Wireless World Research Forum
(WWRF)

france telecom
R&D

Network Handled Seamless IP Mobility for Vertical Handovers Between Cellular Networks and WLAN Hot Spots

Eric Njedjou[1], Philippe Bertin[2], Paul Reynolds[3]

[1,2] France Telecom R&D, 4 Rue du Clos Courtel, 35512 Cesson Sévigné
[1]email: eric.njedjou@francetelecom.com
[2]email: philippe.bertin@francetelecom.com
[3] Orange SA, Bradley Stoke, Bristol BS32 4QJ
email: paul.reynolds@orange.co.uk

Subject Area: WG3 – Cooperative & AdHoc Networks

Objectives of the required research

In the future of Wireless IP Communications, mobile users will be provided with multiple radio interfaces equipped laptops and hand held devices as PDAs. Smart devices that possess a multi-radio technologies capable PCMCIA or flash card will also become pervasive.

These capabilities will give users the ability to access services on the Internet regardless of the radio coverage type (GPRS, UMTS, CDMA1xRTT, 802.11…) they are present in. In such areas as airports, Conference halls and hotels, several radio access networks of different technologies would even be simultaneously present. The user would then have the ability to watch a video news flash on WLAN then move to GPRS/UMTS for web browsing. It could also become possible for the same customer to start file transfer on WLAN and complete the operation on GPRS/UMTS either because he moved away from the WLAN coverage or because he suddenly lost its WLAN connectivity

In order to anticipate mobile users needs to come, it is becoming essential to design optimal ways by which the mobile device can seamlessly change its access link and therefore its point of attachment to the Internet, while the user is moving between radio access points of the various available technologies. It is even more challenging to enable such an inter-technology handover in a way that is convenient to both the customer's profile requirements and the network operator resources utilization.

The aim of this document is to propose a generic mechanism to efficiently handle terminals IP handovers in Beyond 3G architectures where by moving devices will connect to the Internet by a variety of Access Networks each of different technologies.

State of the art in Handoffs techniques using IP mobility

Moving between heterogeneous radio access networks generally implies changing the IP subnet the device's IP layer is attached to. A user would for instance be handing over between a 3G cellular network and a WLAN Hot Spot operated by the same network provider but with access networks not sharing the same IP subnet range. Such a handoff will obviously requires that the mobile node changes its current IP subnet of attachment.
Mobile IP protocols ([MIPv4] and [MIPv6] standardized within the IETF make it possible to change the mobile node's current subnet (i.e access router) while maintaining on-going higher level connections, and especially TCP/UDP level connections.

Before the Mobile IP handover occurs (layer 3 handover), the layer 2 handover or attachment to a new link has to be completed first. In the case where the connectivity on the old link (link of attachment before the handover) is lost by the device, all the packets sent or destined to it will be lost until the Mobile IP registration on the new link has been completed. Therefore, to reduce the packet loss, the layer 3 handover needs to be prepared and/or anticipated.

This is the objective of low latency handoff protocols under specification within the IETF Mobile IP Working Groups: [FMIPv4] and [FMIPv6]. These fast handoff protocols are designed to help minimize the loss of IP packet between the time a mobile node looses its current link connectivity and the time IP packets really start to be forwarded to/from its new link. In these specifications, link-layer hints are assumed to be the triggers used in anticipating the IP subnet change following a layer 2 handover. These are information available from one or multiple link-layer interfaces of a device, that can inform the network layer that some events (link up, link down, link about to go down) might have happened and that may force or prompt the device to look for a new link, or that may even decide it to move to that new link even if the current one is still available.

The previous protocols well apply when the change of IP subnet is not the result of a change of link layer technology and therefore of Access Network type, topology and configuration. Effectively, the handoff process in the above specifications ([FMIPv4] and [FMIPv6]) is partially executed between the previous and new access routers, that both run the handoff protocol. This implies that they are not located two many hops away from each other (because the handoff should be done expeditiously) and more importantly, that they are equivalent nodes, which is not automatically always the case; an example is between a 802.11 access router and a GGSN (GPRS router). New mechanisms are therefore needed that will provide an abstraction of the link technology and also provide new criteria that will be useful for performing the seamless possible handover between any two heterogeneous networks, regardless of their type.

