Mobility control beyond 3G systems

M. Teughels, E. Van Lil*, A. Van de Capelle

K.U. Leuven, ESAT - Telemic,
Kard. Mercierlaan 94, B-3001 Heverlee, Belgium
http://www.esat.kuleuven.ac.be/telemic

Abstract

This paper deals about the issues of mobility control in mobile networks and the consequences for future generation networks. Mobility involves the update of the location information of the mobile user, as well as the redirection of the dataflow. Especially the latter is very challenging for network developers. Answers to this problem have been proposed for connection oriented networks, but for connectionless networks there is still no satisfying proposal, yet. Hence this has consequences on architectural options for future networks in general, taking the impact of mobility in consideration.

1 Introduction

Mobile services gain in importance in today's networks [1]. Providing mobile services however involves many aspects. There is the issue of the wireless link and the multiple access, as a wireless channel is by nature a broadcast channel. For these issues answers have been proposed extensively, and standards have been created [2]. On the network layer, the mobility has several implications too. The security is one of them. Also the location management and addressing has to be solved. Close to these topics is the rerouting of on-going communication due to terminal mobility. This is the responsability of the mobility control protocol. Evolutions in these protocols will be discussed in this paper.

Second generation mobile networks, of which GSM is the most widespread system, are focused on voice services [3]. As the primary service of these networks is point-to-point voice services, a connection-oriented architecture seems obvious. When the user migrates through the network, the connection is rerouted. For the location service, a HLR (Home Location Register) keeps up to date information on the mapping of the terminal address (the phone number) and the geographical position (the location area address). Both functions are separated: the rerouting of the data-flow is done with a hand-over for the connection, while the location update means informing the HLR.

Some extensions of these second generation systems, like WAP and GPRS are currently added to support data connections: the object is to offer connectivity to the Internet [4]. Allthough the Internet is a connectionless best-effort network, the WAP and GPRS access to these networks is offered in a connection-oriented fashion. Hence the user mobility remains limited to the rerouting of these connections. Also mobile routing is avoided: the mobile terminal does not change its IP address when roaming through the network, but this address is mapped to its connection at start-up and only the connection has to be rerouted.

Hence in second generation systems, allthough the service can be a best effort datagram service, the architecture is a connection-oriented one. The status of the mobility control protocol and the architecture of the Third Generation systems will be discussed next. This discussion serves as imput for the study of the mobility control issues in networks beyond Third Generation systems.

It is expected that the mobile access to the Internet will outnumber the fixed access: from that moment on the mobile network will be the driver of the Internet [1, 5]. Hence the mobility issues can have implications on the evolution of fixed network architectures as well.

*Emmanuel is also with FWO, Fund for Scientific Research, Flanders, Belgium.

2 Mobility Control in Third Generation

Third generation systems, grouped in the IMT-2000 proposals, promise integration with the current Internet. The Internet is a best effort datagram service, where IP is the defacto standard network protocol. In IP the destination address is used to route the datagram from source to destination. This address is therefore inserted in the header of all datagrams and inspected by all routers on the path from source to destination.

The address is composed of a network and a host address, and the network address is used to route the datagrams. This system is not prepared to deal with individual hosts connecting to other networks. Also the latency of the routing protocols to update route changes is substantial and prevents the routers in the network to have an updated view on the location of all hosts. Finally the network address is used for both terminal identification and geographical location. Different proposals to overcome these problems have been published recently [6].

2.1 Macro-mobility

2.1.1 Mobile IP

MobileIP tackles the problem of the double signification of the IP address of a host: the terminal identification, as well as the route computation [7, 8]. Therefore a mobile terminal obtains a care-of address, with a meaningfull routing signification, but that can be dynamically assigned to different hosts.

All terminals in the network can however not be aware of the roaming of the terminal. Therefore in the home network of the terminal (where the IP address coincides with its routing signification), a HA (Home Agent) intercepts all datagrams for the mobile host. Whenever the latter migrates to a new network, it obtains a new care-of address and informs the HA of this new address. Hence this HA is able to forward the datagrams to the mobile host using IP tunneling.

