IP Flow Mobility: Smart Traffic Offload for Future Wireless Networks

Antonio de la Oliva*, Carlos J. Bernardos*, Maria Calderon*, Telemaco Melia† and Juan Carlos Zuniga‡
*Universidad Carlos III de Madrid, E-mail: {aoliva,cjbc,maria}@it.uc3m.es
†Alcatel-Lucent Bell Labs, E-mail: telemaco.melia@alcatel-lucent.com
‡InterDigital LLC., E-mail: JuanCarlos.Zuniga@InterDigital.com

***Abstract*— The recent proliferation of smartphone-based mobile Internet services has created an extraordinary growth in data traffic over cellular networks. This growth has fostered interest in exploring alternatives to alleviate data congestion while delivering a positive user experience. It is known that a very small number of users and applications cause a big percentage of the traffic load. Hence, adopting smarter traffic management mechanisms is one of the considered alternatives. These mechanisms allow Telecom operators to move selected IP data traffic, for instance between the cellular infrastructure and the WLAN infrastructure, which is considered a key feature in the latest 3GPP and IETF specifications. This paper presents and compares two possible approaches to IP flow mobility offloading that are currently being considered by the IETF. The first one is based on extending existing client-based IP mobility solutions to allow flow mobility where the user terminal is fully involved in the mobility process, and the second one is based on extending current network-based IP mobility solutions where the user terminal is not aware of the mobility.**

I. Introduction and Motivation

In the past few years we have been witnessing an extraordinary data explosion over cellular networks. Telecom operators have been carefully monitoring the disconnection between the Average Revenue Per User (ARPU) and the associated Cash Costs Per User (CCPU) and, despite the remarkable volume increase of broadband data over mobile networks, the mobile data revenue is falling fast.

There are a number of reasons for such disconnection between data explosion and revenue growth, including among others, terminal subsidies, marketing and sales costs, new services and new content creation, staled data plans and tariffs, network capacity or network coverage and management. In the context of network operational expenditure cost, efficient technology solutions seem to be the most promising approaches. Smaller installation footprints, reduced power consumption and transmission costs, efficient use of multi-radio bandwidth, simplified network management, reliable and cost effective coverage are just examples of the plethora of existing solutions.

Presently, the typical scenario is a user equipped with a dual mode mobile phone (e.g., integrating 3G/4G and WiFi radio devices) attaching to the available networks either sequentially or simultaneously. The latter case is commonly referred as multi-homing case, that is, the user can receive data over different networks (WiFi or 3G/4G) simultaneously.

Part of the research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-ICT-2009-5) under grant agreement n. 258053 (MEDIEVAL project)

In this work, starting from the above considerations, we focus on an emerging technology, referred hereafter as IP flow mobility. This technology allows a Telecom operator to seamlessly and selectively switch over a single IP flow (e.g., user application) to a different radio access, while keeping all other ongoing connections for this and the rest of the users on both radio accesses untouched. The technology is currently being standardized in the IETF and it has been adopted by 3GPP as technique for seamless WiFi offload (e.g., moving a video streaming flow from a 3GPP to a WiFi access network) and consequently, reducing the traffic load of the 3GPP network. IP flow mobility technology has the following key advantages: *i)* it allows the user to enjoy high bandwidth connections in proximity of WLAN hotspots while being always reachable from the Internet, *ii)* it allows the operator to manage the bandwidth in the presence of greedy user connections, and *iii)* it allows the operator to provide different levels of service by applying different policies for different users, tariffs and specific traffic types evolving from a simple pipe provider to a high leverage network provider. The access and core networks are therefore capable of classifying data traffic traversing their nodes and, in agreement with the mobile devices, can apply policies to deliver the best Quality of Experience (QoE) possible.

This article analyzes and compares the two possible approaches to IP flow mobility, namely client-based and network-based IP flow mobility. The former, relies on a host centric solution introducing a mobility client in the host and a mobility agent in the core network (Section II). The latter, relocates the mobility client functionality from the host to the network thus making the mobile device agnostic to any IP mobility signalling (Section III). The article, summarizes the key functional boxes and associated protocol operations, discusses the pro and cons of each solution. The paper also generalizes the adoption of network-based solutions in the context of 3GPP and the use of alternatives network-based mobility protocols (e.g., GPRS Tunnelling Protocol).

