth analyzing thoroughly all possible communication scenarios in heterogeneous networks environment in order to find out possible work phases that would need to undertake possible management schemes. What critical issues and tasks could be relevant in different scenarios and phases, what could be the participating entities, and when and how to participate are some issues need to be highlighted. Moreover, the needs and possible means of incorporating context-awareness, based on a well-defined context model, in different phases especially in HO mechanisms should be investigated.

The rest of the paper is organized as follows. Section 2 highlights the related work. Section 3 outlines a detailed context model. Section 4 analyzes possible scenarios in heterogeneous networks environment from both terminal and network sides. Finally, section 5 concludes the paper.

2. Related Work

References [1] and [2] describe a platform and algorithm for realizing context-aware vertical HO but they concentrate mostly on the upper-layer (service-layer). A clear formulation of strategies in different scenarios involving relevant layers on the terminal side and interworking of layers are not addressed. Reference [2] also highlights a detailed context model, however, many relevant context elements are still not identified, whereas reference [1] lacks a detailed context model. References [3] and [4] describe a context management platform and generic methods for context representation and transfer. Both of them do not address a detailed context model and use of context-awareness in strategies developed on analysis of possible communication scenarios.

3. Context Model

According to [5] *context* means the information on the current state of a service required to re-establish the service on a new subnet without having to perform the entire protocol exchange with the mobile host from scratch. Information or intelligence residing on the terminal side as well as on the network side could be collectively called as *context information*. Context information would be distributed among different entities like the mobile host (MH), the MH's home network, the current visiting network, and the future visiting network.

Context information could be classified based on their frequency of changes and based on their placement. In the former case, context information 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 the contexts that change quite frequently and may loose accuracy over time are dynamic context information.

3.1. Context Model for MH

Static model: **Device capability** would include display size, resolution, battery life, RAM, processor speed, and multimode capability. **Services** could be grouped as **service types** (conversational/real-time, interactive, and streaming services), **QoS requirements** (service precedence, delay, mean throughput, peak throughput, and reliability), and **other service requirements** (direction of communication, always-on/on-request, error tolerance, unicast/multicast/broadcast, symmetry/asymmetry). **Subscription profiles** would consist of subscriber's universal identity, chosen services from providers, subscription types and options (low/normal/high priority group, individual/corporate subscriber, billing options, privacy options, and service options). **User preferences** could be grouped as **interface preferences** for multimode terminal and **service preferences** (precedence of service types, billing constraints, QoS preferences).

Dynamic model: **Running application types** would define the service profile of currently running applications. **Current user location** could be obtained through *Global Positioning System (GPS)*. **Reachable access points (APs)** would identify currently available networks and addresses of the reachable APs. **Device status** would consist of currently active interface and battery status. **Dynamic user profiles** would define user identity assigned temporarily by visiting network, addresses of entities supporting the MH, etc. Lastly, **expected communication duration** would describe the expected time duration of a session a user is about to initiate.

3.2. Context Model for Network

Static model: **Subscription profiles** would be similar to that described for the MH. **Network services** would define services and service profiles provided by network. **Network capability** would include network coverage, mobility support, transmission mode (circuit-switched/packet-switched), traffic characteristics (unicast/multicast/broadcast), availability (always-on/on-request), available bandwidth, error tolerance, real-time capability, QoS support, security, etc. **Service provider's profiles** would consist of provider's identity, policies, charging models, roaming agreement models (on-demand, settled before, or mixed), etc. **Network location information** would include location information of the serving AP as well as of the neighboring APs. **Neighboring APs** would define addresses of neighboring APs.

Dynamic model: **Current QoS parameters** would define the current status of network QoS parameters. **Traffic load status** would identify the current traffic load in the network. **Dynamic user profiles** would be similar to that described for the MH.

4. Communication Scenarios

In the communication scenarios in heterogeneous networks environment, the *primary participating entities* in all relevant tasks involved in different phases should be the MH and the serving network. The *secondary participating entities* should be the networks reachable from the MH and the neighboring networks of the serving network. Considering the means of involvement of the primary entities in different phases all relevant tasks could be analyzed and classified based on four different cases. In the **mobile-initiated case t**he MH would initiate different tasks, whereas in the **network-initiated case** the serving network would initiate different tasks. The MH would execute different tasks in the **mobile-controlled case**, whereas the serving network would execute different tasks in the **network-controlled case**. Additionally, supplementary tasks in the above cases could be further grouped together into two categories. **Assistance from terminal side** would include the MH's assistance in different tasks initiated and controlled by the serving network. Similarly, **assistance from network side** would include the serving network's assistance in different tasks initiated and controlled by the MH.