In figure 1 this configuration is plotted. In the figure a mobile host is connected to a visited network. It obtains a care-off address from a local Foreign Host, and informs its Home Agent, in its home network, of its care-off address. A corresponding host can send datagrams to the original IP address. These frames arrive are routed to the home network, where the Home Agent intercepts them and tunnels them towards the mobile terminal. The tunnel can be used for the dataflow sent by the mobile terminal as well, but the mobile terminal can also directly send the frames to the corresponding host (the dashed lines show the alternatives). In the latter configuration, triangular routing is introduced: up- and downstream are forwarded along different routes.

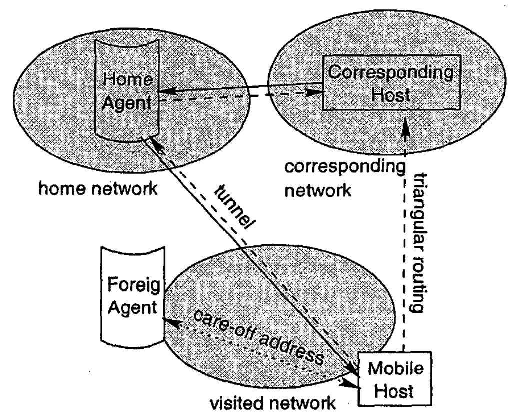

Figure 1: Data flow in Mobile IP network

For this solution many problems have been reported. The triangular routing is one of them: the datagrams toward the mobile host are transmitted to the HA, but the flow in the other direction do not have to pass through this node. Hence a triangle in the dataflow is obtained. Also if the HA is in a remote network, thus when the mobile terminal migrates to a remote network, there is a substantial delay in the end-to-end transmission. Also the protocol is not designed to deal with the mobility, while the terminal is engaged in active sessions.

2.1.2 Hierarchical Mobile IP

To shortcut the location update delay, a chain of Home Agents can be implemented. When the mobile terminal moves to a remote network, and from there to a neighbouring networks, an extra HA can be usefull in the new area: the original HA forwards it to the second one, and only the latter needs to be aware of the exact network location of the mobile terminal [9]. With Hierarchical Mobile IP, more levels of mobility can be introduced in the network.

However the tunneling and poor routing becomes even worse with this alternative. The datagrams are transported from one tunnel to a following one, and the path towards the mobile terminal grows with every hierarchy level. The proposed routing results in an increased load in the network, as the shortest path between the end-points is not used any more. The tunneling of all mobile destined frames requires substantial processing power at the HA. Also the down-link has to be tunneled through the HA, as the other side cannot deal with a connection where the other party changes its address during the session. Hence the tunneling through the HA must hide this.

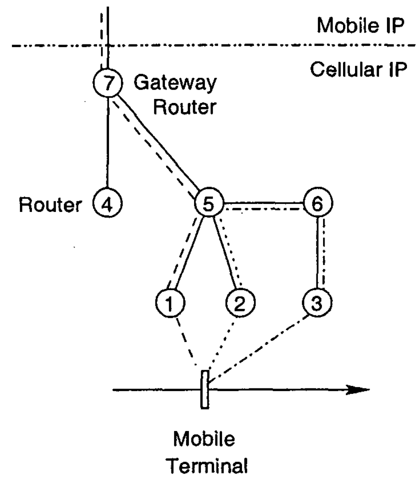

Figure 2: Data flow in Cellular IP network

2.2 Micro Mobility

With Mobile IP the mobile terminal obtains a new address whenever it registers with another network. This is a solution for the macro mobility, e.g. when the computer is powered on whithin a remote network. For the micro-mobility however, when the mobile terminal moves from one access point to another while active, this solution is not fullfilling. This is the hand-over problem. Hence whithin the MobileIP Working Group of the IETF (Internet Engineering Task Force) various alternative proposals have been published. Cellular IP is such an alternative.

2.2.1 Cellular IP

Cellular IP is designed to allow the mobile terminals to roam through the network, keeping their original IP address [10]. The routers in the network keep track of the location of the terminal inspecting the upstream dataflow. These datagrams transmitted by the mobile terminal are always routed towards a gateway in the network, and packets destined for the mobile are forwarded along this route.