II. Flow Mobility in Client Based IP Mobility

Client-based IP mobility solutions require the user terminal to be involved in the management of the mobility, by running a specialized stack that is able to detect, signal and react upon

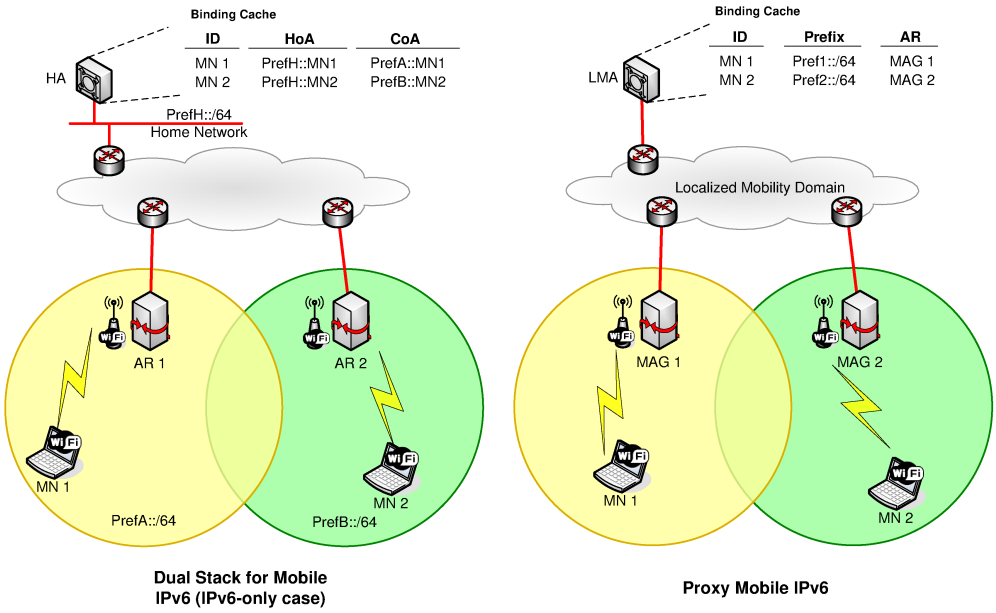

Fig. 1. Dual Stack for MIPv6 and PMIPv6 overview

changes of point of attachment. Dual Stack for Mobile IPv6 [1] is standardized by the IETF to provide basic client IP mobility support.

A. Dual Stack for Mobile IPv6

The Mobile IPv6 Support for Dual Stack Hosts and Routers specification [1] – also known as DSMIPv6 – is based on Mobile IPv6 (MIPv6) [2], extending its basic functionality to also support dual stack IPv4/IPv6 scenarios. Mobile IPv6 (MIPv6) [2] enables global reachability and session continuity by introducing the Home Agent (HA), an entity located at the Home Network of the Mobile Node (MN) which anchors the permanent IP address used by the MN, called Home Address (HoA). The HA (see Fig. 1) is in charge of defending the MN's HoA when the MN is not at home, and redirecting received traffic to the MN's current location. When away from its home network, the MN acquires a temporal IP address from the visited network – called Care-of Address (CoA) – and informs the HA about its current location. An IP bi-directional tunnel between the MN and the HA is then used to redirect traffic from and to the MN. There is also optional support to avoid this suboptimal routing and enable the MN to directly exchange traffic with its communication peers – called Correspondent Nodes (CNs) – without traversing the HA. This additional support is called Route Optimization (RO), and allows the MN to also inform a CN about its current location.

DSMIPv6 extensions add to basic Mobile IPv6 the capabilities required to support the registration of IPv4 addresses and the transport of both IPv4 and IPv6 packets over the tunnel with the HA. These extensions also enable the mobile node to roam between IPv4 and IPv6 access networks.

B. Flow mobility extensions for Mobile IPv6

The basic Mobile IPv6 specification and the extensions defined to enable IPv4 operation provide a very limited multi-homing support, as each permanent address (home address) can only be associated to a single temporal address (care-of address). Therefore, the only possible scenario in which a mobile node can use more than one care-of address simultaneously is that in which the node is using different home addresses (one per care-of address). This limits the scope and usability of this basic solution as it prevents different flows to be routed to different care-of addresses, and consequently, does not support a scenario in which a mobile node is reachable – via a single home address – through different physical interfaces.