Proposed approach for inter-technology IP handoff management

Outline

The first section will detail simple terminal handover using Mobile IP between two subnets served by attachment points of two technologies (illustrated by an example featuring 802.11b and GPRS), with the intent of showing that the handover is not optimal using Mobile IP alone. In a second section we will introduce the link layer hints parameters, abstracted from different technologies, that can conveniently assist the network in optimising the mobile device handover between the previous subnets. The third section will present other handoff decision factors, likely to reside at the network side, that will reflect the service provider handover policy and will be of precious complement in performing an effective target selection for handover. The fourth section will discuss architecture choices for the location of a seamless mobility management function or entity within the network. A fifth section will illustrate a possible service scenario of inter-technology handover using the presented approach.

Simple Mobile IP enabled vertical handover between GPRS and 802.11b

We assume that the mobile device runs a Mobile IP protocol ([MIPv4] or [MIPv6] and uses two different interfaces to access the GPRS and 802.1b networks.

<u>GPRS to 802.11b handover</u>

Let's say the mobile node is attached to the GPRS network and the IP address obtained via the PDP context is currently the care-of-address registered to the Home Agent. When the device moves towards a Hot Spot area, the establishment of a new IP address generally takes place by auto-configuration (stateless or stateful) following the attachment to the new link-layer and sending of IPv4 DHCP Discover Requests, or reception of ICMPv6 Router Advertisements. This new IP address is a second care-of-address for the Mobile IP stack, and is automatically registered to the Home Agent through a Binding Update message. The handover to the 802.11b access occurs as soon as the Binding Ack message is received. Effectively, at its reception, the 802.11b care-of-address becomes active while the GPRS one turns inactive.
Therefore, the switch to the newly attached access network is solely based on the acquisition of the new care-of address on that access network. Yet, the Wifi network might be overloaded (which might not be convenient to both the user entering the Hot Spot and the network operator because the first will not be served efficiently and the second would not be able to balance its access resources). Further, the Wifi access may even not correspond to the user capability in term of access cost.

<u>802.11b to GPRS handover</u>

The device is supposed to be currently attached to the WLAN access network with the IP address obtained as indicated above being the care-of-address registered to the Home Agent. The device is moving outside of the Hot Spot area. The 802.11b connectivity will progressively get poorer until loss is encountered. As a result, a handover to GPRS (active care-of-address becoming that of GPRS) will take place. Effectively, once a previously active

care-of-address happen to disappear, Mobile IP sends a registration to the Home Agent with the next care-of-address on the node's list. For this to be possible, the mobile node needs to keep its GPRS IP attachment.
In this Mobile IP handover, IP packets exchanged between the time the 802.11b link connectivity has gone and the time the registration update is really done with the GPRS care-of-address will be lost.

Yet, if some link-layer hints were extracted from the WLAN interface, it would have been possible to anticipate the loss of 802.11b link connectivity by monitoring the associated radio signal quality and strength. In that way, the handover to the GPRS would have been executed as soon as the monitored parameters reach a certain low threshold, positioned so that the 802.11b link is still established. This would have resulted in reducing the packet loss. Furthermore, it's possible using such hints, to reduce power consumption on the node by activating the GPRS interface and PDP context only when such a threshold is reached.

Other shortcomings

Further, Mobile IP is not meant to achieve "static handover". We designate by "static handover" a handover that is not directed by the movement of the mobile device. For instance, a user whose network profile specifies "browsing the web with any available access" and who is currently benefiting from the Hot Spot coverage could be handed (by the network) over to GPRS in case a user with a profile specifying "browsing the web at the best available speed" arrives in the Hot Spot.

Regarding the drawbacks early stated, some new factors, we detail in the next sections, are to be taken into consideration when achieving inter-technology handovers using Mobile IP.