In the following scenarios, 3GPP cellular and DVB-H networks have been assumed to have ubiquitous coverage whereas other access technologies like WLAN or Bluetooth have been assumed to be overlaid as *hot-spots*. Referring to different kinds of management mechanisms, as described earlier, location management and accounting management have been excluded from the discussion here on.

4.1. Scenario 1: MH is Stationary and Not Communicating

In this scenario, it is not required to take any action from any side pertaining in the scenario.

4.2. Scenario 2: MH is Stationary and Starts a Session as Initiator

In this scenario, the MH would be under coverage of more than one access networks, as shown in Fig. 1. The important issues here should be like how to know about user preferences, application requirements, and available networks, which network the MH should connect to and how to connect, what to do if the MH fails to connect due to some reason, and finally, how to initiate the application.

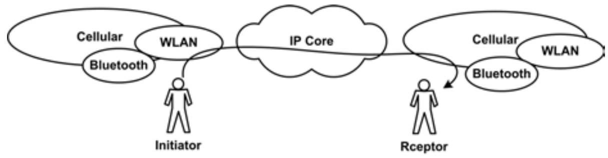

Figure 1. Scenario 2 and scenario 3

Tasks in terminal-initiated and -controlled case:

The grouping and mapping process should be a general task applicable to all scenarios. Grouping defines that static and dynamic contexts on the terminal and the network sides should be assorted according to their relevance to different mechanisms. Possible mechanisms could be application startup and re-initiation, HO prediction, HO initiation, HO decision making, HO execution, connection establish, etc. Mapping means establishing relationships among different groups of data intelligently so that they can be translated into a minimum data set compatible to specific task requirements. Mapping could be performed among different context parameters within a group, or among different groups, or among user preferences and different context groups. User preference inputs should be taken in as much user-friendly and user-perceivable way as possible. Context groups should be stored in databases in specific formats. Context format should be such that different groups are comparable with similar groups of information coming from other entities (e.g. network side). Stored data should be easily accessible by different software modules. Dynamic group of data should be dynamically re-writable. The mapping and grouping process is illustrated in Fig. 2. *User preferences*, *other service preferences*, *device profiles*, and *user profiles* are mapped into defining *application profiles* for each of the three types of services identified individually by *application type*. Any application running inside the MH would have predefined *application type*. Thus, any application, during runtime, could be paired with individual set of *application profiles* based on its *application type*.

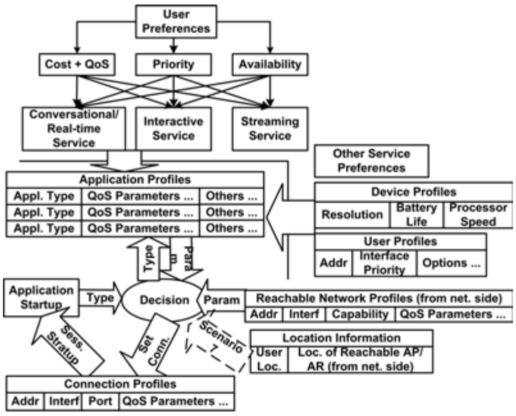

Figure 2. Mapping and grouping of contexts at MH (scenario: session startup)