In order to enable flow mobility in a client mobile IP context, the IETF has standardized the basic components that are required. These components are: *i)* multiple care-of address registration support, *ii)* flow bindings support, and *iii)* traffic selectors definition. We next explain in further detail how each one of these pieces works, pointing out the basic functionality they provide and how each component fits in the overall flow mobility solution.

Basic Mobile IPv6 protocols provide the tools to bind a home address to a single care-of address. Since flow mobility requires the ability of receiving traffic destined to the same home address via different care-of addresses, Mobile IPv6 needed to be extended to support the registration of several care-of addresses with the same home address. This is the purpose of the Multiple Care-of Addresses Registration extensions, standardized in the RFC 5648 [3]. These extensions allow a mobile node to register multiple care-of addresses for a home address and create multiple binding cache entries. In

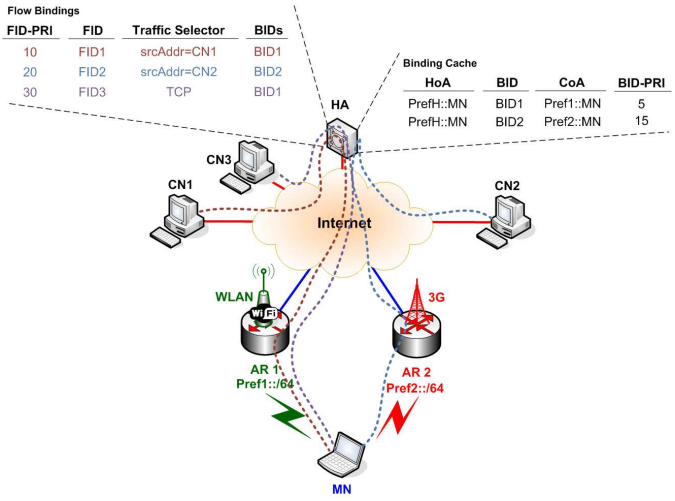

Fig. 2. Overview of the flow mobility extensions for Mobile IPv6

order to do so, the Binding Update (BU) message defined by Mobile IPv6 is extended with a new mobility option used to carry a care-of address and a number to uniquely identify the binding entry, called Binding Identification (BID) number. A mobile node can include a number of these new mobility options in the BU message, triggering the creation of multiple binding cache entries in the home agent, each of them identified by the respective BID. Note that the binding cache and binding update list structures are also extended to support the multiple care-of address registration. Fig. 2 shows with an example how the flow mobility extensions for mobile IPv6 work. A mobile node (MN) – identified by its home address `PrefH::MN` – is simultaneously attached to two different heterogeneous access networks (WLAN and 3G), therefore configuring two care-of addresses (`Pref1::MN` and `Pref2::MN`). Thanks to the use of the multiple care-of addresses registration extension, the MN is able to register its two care-of addresses at the home agent. Note that although we are always referring to the registration at the home agent in this example (and in the explanation of the different extensions), the protocols are also defined for its use in the registration with correspondent nodes.

In addition to the capability of associating a single home address with different care-of addresses, the ability to use and control them simultaneously is required. This is the goal of the second set of extensions, the Flow Bindings in Mobile IPv6 and NEMO Basic Support – standardized in RFC 6088 [4] – which allows mobile nodes to bind one or more IP flows to a specific care-of address. With this extension, a mobile node can instruct the home agent (or the correspondent node) how to route inbound packets (i.e., to which care-of address packets of a specific flow should be sent). Note that the mobile node also needs to have support to be able to route outbound packets via different care-of addresses, being that packet forwarding coherent with the inbound policy signaled by the mobile node. The flow bindings specification basically defines a set of Mobile IPv6 options and sub-options allowing the mobile node to associate a particular IP flow (which is also assigned a Flow Identifier, called FID) with a particular care-of address (identified by its BID). These bindings between IP flows and entries in the binding cache are stored in a different conceptual list, that is looked up in order to determine which entry of the binding cache has to be used to forward a data packet. This list basically includes the FID, a traffic selector that is used to assign packets to flows (i.e., a flow is defined as a group of packets matching a traffic selector), and a FID priority – used to break the tie between overlapping flow bindings.

The last above-mentioned extension required to enable IP flow mobility is the definition of traffic selectors for flow bindings, standardized in RFC 6089 [5]. This extension basically defines binary formats for IP traffic selectors to be used in conjunction with the flow binding extensions, so IP flows can be identified according to different criteria (such as 5-tuple, transport protocol, etc).

