Mobility Testbed for 3GPP2-based MMD Networks

Ashutosh Dutta, Kyriakos Manousakis, Subir Das
Telcordia Technologies
Tsunehiko Chiba, Hidetoshi Yokota, Akira Idoue
KDDI R&D Laboratories, Japan
Henning Schulzrinne, Columbia University

Abstract: Wireless service providers strive to preserve the quality of service and user experience for the mobile users. Several standards bodies are defining architecture that can be used as a platform to provide secured and seamless services to these roaming users. However, placement of several functional components and their interaction at several layers contribute to the additional issues and affect the optimal operation. Testbed realization of any standardized architecture can help investigating the underlying networking issues. In this paper, we describe a mobility testbed implementation based on one of the architecture alternatives of 3GPP2, where the outbound signaling servers are distributed around the edges of the network. We experiment with three different handoff techniques and analyze the associated experimental results. Analysis of these experimental results and experiences obtained from the testbed implementation can be helpful to any service provider that plans to deploy the distributed version of the MMD architecture.

1. Introduction

Ubiquitous computing is gradually becoming prevalent in all walks of life. Thus a roaming user needs secured and seamless mobility access as it moves across heterogeneous access networks such as IEEE 802.11, CDMA2000, WIMAX and GPRS. However, additional complexity of the underlying networks, variety of security and bandwidth requirements of each of these types of networks make it difficult to achieve the desired quality of service. In an effort to provide cellular-like dependable but more flexible service to the roaming users over an IP network, several standards bodies such as 3GPP, 3GPP2, ITU-T, IEEE and IETF are working together to define a set of protocols and architectures that can be used by the service providers. 3GPP focuses on defining the architecture and associated functional elements called IMS (IP Multimedia Subsystem) suitable for WCDMA networks [1]. Similarly 3GPP2 has defined an IMS equivalent architecture called MMD (Multimedia Domain) that can be used over CDMA2000 networks [2]. However there has not been any wide scale deployment based on these architectures. Thus it is useful to build the pilot testbeds based on these architectures, analyze the performance and recommend any modifications that may be needed to optimize. Results of experience obtained during the implementation and performance analysis can also be useful to the wireless service providers who plan to build these networks based on IMS/MMD architecture. We highlight some of our experience while building a specific 3GPP2-based architecture. We describe the basic functional components of the testbed and the results of the implementation, but focus our discussion on security and mobility optimization issues.

The rest of the paper is organized as follows. Section 2 provides an overview of the IMS/MMD testbed architecture at a functional level. Section 3 highlights several issues that might affect the

optimization of MMD networks. Section 4 describes the MMD testbed and analyzes three different types of handoff scenarios. Finally, Section 5 concludes the paper.

2. Mobility Management of MMD Architecture

In this section we discuss several physical and functional components of MMD (Multimedia Domain) architecture that form the mobility framework. In a regular CDMA2000 environment, mobility is taken care of at each level of the hierarchy. In a data oriented cellular environment, MN (Mobile Node) usually initiates a data call via BTS (Base Transceiver Station). The BSC (Base Station Controller) responsible for this BTS forwards the call to the associated PCF (Packet Control Function). The PCF selects a PDSN (Packet Data Serving Node) based on certain unique characteristics of the mobile and establishes a GRE 9Generic Routing Encapsulation) tunnel with the PDSN [3]. When the mobile moves between two BTSs within one PDSN, a new PPP session is not warranted. However, if the mobile moves between two BTSs that are controlled by two different PCFs, each PCF may choose a new PDSN in the hierarchy to terminate the PPP session. Besides these core network elements, MMD architecture uses other functional elements at network layer and application layer that take care of mobility binding, signaling, security, and quality of service. Mobility at PDSN layer is taken care of by network layer mobility such as Mobile IP [4], [5], application layer mobility [6] can be used as well.

At a functional level, an MMD architecture primarily consists of several signaling entities such as P-CSCF (Proxy-Call Session Control Function), I-CSCF (Interrogating CSCF), S-CSCF (Serving CSCF), and HSS (Home Subscriber Server) [2]. However, 3GPP2 does not mandate the placement of P-CSCFs in any specific part of the network. Thus an access network provider can place the P-CSCF based on its specific requirement and ease of management. However, placement of P-CSCF may affect the operation of the network in terms of scalability and handover optimization. Authors provide a gap analysis of alternate architecture based on placement of P-CSCF in [7].