Figure 2 is included to explain the dataflow in a Cellular IP network. The routers in the figure are plotted with numbered circles, and only the routers of the Cellular IP network are included. The rest of the world is connected using the Mobile IP protocol. When the mobile terminal moves from an access point connected to router 1 to one connected to router 2 and from there to router 3, the proposed protocol is used to reroute the dataflow. When upstream frames are sent using router 2, router 5 notices this and updates its routing table entry for the mobile host. After a time-out period, the entry in router 1 expires. The gateway router, number 7, does not notice any difference. Subsequently, when router number 3 is used this causes again an update at router number 5 and the hand-over is executed. Hence, only the routing tables in the routers involved in the hand-over are updated. Routers 4 and 7 do not notice anything of the mobility from router 1 to router 3.

To shortcut the latency between the relocation and the rerouting of the data traffic a soft hand-over is implemented: the mobile terminal is allowed to send its upstream data along the new path, while still receiving the down-link data along the original path. When the new path is recorded up to the gateway, no datagrams will be sent along the old path any more and this can therefore be released. Also, if the application results in asymmetric traffic, with almost no upstream data, the mobile terminal must send route update packets to allow the network to register its actual location.

Not accounting for the gateway, this solution is not scalable. The routers in the network must keep track of all the individual hosts in their area. Thus, instead of routing by the network address, each individual host requires a dedicated entry in the routing tables. Also source routing is required for the uplink. Current routers are designed to deal with plain routing only and all options, like source routing, result in poor performances. Finally all upstream datagrams must be inspected at all routers to record the actual location of the mobile terminal. Thus the route update procedures whithin the router are called for all these packets, with a severe performance degradation as a result.

The idea of Cellular IP has been implemented as a link layer protocol, e.g. as a bridging protocol between IEEE 802.11 Wireless Ethernet access points. These protocols are designed to build routes to individual hosts based on their MAC address and have showed to work excellent whithin one subnet. However scaling this solution for larger internets puts severe contraints on the processing power of the routers.

2.2.2 Other proposals

Hence the micro-mobility in IP is still an issue. As explained, whithin an IP subnet it can be solved using bridging protocols. Hand-overs between subnets are however also required. The implementation of Mobile IP for the micro-mobility is not straighforward. The amount of proposals by the IETF for this problem is an indicator of the complexity of the problem.

Many proposals are based on MobileIP or derivates like Hierarchical MobileIP. In order to increase the hand-over speed, a soft forward handover is proposed, where the mobile terminal prepares the new route before the original is lost [11, 12, 13]. In order to do so, it registers at the new access point while communicating through the original one. Hence the new tunnel can be prepared and when established, the old tunnel can be replaced for the new one. The speed of the handover can be further increased using a local anchor in the route, e.g. with Hierarchical MobileIP.

Other proposals, like HAWAI, are only valid whithin one subnet [14]. However, as already discussed, this problem is solved with the wireless bridging protocols. Anyway, the intention to use IP as the network layer for the Third Generation systems is obvious, with proposals to use the air interface specification to provide the initial link layer connections when registering with a new access point [15].

3 Mobility Control beyond Third Generation

The fourth generation systems, called "Systems beyond the Third Generation", focus the scenario where multimedia services are provided by a mobile Internet. To provide these services, the Internet is currently evolving as well. The evolution introduces the conept of flows, or soft-state connections, in the connectionless IP network. Hence when a multimedia session starts, the network recognises this and prepares an and-to-end soft-state connection. This is the so-called layer 3 switching, with MPLS as one of the current proposals to solve this problem [16].

There is a proposal combining MobileIP and MPLS, recognising the scalability issues with MobileIP [17]. In this proposal an MPLS soft-state connection is used to replace the tunnel between the HA and the mobile terminal. However there are still many unsolved issues about MPLS, like the interoperability between different networks, and the poor routing remain. Also the hand-over problem has not been solved: if during a session the mobile terminal moves from one subnet to the next one, it must replace its care-off address. The corresponding host cannot deal with changing source addresses. To solve this problem, tunneling can be re-introduced ... Also MPLS offers soft-state connections. As these may time-out in the network, the destination address of the packets is used to recover the transmission. Therefore the care-off address must be used, reïntroducing the necessity of tunneling at the HA once again.