If we refer back to the example shown in Fig. 2, the use of the IP flow mobility extensions allows for example to influence which data path is followed by the different traffic that the mobile node is sending/receiving. In this example, any traffic sent by CN1 is forwarded by the home agent to the care-of address that the mobile node has configured from the WLAN access. Traffic sent by CN2 is similarly received by the mobile node via its 3G interface. Any TCP traffic not sent by CN1 or CN2 is received via WLAN (note that here the use of the FID-PRI). Finally, any traffic not matching any of these rules is forwarded by the home agent to the WLAN interface of the mobile node, as indicated by the binding cache entry with the highest order BID priority (BID-PRI).

In addition to these basic protocol components, complementary support might be needed to deploy a complete IP flow mobility solution in an operator's network, such as a framework to transport policies from the operator to the mobile node. The Access Network Discovery and Selection Function (ANDSF) framework defined by the 3GPP or the Policy and Charging Control (PCC) support can be used/extended for that purpose.

III. Flow Mobility in Network Based IP Mobility

Network-based IP mobility solutions locate the mobility management control of the terminal in the network. In this way, the terminal is not required to perform any kind of signalling (e.g., binding updates) to react upon changes of its point of attachment to the network, being these changes transparent for the mobile terminal IP protocol stack. Proxy Mobile IPv6 (PMIPv6) is the protocol standardized by the IETF to provide network-based IP mobility support. Although this protocol provides basic multi-interface functionality, in its current state it is not able to provide full flow mobility granularity, hence extensions to support it are required and are being standardized at the IETF NETEXT WG [6].

A. Proxy Mobile IPv6

Proxy Mobile IPv6 (PMIPv6) [7] is a network-based mobility management protocol. This means that the MNs are provided with mobility support without their involvement in the mobility management and IP signalling, as the required

functionality is relocated from the MN to the network. In particular, movement detection and signaling operations are performed by a new functional entity – called Mobile Access Gateway (MAG) – which usually resides on the Access Router for the MN (see Fig. 1). In a Localized Mobility Domain (LMD), which is the area where the network provides mobility support, there are multiple MAGs. The MAG learns through standard terminal operation, such as router and neighbor discovery or by means of link-layer support, about an MN's movement and coordinates routing state updates without any mobility specific support from the terminal. The IP prefixes (Home Network Prefixes) used by MNs within an LMD are anchored at an entity called Local Mobility Anchor (LMA), which plays the role of local HA of the LMD. Bi-directional tunnels between the LMA and the MAGs are set up, so the MN is enabled to keep the originally assigned IP address despite its location changes within the LMD. Through the intervention of the LMA, packets addressed to the MN are tunnelled to the appropriate MAG within the LMD, making hence the MN oblivious of its own mobility.

As previously explained, the standard PMIPv6 protocol allows basic multi-homing capabilities, that is, the MN is able to attach to the network using multiple interfaces. In the current specification, for each of the attachments the LMA creates a different mobility session and can provide one or several home network prefixes (HNP) to each interface. The basic functionality provided by PMIPv6 enables the LMA to move the complete set of prefixes associated to one interface to another, but it does not support the movement of an arbitrary number of prefixes from one interface to other (i.e., not the complete set) or just a single IP flow identified by any other mechanism different from the prefix used at the MN to route the flow. In order to support full flow mobility granularity, the PMIPv6 protocol must be extended to: *i)* span one mobility session across multiple MN interfaces, *ii)* allow the MN to configure the same home network prefixes on multiple interfaces and *iii)* transfer the policies between the MN and the network to install the required filters in the LMA/MAG for flow routing.

In the following section we analyze how each of these issues is being addressed in the current standardization efforts.

B. Flow mobility extensions for Proxy Mobile IPv6

Although the basic specification of PMIPv6 provides limited multihoming support to multimode devices, it does not include the ability to move selected flows from one access technology to another. This functionality is currently being developed by the IETF NETEXT WG[1] as described in [6]. The rest of this section focuses on the description of the key concepts behind the flow mobility support for PMIPv6.