Link layer hints

With the development of new protocols managing efficiently host mobility at the IP layer, a need appears for providing tight interactions with link layers, either in terminals and in networks (e.g. between an access point and an access router). The goal for such inter-layers communications is to optimize network layer operations (handover anticipation, target attachment point selection, link status detection…) based on information available at lower layers (radio signal strength, air interface load…). This is particularly needed for Mobile Nodes integrating several link layer interfaces with heterogeneous technologies in order to control the selection of the right radio module.
In order to envisage such inter-layers communications, link layers hints are expected to provide notifications from the link to the network layer when changes occur in link environment. For example, such events can be the detection of a new point of attachment, the start/end of a handover procedure for a Mobile Node, or the new attachment of a Mobile Node for an Access Point. Link hints will also provide the necessary framework to manage communications over several interfaces in a given Mobile Node.
In order to standardise link hints, it is necessary to define information elements that they shall transport. These information elements shall have the necessary abstraction level to be mapped over several technologies. A first set and classification of parameters for link hints are defined in [PAR]. The table below gives an example of how those parameters could be mapped on several technologies at a Mobile Node.

	IEEE 802.11	Bluetooth	GPRS
MN static parameters			
MN interface Type List	802.11g,a	Bluetooth	GPRS
MN interface type options	802.11i	BNEP – RFCOM	-
MN hardware ID	H.MAC addr	H.MAC addr	IMEI
MN configuration parameters			
MN current interface type	IEEE 802.11a	Bluetooth	GPRS
MN default network ID	SSID	X	PLMN
MN maximum Tx power	50mW	1mW	2W
MN data rate	6, 12, 24Mbps	1Mbps	30kbps
MN security level	EAP TTLS	Authentication	SIM auth.
MN frag. Threshold	max MTU	max MTU	max MTU
MN link environment parameters / MN interface status			
MN current network ID	SSID	X	PLMN
MN current attach.ID	BSSID	H.M.addr of AP	cell-id
MN power mode	PSP	park/hold/...	ready
MN measured bandwidth	x Mbps	x Kbps	x Kbps
MN Bit error rate	probability	probability	probability
MN packet error rate	probability	probability	probability
MN current data rate	x Mbps	x Kbps	x Kbps
MN curr. transmit power	x mW	x mW	x mW
MN curr. radio link qual.	x dBm	x dBm	x dBm
MN interface status	ON || OFF	ON || OFF	ON || OFF
MN L2 handover status	YES || NO	YES || NO	YES||NO
MN interface load	x packets	x packets	x packets
MN noise level	x dBm	x dBm	xdBm
MN link environment parameters / MN available attachment status			
MN other network ID	SSIDs in range	X	PLMN(s)
MN available attach. IDs	BSSIDsin range	H.M.addr of AP	cell-id
MN net. adv. Frequency	beacon period	X	X
MN net. config. data rate	6, 12 Mbps	1Mbps	30 Kbps
MN avail. radio link qual	x',x''... dBm	x',x''..dBm	x',x''..dBm

Fig. Link hints parameters

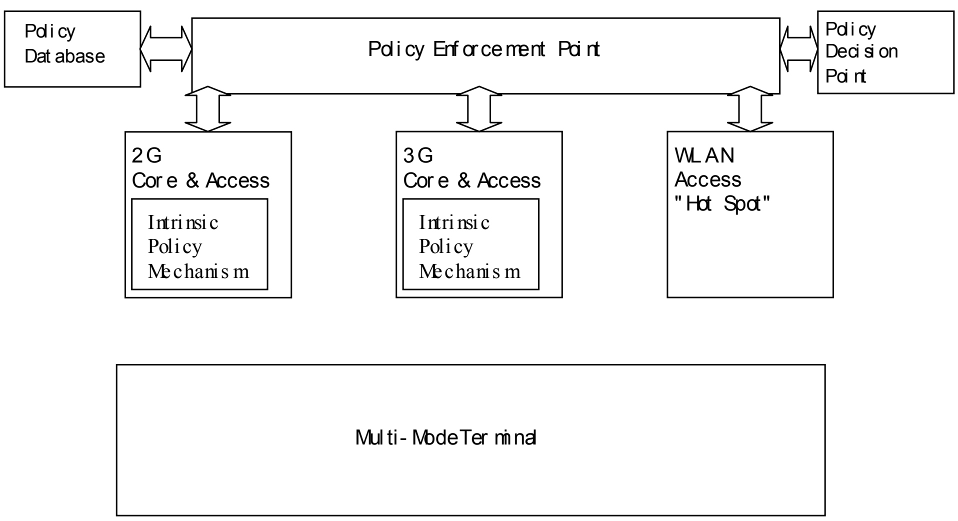
The IETF [DNA] Working Group has initiated an effort to cataloguing link hints existing in a couple of well known technologies and will define an abstraction which should be technology independent. However, [DNA] is limiting the use of these link hints to optimising the detection of network attachment.