Here, we focus our discussion on implementation of the specific architecture where P-CSCF is located in each visited sub-network.

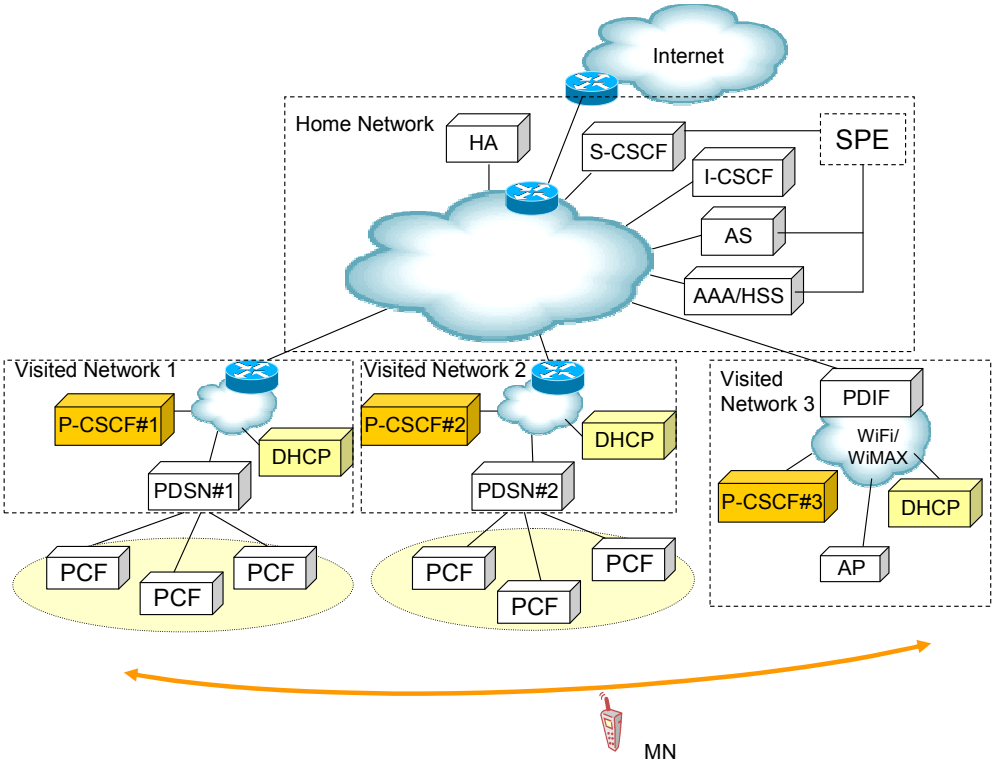

Figure 1: Functional elements of the distributed MMD architecture

Figure 1 shows one such functional architecture where P-CSCFs are distributed across the visited networks. In this specific architecture, there are four network domains labeled as Home Network, Visited Network 1, Visited Network 2 and Visited Network 3. Two of the network domains have CDMA2000 access while the third domain supports either WiFi or WiMAX. A mobile terminal can access its IMS services in any of the visited network domains. Roaming service and mobility is supported by a combination of SIP components such as P-CSCF, S-CSCF, I-CSCF and Mobile IP components such as HA (Home Agent) and FA (Foreign Agent). We provide description of some of the functional components.

MN: MN is the mobile node that moves across the networks.

DHCP Servers: DHCP servers are configuration agents and help the mobiles with the configuration of network layer identifiers such as IP address and address of P-CSCF and DNS servers. Ideally each visited network is equipped with a DHCP server.

HA: HA serves as the anchor point for a mobile and maintains the mapping between mobile's home address and the new care-of-address that mobile obtains in each network. This care-of-address is obtained either from a DHCP server, or from a foreign agent or by means of stateless auto-configuration. In normal case the signaling and media do traverse via the HA that contributes to the additional delay because of trombone routing.

S-CSCF: The S-CSCF is the central node of the signaling plane. It is a SIP server and performs session control functions. It is always located in the home network. It uses the HSS or DNS server to locate the appropriate P-CSCF for outgoing intra-domain calls and the appropriate I-CSCF for inter-domain calls. The S-CSCF in the testbed is implemented as Back-to-Back User Agent (B2BUA), with one UA receiving incoming calls and a second UA sending the invitation to the terminating end.