Flow mobility assumes simultaneous connection to the same PMIPv6 domain through different interfaces. The simultaneous use of different attachments to the network increases the complexity of the solution due to two main reasons:

- In order to support flow mobility, the MN must be able to send and receive traffic to/from any prefix associated to it through any of its interfaces. This functionality can be provided by different mechanisms. Two of the mechanisms that have been studied at the IETF are the Weak Host Model and the Logical Interface. On one hand, the Weak Host Model [8] corresponds to the implementation decision taken while designing the IP stack. In a mobile node implementing the Weak Host Model, the IP stack accepts any locally destined packet regardless of the network interface on which the packet was received. On the other hand, the Logical Interface is a software entity which presents one single interface to the IP stack, and hides the real physical interface implementations (e.g., modems). Hence, the IP stack binds its sessions to this Logical Interface and it is oblivious of the actual physical interfaces receiving or sending packets. One of the principles of PMIPv6 is to achieve a mobility solution in which the IP stack of the mobile node is completely unaware of the mobility. In order to maintain the MN's IP stack unaware of mobility while providing flow mobility support, the IETF has chosen to rely on the concept of Logical Interface (LIF) [9].
- In the general case, through the use of flow mobility, the MN will be able to receive any traffic destined to any of its IPv6 addresses through any of its interfaces. This represents a problem at the MAG level, since in order to support flow mobility, the MAGs must be able to forward any prefix associated to the MN even if this prefix was delegated by a different MAG. This situation is being solved by the IETF through the addition of extra signalling to the standard PMIPv6 so that the MAGs can be configured appropriately.

In the following we explain in detail the solution to both issues presented above.

1) Logical Interface: The Logical Interface is a software entity that hides the real physical interface implementation to the host IP layer. Its use allows the MN to provide a single and permanent interface view to IP and the layers above, that can bind to this interface in order to establish any remote communication. Internally the LIF is able to leverage several functionalities such as inter-technology handover, multihoming or flow mobility, while presenting always the same IP address (or set of IP addresses) to higher layers. Even more, its implementation does not require any modification to the IP stack at the MN, so it remains standard.

This interface is implemented as a logical entity that bonds several physical interfaces (e.g., WiFi and 3G) into a unique interface, which is used by IP and higher layers. The LIF hides to the IP layer the physical interface used to actually send each data, hence a movement of a flow from one interface to another is transparent to the IP and higher layers. Even more, it supports sequential attachment of interfaces as they come up, so the flow mobility features can be started in order to offload some interface or network (e.g., 3G offload) as soon as a new interface becomes active (e.g., a WiFi interface associates with an Access Point), without the higher layers being aware of it. The LIF is sometimes referred to as Virtual Interface.

2) Signalling extensions to PMIPv6: As explained above, signaling extensions to PMIPv6 are required in order to