The **monitoring and probing** process means that the multimode terminal would monitor (layer-2 or layer-3 monitoring, or both) each of its interfaces for shared contexts advertised by available networks, or the MH would utilize layer-2 or layer-3 probing for context information from reachable networks. The MH may also involve some intelligent method to measure, specially, network QoS parameters from available layer-2 or layer-3 signal as such contexts would not be advertised explicitly by the networks. Context information of each available network should be stored in the MH in specific format (*Network Profiles* in Fig. 2). Required criteria of the stored network data have already been mentioned in the grouping and mapping process. If multiple sets of network context information are provided by a single network then each set of context information should be treated as if they represent different networks, although interface and network IDs should be the same (concept of *virtual network*). The **network selection** process would deploy a decision module with intelligent algorithms in order to select appropriate network based on context information on the terminal as well as on the network side. It would need data retrieval from different databases. The decision module would be triggered by application startup (see Fig. 2), HO prediction, availability of new network, degradation of current radio-link, etc. In case of multiple applications running together, intelligent resource management algorithm like *load-balancing* should be employed while choosing a network. The decision module would also consider current device status like battery power before deciding whether to serve the service request. The **connection** procedure should include security mechanism, if not yet authenticated, and connection setup with the selected network including registration, if not yet registered. There should be provision (*roaming agreement*) of connecting through a provider's network that is not the MH's service provider. During connection phase QoS management mechanisms (resource negotiation and reservation) with the serving network should also take place. Therefore, a QoS framework would be needed to guarantee negotiated QoS throughout the session. An appropriate transport protocol (UDP, TCP, RTP, SCTP, etc) would need to be chosen in this phase. Connection procedure would be based on appropriate mobility protocol at the network-layer. Relevant contexts of an established connection, used for session startup, should be stored (*Connection Profiles* in Fig. 2). Home registration, if required, should take place after the session has been started. Appropriate **backup policy** would be needed in case it is not possible to connect to the selected network. Possible backup solution could be re-trying, application shutdown due to unavailability, or selection of alternative network, if already available (next best-matched network), or search for alternative network, if already not available, etc. Finally, the **application startup** process would commence applications based on negotiated resources and connection contexts already stored.

Assistance from network side:

The **context management** would include context gathering and distribution. Each available network would inform of its own context data directly to the MH (advertisements or probing), or the serving network would gather contexts from neighboring networks and inform the MH. In the latter case, the MH should also verify which networks are reachable, because not all neighboring networks of the serving network would be reachable by the MH. The serving network should take part in the **QoS management** mechanisms with the initiator based on its policies (resource management and admission control). Finally, the serving network should, if required, take part in **security management**, and **registration** and **connection** mechanisms with the initiator. It should also enable **session establishment** with the receptor.

4.3. Scenario 3: MH is Stationary and Starts a Session as Receptor

In this scenario, the MH would be under coverage of more than one access networks, as shown in Fig. 1. The important issues here should be like should the MH accept the session request and how to take part in connection establishment.

Tasks in terminal-initiated and -controlled case:

The **connection acceptance decision** algorithm should decide whether to accept a connection request based on available resources. The **connection** procedure would be based on negotiation with requesting peer. Relevant mechanisms involved in the process should be similar to those described in section 4.2 (scenario 2), except that the receptor is assumed to be already registered with its serving and home network.

Assistance from network side:

The serving network should take part in **QoS management** and **session establishment** with the receptor and the initiator, respectively.

4.4. Scenario 4: MH is Stationary and within a Session

In this scenario, the MH would be under coverage of one or more alternative networks (WLAN, Bluetooth, etc) while having an ongoing session through cellular or DVB-H network, as shown in Fig. 3. The important issues here should be like how to ensure that always the best connection is available, how to select an alternative connection, if available, how to perform HO, if required, what to do if the MH fails to connect due to some reason, and finally, how to re-establish the session, if required.

Task in terminal-initiated and -controlled case:

The **grouping**, **mapping**, **monitoring**, and **probing** processes should be similar to those described in section 4.2 (scenario 2). Additionally, it would also be needed to monitor the radio link of the serving network. During **network selection** process a decision module, as described in section 4.2 (scenario 2), would select an appropriate network. The module would be triggered by degradation of radio link of existing connection, or availability of alternative networks. The **vertical HO** procedure would be initiated if the decision module indicates that switching to an alternative network is necessary (mobile-initiated and controlled HO). The HO procedure would be based on appropriate mobility protocol at the network-layer (inter-domain HO, macro mobility). The **connection**, **session re-establishment** and **backup** procedures, as described in section 4.2 (scenario 2), would follow the HO procedure.

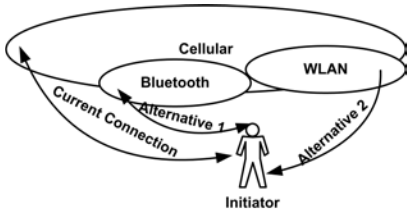

Figure 3. Communication scenario 4

Assistance from network side:

All required assistance from the network side should be similar to those described in section 4.2 (scenario 2). Besides, assistance would be required during possible **soft HO** mechanism. The serving network would provide support for discovering contexts (addresses and others) of candidate APs/ARs on request from the MH. Also, the serving network would provide means of transferring contexts (security context, negotiated QoS data, etc) to the future AP/AR as requested either by the MH or by the AP/AR.