Network Operator handover policy

The aim of a competent network provider is to attempt to correlate business decisions to actuality within the network. As networks evolve and more access technologies develop, the criteria for selecting the optimum technologies to meet service and business policy requirements will become more complex. Accesses may be added to both the network and the mobile terminals independently of each other so that they have no 'knowledge' of each other. A subscriber does not want to be faced with a decision each time they establish a session as to which is the best technology to use for that session, in that area, at that time of day. The task of initially selecting the most suitable access technology and for maintaining the optimum connection and essentially managing the terminal mobility, is one that should not burden the subscriber, but should be under the control of a separate entity that is capable of undertaking these tasks. This entity is the Policy Server

A Policy Server is envisaged as being implemented as a policy decision point and policy enforcement point, which could reside either within the mobile terminal or the network. Whilst a mobile terminal based policy server has an awareness of the immediate environment in which it is operating, and of the particular application requirements, it does not possess the knowledge regarding network capabilities and capacity that may or may not be available. A network based policy server on the other hand, knows precisely what network resources are available from particular accesses at particular times and is in a position both to provide resources to meet the subscribers needs and to balance the network load to ensure capacity for all subscribers.

Such a stand-alone policy server is by definition, a hierarchical element information, communicating not only with the policies within the individual accesses (see figure 1). The advantages of such an architecture are its independence of any particular technology, its ability to address the network operator’s requirements and its capabilities to address the requirements of the subscriber, as defined in their profile. The later are located in a database, owned by the network operator, associated to a policy server. These profiles list the parameters that could be used to decide an inter-technology handover. They are updated either by the network administrator when subscription changes occur or dynamically by the system when the subscriber environment change (lost of signal, new access network available in proximity. Where necessary, the policy server can also communicate with the mobile terminal itself, to obtain either environmental or application level information such as quality of service requirements.



Example of a Hierarchical Policy System

Architectures for the location of the handover management function

A seamless handover should by essence be an expeditious process. The first reason for this is the necessity to achieve absolute reduction of packets loss. A second reason is that factors used to trigger the handover process can be derived from events happening unexpectedly ("link down" event is an example).

As such, the handover decision ought also to be taken very rapidly. As a result, the inter-technology handover manager should be located so as to be able to get all the decision factors (link hints, access networks load information…) within an appropriate lap of time. It has to be noted that after gathering the necessary information, the handover manager has to run the selection algorithm and then send the decision back to the terminal.

It then appear clear that locating the function in a node residing network clouds away from the mobile terminal won't happen to be very effective as this will lead to enormous delays and de-synchronization with the network (a node receiving a decision that is inconsistent with the current available accesses). Therefore, In place of always having an hierarchical technology independent element to enforce the operator policy, it could be alternatively suggested to have the handover management performed in one access network and preferably the one that can allow the mobile node not to be too many hops away from the IP node dedicated for that management function. In this case, pretty fast handover procedures are able take place.

Nevertheless, some situations might not justify any urgent need to perform a handover. For load balancing purposes for instance, the network operator might decide to move some users from Wifi to GPRS based for example on the amount of time they've been using the 802.11b access (fair access). In this example, the handover manager need not be in the very vicinity of the mobile terminal. It can be located in a cloud kept separate from the access networks as shown in the previous section. Such a topology may be even envisaged for a better monitoring and assessment of the load as reported from every access network.