[1] `http://datatracker.ietf.org/wg/netext/`

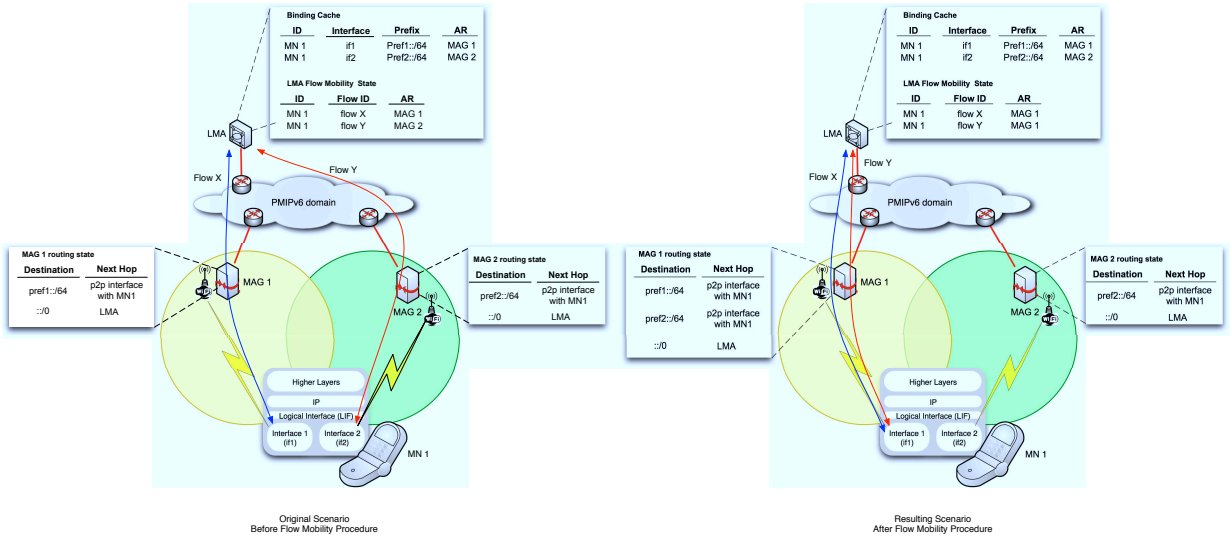

Fig. 3. PMIPv6 flow mobility operation

provide the MAGs with the information regarding the different prefixes used by the MN. This information exchange is needed since, in general, a MAG will not forward traffic from/to a prefix that has not been delegated by it to the MN.

In [6] several cases showing the possible configurations for the combinations of prefixes and interfaces are detailed. The IETF currently focuses on two scenarios: *i)* the movement of a specific flow from one interface to another (e.g., a video-conference where the voice is going through a reliable interface such as 3G and the video through a high bandwidth link such as WiFi, but both flows are addressed to the same prefix), and *ii)* the movement of a complete prefix and all the communications using it, to another interface. The former is often referred to as ”handover with full flow granularity” while the later one is referred to as ”partial handover”.

Both cases face the problem of requiring the target MAG to get knowledge regarding the prefixes through which the MN is receiving traffic. Flow mobility signaling takes place whenever the LMA decides to move a flow from one access to another. At the time of movement, either the prefix is already known at the target MAG or the LMA must advertise it to the MAG which is going to receive traffic addressed to this prefix. In the case the MAG already knows the target prefix, the LMA simply switches the flow to the target MAG, and no extra signaling is required. In the case signaling is required, the IETF is defining new messages to manage the notification to the MAG of the new flow/prefix to be forwarded.

Fig. 3 shows an example of the initial and resulting routing state of the network upon a flow mobility procedure is completed. Let us suppose the following scenario; An MN (MN 1) is attached to the network through two interfaces `if1`, connected to MAG1, and `if2`, connected to MAG2 and each one receives a prefix, `pref1::/64` for `if1` and `pref2::/64` for `if2` respectively. The MN is receiving two flows, Flow X and Y. Flow X is addressed towards `pref1:lif` (being `lif` the resulting EUI64 identifier of the Logical Interface) and is forwarded through MAG1, while Flow Y is addressed to `pref2:lif` and is forwarded though MAG2. Following this configuration, the LMA has a conceptual data structure called the Flow Mobility Cache containing the mapping of flows and corresponding MAGs. This mapping can be based on any of the flow identifiers defined in [4].

At some point of time the LMA decides to move Flow Y from MAG2 to MAG1. The decision can be based on application profiles, policies or triggered due to network congestion, for instance. In order to do so, the LMA needs to signal MAG1 that Flow Y is going to be forwarded through it. Through some signaling message, the LMA is able to install state in MAG1 regarding the identification of the flow and the identity of MN 1. Once this state is installed on MAG1, the LMA modifies the mapping stored in its Flow Mobility Cache, indicating that Flow Y is routed through MAG1 and starts forwarding the packets towards MAG1. The final state after flow mobility completion of the routing configuration on the network is also presented on Fig. 3.

IV. IP Flow Mobility adoption in 3G architectures

The 3GPP System Aspects Working Group 2 (SA2) has specified the architecture to support simultaneous Packet Data Network (PDN) connections across different accesses for mobile devices equipped with multiple interfaces. Technical Specification 23.402 [10] defines the simultaneous usage of the 3GPP access and non-3GPP access. In this view the mobile customer can receive/send data from/to a 3GPP cellular bearer while taking advantage of a non-3GPP radio access such as WiFi or WiMAX. The following examples give a general overview of the scenarios being considered in 3GPP:

1) A premium customer is connected through 3GPP cellular access as well as the domestic WiFi. He is having several simultaneous flows including a telephony call, a media file synchronization, a video streaming and a p2p

download. Based on operator's policies or user's profile the telephony call and the video streaming are routed via the 3GPP access, while the other two flows are best effort and are routed via the non-3GPP technology.