Tasks in network-initiated case:

The **context gathering** process (layer-2 or layer-3 process, or both) would enable the MH to forward all relevant contexts (*application profiles*, as shown in Fig. 2) to the serving network as soon as it is registered. The serving network should also be aware of each application type currently running on each visiting MH (see Fig. 4). It should also gather context information of the networks that are currently reachable by the MH. The MH would transfer contexts of its reachable networks, or the serving network would collect them when the MH sends information about the reachable networks. Relevant contexts on the terminal or network side could be grouped according to the idea described in section 4.2 (scenario 2). The network should maintain a unique set of context data for each of the hosts visiting the network. The **monitoring** process, in this case, means that the serving network would monitor each visiting MH (layer-2 or layer-3 monitoring, or both). The **network selection** process would deploy a decision module similar to that described in section 4.2 (scenario 2). Individual decision module would be needed for each visiting MH, or a centralized one for better scalability. Efficient scheduling mechanism should be employed in case of centralized decision module so that no MH should wait too long for decision in scenarios where multiple decisions would need to be processed simultaneously for serving multiple MHs. The decision module would be triggered by availability of new network, degradation of radio-link, etc. The serving network should employ intelligent resource management algorithm like load-balancing and initiate a vertical HO by informing the MH.

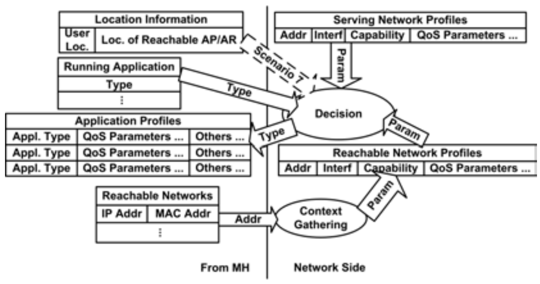

Figure 4. Context gathering and grouping on network side

Assistance and control from terminal side:

The MH should transfer relevant context to the serving network as discussed above. It should execute *vertical HO* with appropriate *backup* and *session re-establishment* mechanisms as soon as it receives HO triggering from the serving network. The serving network should assist it in different mechanisms in the same way as described terminal-initiated and -controlled case.

4.5. Scenario 5: MH is Moving and Starts a Session as Initiator

This scenario should be the same as scenario 2.

4.6. Scenario 6: MH is Moving and Starts a Session as Receptor

This scenario should be the same as scenario 3.

4.7. Scenario 7: MH is Moving and within a Session

In this scenario, the MH would have an ongoing session over some access network while moving. The movement of the MH may result in three different cases discussed below:

Case 1 – Availability of alternative networks:

Let us consider the critical scenario as shown in Fig. 5, where the MH arrives at the intersection region of three different networks. According to the decision module (either on terminal or network side), the MH has three options – (i) move to network-B (1st preference) (ii) stay in network-A (2nd preference) (iii) move to network-C (3rd preference). If the MH considers contexts like user location and movement, location of APs/ARs/BTSs, network coverage, etc it should decide moving towards the coverage of the least preferred network (network-C). Because, if the MH decides to switch to network-B due to its direction of movement there must be another vertical HO in very short time, which would incur unwanted overhead. On the contrary, the MH may want to remain connected through network-A as long as possible. Then for seamlessness it is also critical to initiate the HO process well before the time when actual HO is mandatory (*proactive HO*). The *time window (T)*, as shown in Fig. 5, represents the time instance when HO should be initiated in advance. T must not be long enough to initiate unnecessary HO and must not be short enough to create unwanted interruption due to unavailability of access network for late HO initiation or because the HO has not yet finished.

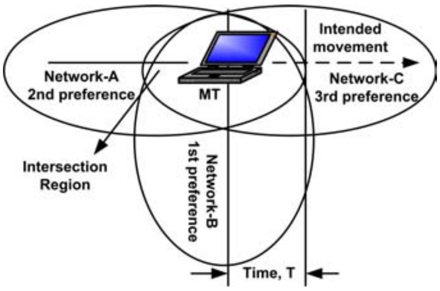

Figure 5. Context-aware vertical HO critical scenario

Required issues and tasks should be similar to those described in section 4.4 (scenario 4), except that additional location contexts would need to be taken into explicit consideration. In the terminal-initiated and -controlled case, contexts like location and coverage of the reachable APs/ARs/BTSs would need to be transferred from the reachable networks to the MH whereas in the network-initiated and -controlled case, contexts like user location, and location and coverage of APs/ARs/BTSs that are reachable by the MH would need to be transferred to the serving network from the MH. The MH and the serving network, respectively, would predict user movement and use it in the HO decision in the former and latter cases.