One very important issue with MPLS is the billing. Unlike fixed backbone networks, where bandwidth is unexpensive, the resources in a wireless environment are very scarce, and therefore expensive. The auctioning of the UMTS licenses in Europe stresses this [1, 5]. Both the billing and the mobility management are solved rather easily in connection-oriented networks: the clear service boundaries (start, stop and type of service) from the SETUP and the RELEASE messaging facilitate the billing, while the confinement of all communication whithin the connections only require to reroute these connections. Also the location management can be done with a HLR, as this information is only required at connection set-up.

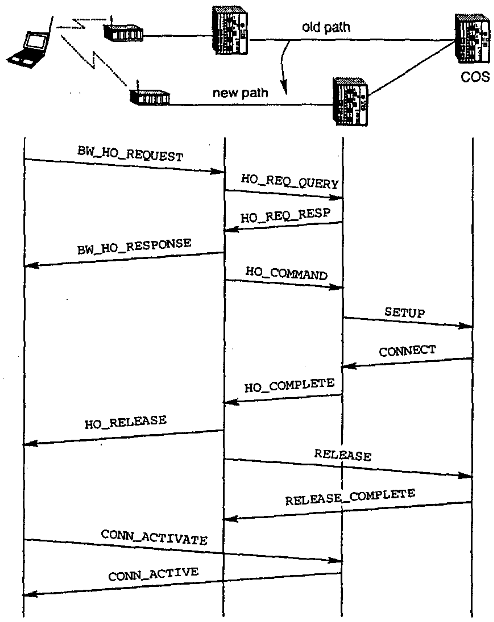

Figure 3: Hand-over signalling in Wireless Mobile ATM

To demonstrate the simplicity of the hand-over in a Wireless Mobile ATM network, the proposal of the ATM Forum can be used [18]. The hand-over is a mobile initiated, network controlled hard backward approach, but a mobile initiated, network controlled hard forward alternative is also defined as a rescue hand-over (when the connection is lost before the new path is prepared). In figure 3 the signalling for the backward hand-over is plotted. When the mobile terminal senses that a hand-over has to be prepared, it informs the original edge switch. This network node searches for a candidate edge switch and contacts this node. The candidate edge switch checks if the required resources are available and replies the result. If it can accept the connection, it prepares the new path with a SETUP message up to the branching point in the network. Once the new path prepared, the new connection can be establised.

Many alternatives have been proposed lately in litterature [19]. Whithout discussing all of them in detail, the conclusion can be drawn that a hand-over can be performed without service degradation noticable by the user. Hence a transparent mobile service can be offered, even for interactive multimedia and video communications, services with the most stringent delay and delay jitter requirements. Also the hand-over procedures can offer lossless mobile services, important for data services [20].

4 Conclusion

Mobile services are gaining importance. Many market forecasts predict that mobile network access will outnumber fixed access, even for data services. Thus the mobile services can have influences on future architectural choices.

To provide these mobile services a connection oriented network can dramatically facilitate this task. This is demonstrated with an overview of the different solutions and proposals for mobility control protocols in broadband multiservices networks, like the target of the Systems Beyond the Third Generation.

As a starting point, it is noticed that the Second Generation systems, both for voice as for data services have opted for a connection-oriented solution. Various proposals for the Third Generation Network are discussed in this paper, but none of them provides a scalable mobility control protocol. As an alternative the simplicity and performance of the connection control procedures in a Wireless Mobile ATM network are detailed.

The authors are aware that the data services are not all served with a connection-oriented solution, as many sessions have lifetimes that are substantially shorter than the time required to prepare a connection. Also it is not stated that it is impossible to provide mobile services in a connectionless network. This paper only adds a new and original contribution to the discussion, whithout the ambition of solving the problem: the solution is probably between both extremes. This paper has not touched the issues with the interactions between TCP flow control and wireless links.

Acknowledgements

This work was funded in part by the generic basic research programme on network technologies of ITA-II (Information Technology Action, Part II) of the Flemish government.

ATM Forum documents, tutorials and specifications are provided for this research by the ATM Forum Student Affiliate Program.

The simulations where performed with OPNET licenses in the framework of the OPNET University Consortium.