Possible seamless mobility service scenario

We have seen that seamless mobility will make use of a combination of intrinsic intra-technology management, inter-technology mobility management and a controlling policy server whose responsibility is to keep an ongoing session on one terminal while changing from access networks. The maintenance of seamlessness takes three forms:

(i) *Non-Contiguous Access Network*: Here there is an actual discontinuity between the coverage of the various access technologies (e.g. WLAN in the office and Wireless xDSL in the home); as such there is likely to be an extended period of time to complete the handover (this can be considered as a type of roaming). However, from a service point of view the session remains and once the terminal has "rebinded" to the new access technology the session continues seamlessly.

(ii) *Contiguous Access Network*: Here there is continuous coverage but supplied by different access technologies (e.g. WLAN in the office and 3G in the open); as such the handover can be optimised to support user requirements, network conditions or application requirements.

(iii) *Overlapping Access Network*: Here the coverage of the access networks overlap each other thus enabling the user, network or application to select the best access mode. The handover in this case can take the form of a user being asked what network he prefers, a network to select the best access based upon network parameters (e.g. congestion or return on capital investment) or an application trigger.

Possible service scenarios representing a real live situation where network controlled seamless mobility could be beneficial for both the user and the operator is as follows: seamlessness may be handled in three different ways, although they will appear neutral to the user.

- **Scenario i** - Here a mobile user has a multimedia terminal that includes the support of VoIP. This user spends a lot of time in places with WLAN service and would like to utilise WLAN for his multimedia calls whenever possible. However, he is now on-the-go and may need to leave the area with WLAN in the middle of a call. Nevertheless, he still would like to maintain his multimedia and VoIP sessions on his terminal without noticeable interruption when he enters another WLAN or cellular coverage.

- **Scenario ii** - Here the user has a WLAN card in his mobile terminal, and can switch between cellular and WLAN as necessary without interrupting the session.

- **Scenario iii** - When he comes back to his office, let us say his terminal is connected to the application through GPRS providing continuity from outside, but in those premises the terminal is exposed to a second access network: WLAN. The terminal will switch networks whilst it is not moving (i.e. network optimisation).

Time frame to get the expected result

We have already submitted two IETF Drafts; the first ([IAH]) tries to motivate the need for handling inter-technology seamless Mobile IP handovers with assistance from the network. The second (PAR] proposes an abstraction of link hints available from different radio technologies in generic parameters, to be used in achieving access network selection. We are currently contributing to a Draft within the [DNA] Working group that is cataloguing link hints from various radio technologies (802.11b, GPRS, CDMA 1xRTT) and defining an abstraction that can help in optimising the detection of network attachment and further network configuration of mobile nodes. We are also looking forward to;

- Contributing within the MIPSHOP Working Group, with the intent of specifying an inter-technology fast handoff process for Mobile IPv6
- Contributing to the 3GPP in its effort to specify the use of mobile IP in achieving seamless mobility between WLAN and 3G access networks

List of References:

[MIPv4] "IP Mobility Support", C. Perkins (Editor), RFC 2002, October 1996.
[MIPv6] "Mobility Support in IPv6", D. Johnson, C. Perkins, and Jari Arkko, draft-ietf-mobileip-ipv6-21.txt, work in progress, February 2003
[FMIPv4] "Low latency handoffs in Mobile IPv4", K. Malki et al., draft-ietf-mobileip-lowlatency-handoffs-v4-00.txt, work in progress, July 2003
[FMIPv6] "Fast Handoffs for Mobile IPv6", MIPv6 handoff Design Team, draft-ietf-mobileip-fast-mipv6-06.txt, work in progress, March 2003

[DNA] "Detecting Network Attachment", Greg Daley, Working Group Charter text, August 2003, work in progress
[CARD] "Candidate Access Router Discovery", Marco Liebsch, Ajoy Singh et al., draft-ietf-seamoby-card-protocol-03.txt, work in progress, August 2003
[IAH] "Motivations for achieving Network Controlled Handoffs using IP mobility between heterogeneous Access networks", E. Njedjou et al, draft-njedjou-inter-an-handoffs-00.txt, work in progress, June 2003
[PAR] "Link Hints Parameters", P. Bertin et al, draft-bertin-link-hints-params-00.txt, work in progress, August 2003