Case 2 – Movement within a network domain:

In this case, there would not be any alternative networks other than the currently serving network. For example, the MH may need to switch between BTSs of a cellular network.

Terminal-initiated and -controlled case: The decision should be fairly simple as only one network is available. Seamless horizontal HO with efficient micro mobility functionalities (HO between BTSs/APs/ARs) would take place in this case. Intelligent decision module should ascertain radio link characteristics and optimum value of T for determining when to initiate HO.

Network-initiated and terminal-controlled case: The serving network would take the HO decision and inform the MH where and when to switch. The HO execution should take place at the MH.

Case 3 – Loss of coverage of current network:

In this case, if there are alternative networks available the MH should select (or the network may select) a preferable network to switch to, before it is disconnected from the current network due to lack of coverage. In this sense, this case is similar to case 1. In case no alternative networks are available then the ongoing session needs to shut down, however, considering almost ubiquitous coverage of cellular networks this could be considered as nearly impossible scenario.

4.8. Comparison between Different Cases

TABLE 1

COMPARISON CHART – TERMINAL-INITIATED & -CONTROLLED AND NETWORK-INITIATED & TERMINAL -CONTROLLED CASES

Terminal-Initiated & -Controlled Case	Network-Initiated & Terminal-Controlled Case
1. Relatively few contexts would need to be transferred from the network side. Less use of radio resources.	1. A large amount of contexts would need to be transferred from the terminal side. Wastage of radio resources.
2. Individual, simple context management and small data storage.	2. Centralized, complex context management and large data storage.
3. Individual monitoring and independent actions.	3. Central monitoring and scheduled actions for hosts.
4. Should be lot faster and efficient.	4. Would incur additional delay due to huge amount of data processing.
5. Should be scalable.	5. Would not be scalable. Amount of data and workload would limit scalability in spite of having adequate amount of other resources.
6. Should be flexible. Each device would have its own algorithm, software architecture and modules.	6. Would be less flexible.
7. In case of location context, it would guarantee individual privacy.	7. In case of location context, it would not guarantee individual privacy.
8. Would hinder efficient resource management (e.g. load-balancing) at the network side.	8. Resource management policy based decision would be realized easily.

5. CONCLUSIONS

In this paper a detail analysis of context model and possible communication scenarios in context-aware vertical handover in heterogeneous networks have been discussed. The scenario analysis identifies all relevant issues, tasks, and requirements. This analysis would provide the fundamental background for design and implementation of different mechanisms in order to achieve ubiquitous mobility in heterogeneous networks.

REFERENCES

[1] S. Balasubramaniam, J. Indulska, “Handovers between Heterogeneous Networks in Pervasive Systems”, *IEEE Proc. ICCT 2003*, April 2003.

[2] S. Balasubramaniam, J. Indulska, “Vertical Handover Based Adaptation for Multimedia Applications in Pervasive Systems, *Joint Int. Workshop on Interactive Distributed Multimedia Systems and Protocols for Multimedia Systems (IDMS/PROMS 2002)*, 2002. Lecture Notes in Computer Science, LNCS 2515, pp. 61-72.

[3] Q. Wei, C. Prehofer, “Context Management in Mobile Networks”, *Proc. of ANwire Workshop*, Paris, France, November 18, 2003.

[4] C. Prehofer, N. Nafisi, Q. Wie, “A Framework for Context-Aware Handover Decisions”, *IEEE proc. of PIMRC 2003*, Sep. 2003, Vol. 3, pp. 2794-2798.

[5] J. Kemp *et al.*, “Problem Description: Reasons for Performing Context Transfers between Nodes in an IP Access Network”, *IETF Network Working Group*, RFC 3374.

[6] “Services and Service Capabilities”, *3GPP TS 22.105 v. 3.6.0*, 1999.

[7] Frank A. Zdarsky, Jens B. Schmitt, “Handover in Mobile Communication Networks: Who is in Control Anyway?”, *Distributed Computer Systems Lab*, University of Kaiserslutern, Germany.