2) When the premium user moves out of reach of the domestic WiFi, the IP flows on this access are moved to the 3GPP access to ensure seamless service continuity. By means of the multiple PDN support the network will then be able to handover flows while providing uninterrupted services. If later the user moves again in the domestic WiFi coverage best effort flows can eventually moved back to the WiFi connectivity.
3) In addition to the traditional radio coverage problem, the core network might implement methods to dynamically perform traffic steering by exploiting the multi-link diversity and redirecting selected IP flows to the most suitable access network. In this case, for the premium user, the network can steer the IP flow to redirect for instance a video download from the 3GPP to the WiFi access in case the end-to-end QoE measure over the 3GPP access does not meet expectations.

Considering these scenarios, several system requirements can be derived:

- Service continuity should be provided when the MN roams across different accesses.
- Flows should be redistributed across different accesses while connected.
- The MN should be able to exploit multiple radio accesses when possible.
- Different types of services should be provided to customers, i.e., operator-based and non-operator-based.
- Flows can be moved from one access to another in case the MN looses connectivity.
- The Telecom operator should be able to control the simultaneous usage of accesses.
- Changes in the capabilities of the difference accesses (e.g., network congestion) can trigger flow mobility.
- The operator should be able to control flow mobility.

Based on the above, 3GPP investigates two possible alternatives:

1) *DSMIPv6 client-based solution*. This approach is being adopted in 3GPP release 10 and uses the DSMIPv6 protocol stack described in Section II with the extensions for flow mobility specified by the IETF.
2) *PMIPv6/GTP network-based solution*. These solutions exploit the network-based management mobility paradigm and propose supporting multi-homing according to the logic specified in [6]. PMIPv6 extensions are currently being discussed in IETF as described in section III. Also, GPRS Tunneling Protocol (GTP) provides a pre-existing network mobility management alternative to PMIPv6. A slight modification of the 3GPP architecture to accommodate network-based IP flow mobility might be needed and these extensions are being discussed in 3GPP for release 11 and beyond. Both PMIPv6 and GTP solutions would rely on the above-mentioned LIF concept implemented at the terminal.

V. Conclusion

In this article, we have presented and discussed the advantages and drawbacks of the two approaches to enable IP flow mobility that are being standardized in the IETF and 3GPP standards development organizations (SDOs), namely client-based and network-based IP flow mobility. At this stage it is still hard to forecast the evolution of the technology market, however it is clear that Telecom operators are seeking for low cost solutions addressing the smart traffic steering problem beyond classical IP routing functionalities. From the two approaches, network-based flow mobility seems to be a more promising technology that can help Telecom operators that have heterogeneous access networks to extend their network capacity and tier services offerings at low cost, relying on simple software constructs and without modifying the core protocol stack at the terminal.

References

[1] H. Soliman, "Mobile IPv6 Support for Dual Stack Hosts and Routers," RFC 5555 (Proposed Standard), Internet Engineering Task Force, June 2009.

[2] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6," RFC 3775 (Proposed Standard), Internet Engineering Task Force, June 2004.

[3] R. Wakikawa, V. Devarapalli, G. Tsirtsis, T. Ernst, and K. Nagami, "Multiple Care-of Addresses Registration," RFC 5648 (Proposed Standard), Internet Engineering Task Force, Oct. 2009.

[4] G. Tsirtsis, G. Giarreta, H. Soliman, and N. Montavont, "Traffic Selectors for Flow Bindings," RFC 6088 (Proposed Standard), Internet Engineering Task Force, Jan. 2011.

[5] G. Tsirtsis, H. Soliman, N. Montavont, G. Giaretta, and K. Kuladinithi, "Flow Bindings in Mobile IPv6 and Network Mobility (NEMO) Basic Support," RFC 6089 (Proposed Standard), Internet Engineering Task Force, Jan. 2011.

[6] C. J. Bernardos, "Proxy Mobile IPv6 Extensions to Support Flow Mobility," Work in Progress, draft-bernardos-netext-pmipv6-flowmob-02, Internet Engineering Task Force, Feb. 2011.

[7] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy Mobile IPv6," RFC 5213 (Proposed Standard), Internet Engineering Task Force, Aug. 2008.

[8] R. Braden, "Requirements for Internet Hosts - Communication Layers," Internet Engineering Task Force, RFC 1122 (Standard), October 1989.

[9] T. Melia and S. Gundavelli, "Logical Interface Support for multi-mode IP Hosts," Work in Progress, draft-ietf-netext-logical-interface-support-01, Internet Engineering Task Force, Oct. 2010.

[10] 3GPP, "3GPP TS 23.402; Architecture enhancements for non-3GPP accesses," 3GPP, 2011.