References

[1] F. Leite, "IMT-2000: The global standard for third generation wireless communications," in *IEEE International Conference on Third Generation Wireless Communications*, pp. 1–6, IEEE - Delson, San Fransisco, June 2000. Opening Address.

[2] ETSI, "Universal mobile telecommunication system (UMTS)," Tech. Rep. ETSI EG 201 721 V1.1.2, ETSI Guide, 2000.

[3] M. Mouly and M.-B. Pautet, *The GSM System for Mobile Communications.* Mouly/Pautet, Palaiseau, 1992. ISBN 2-9507190-0-7.

[4] C. Bettstetter, H.-J. Vögel, and J. Eberspächer, "GSM phase 2+ general packet radio service GPRS: Architecture, protocols, and air interface," *IEEE Communication Surveys*, vol. 2, pp. 1–13, Third Quarter 1999.

[5] W. Lee, "A total solution for the wireless IP core network," in *IEEE International Conference on Third Generation Wireless Communications*, pp. 9–16, IEEE - Delson, San Fransisco, June 2000. Keynote Address.

[6] J. W. Mark, X. Shen, Y. Zeng, and J. Pan, "TCP/IP in wireless/internet interworking," in *IEEE International Conference on Third Generation Wireless Communications*, pp. 438–445, IEEE - Delson, San Fransisco, June 2000.

[7] C. Perkins, *RFC-2002 IP Mobility Support.* IETF Network Working Group, October 1996.

[8] C. Perkins, "IP mobility support for IPv4, revised." draft-ietf-mobileip-rfc2002-bis-02.txt, July 2000.

[9] E. Gustafsson, A. Jonsson, and C. E. Perkins, "Mobile IP regional registration." draft-ietf-mobileip-reg-tunnel-03.txt, July 2000.

[10] A. Campbell, J. Gomez, C.-Y. Wan, A. Kim, Z. Turanyi, and A. Valko, "Cellular IP." draft-ietf-mobileip-cellularip-00.txt, July 2000.

[11] G. Dommety and T. Ye, "Local and indirect registration for anchoring handoffs." draft-dommety-mobileip-anchor-handoff-01.txt, July 2000.

[12] K. El Malki and H. Soliman, "Fast handoffs in mobile IPv4." draft-elmalki-mobileip-fast-handoffs-02.txt, July 2000.

[13] J. Z. Wang and A. Tang, "Universal mobile IP (UMIP) location management." draft-wang-mobileip-umip-00.txt, March 2000.

[14] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and L. Salgarelli, "Ip micro-mobility support using HAWAII." draft-ietf-mobileip-hawaii-01.txt, July 2000.

[15] Y. Xu, R. Bhalla, E. Campbell, K. Freter, E. McGrath Hadwen, G. Dommety, K. Joshi, P. Yegani, T. Matsumera, A. Tashima, L. D. Hyun, N. Itoh, K. Ohki, B.-K. Lim, P. J. McCann, T. Towle, J. Jayapalan, P. W. Wenzel, C. B. Becker, J. Jiang, S. Shikano, W. Kim, Y. Chang, B. Semper, J. M. Koo, M. A. Lipford, F. Leroudier, and J. Gately, "Mobile IP based micro mobility management protocol in the third generation wireless network." draft-ietf-mobileip-3gwireless-ext-04.txt, June 2000.

[16] E. C. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol label switching architecture." draft-ietf-mpls-arch-07.txt, July 2000.

[17] Z. Ren, C.-K. Tham, C.-C. Foo, and C.-C. Ko, "Integration of mobile IP and MPLS." draft-zhong-mobile-ip-mpls-01.txt, July 2000.

[18] Wireless ATM Working Group, *Draft Wireless ATM Capability Set 1 Specification.* ATM Forum Technical Committee, January 2000. BTD-WATM-01.13.

[19] M. Teughels, I. De Coster, E. Van Lil, and A. Van de Capelle, "Leaf Initiated Join handover evaluation," in *Wireless Networks*, Baltzer Science, September 2000. Selected for publication.

[20] S. Maltha and R. Overduin, "MEDIAN service requirement study," in *Proceedings of the 2nd Mobile Communications Summit, Granada*, pp. 27–29, November 1996.