P-CSCF: The P-CSCF is a SIP proxy and is the first outbound proxy for a mobile in the visited network. The P-CSCF routes REGISTER requests to the I-CSCF and caches the S-CSCF so that it can route the rest of SIP signals directly to the S-CSCF. 3GPP mandates that there should be security association between the mobile and P-CSCF.

I-CSCF: The I-CSCF is another SIP proxy that provides forwarding of messages to the correct S-CSCF, through HSS look-ups.

HSS: The HSS stores information about the subscribers, their addresses, and their services such as user account, contact URI of the user, address of P-CSCF for each mobile, E.164 number etc. HSS is located in the home network and communicates with S-CSCF.

Application Server (AS) - The AS sends messages to the HSS defining the applications deployed on the AS, along with the parameters needed to configure an instance of the service. This description is used by the SPE to generate the system administrators' screens for service configuration and provisioning. The AS also accepts messages from the SPE to configure applications, on a per user basis.

Service Provisioning Environment (SPE) - The SPE is implemented in combination with the HSS, but logically it is a separate component. It provides a mechanism to view the services that are deployed on the AS and system administrators can use it to provision services automatically for each user.

3. Optimization Issues for MMD Network

We briefly describe some of the key issues and metrics that need to be considered for an optimized operation of MMD network.

Context Transfer and Security Association: As the first signaling entity, each P-CSCF keeps the state of certain aspects of an ongoing call. This call state includes certain metrics of the ongoing call such as quality of service (QoS), bandwidth information, calling data record etc. However, in order to maintain the same quality of call in the new network, the existing call state needs to be transferred as quickly as possible to the new P-CSCF. Similarly establishing a security association between the mobile and the new P-CSCF is also needed before media can pass. Thus it is important to devise methodologies that can transfer the call state between the P-CSCFs and expedite the security association proactively. An expedited P-CSCF discovery mechanism can help achieve some of these.

Number of signaling messages over the air: Since the wireless bandwidth is scarce, it is a good design practice to make use of the protocols that can use minimum number of message exchange during the IP address assignment and server discovery. Less amount of round-trip time during these message exchanges also helps to reduce time for server discovery and IP address acquisition.

Number of registrations per P-CSCF: Number of simultaneous registrations a server can handle depends upon the CPU and processing power of the server. Usually mobility rate affects the number of registrations. It also directly relates to the number of call states that a P-CSCF may need to maintain.

Distance between mobile and HA: When the mobile is using MIPv4 to take care of mobility binding, distance between the mobile and the home agent plays an important role. During normal MIP registration and binding update, transient packets are lost in the absence of any fast-handoff mechanism. If a distant home agent is being used to discover the P-CSCF address that also delays the P-CSCF discovery process.

Number of IPSec tunnels for signaling and media traffic: IPSec is one possible way of providing security association for signaling and media traffic. If the authentication and encryption are provided hop-by-hop basis, then a new security association needs to be established every time there is a change in the end-point identifier upon the movement.

Distance between mobile and P-CSCF: Every SIP signaling message to and from the mobile needs to traverse through the P-CSCF and also there is a security association between the mobile and the P-CSCF. Thus there is a tradeoff between having a P-CSCF closer to the mobile and having to change the number of P-CSCFs during a mobile's movement.

User movement pattern: User movement can be intra-domain and inter-domain. During intra-domain movement the mobile's movement is confined within the same DNS and AAA domain. During inter-domain movement both DNS and AAA domains may change and it may involve additional operations and handoff delays.

Packet-to-mobility ratio: Packet-to-mobility ratio depends upon the number of packets communicated during each movement, mobility rate etc. If the mobile performs too frequent handoff and changes its P-CSCF at every move, it will need to re-register and perform MIP update during each move, and that may lead to transient data loss. In such situations it may be advisable to have the P-CSCF closer to the mobile and perform the discovery quickly enough.

4. Testbed Implementation and Results

In this section we describe the testbed that we have prototyped according to the target architecture mentioned in Figure 1. We also present experimental results involving three different handoff scenarios. Figure 2 shows the experimental MMD testbed that emulates an MMD architecture.

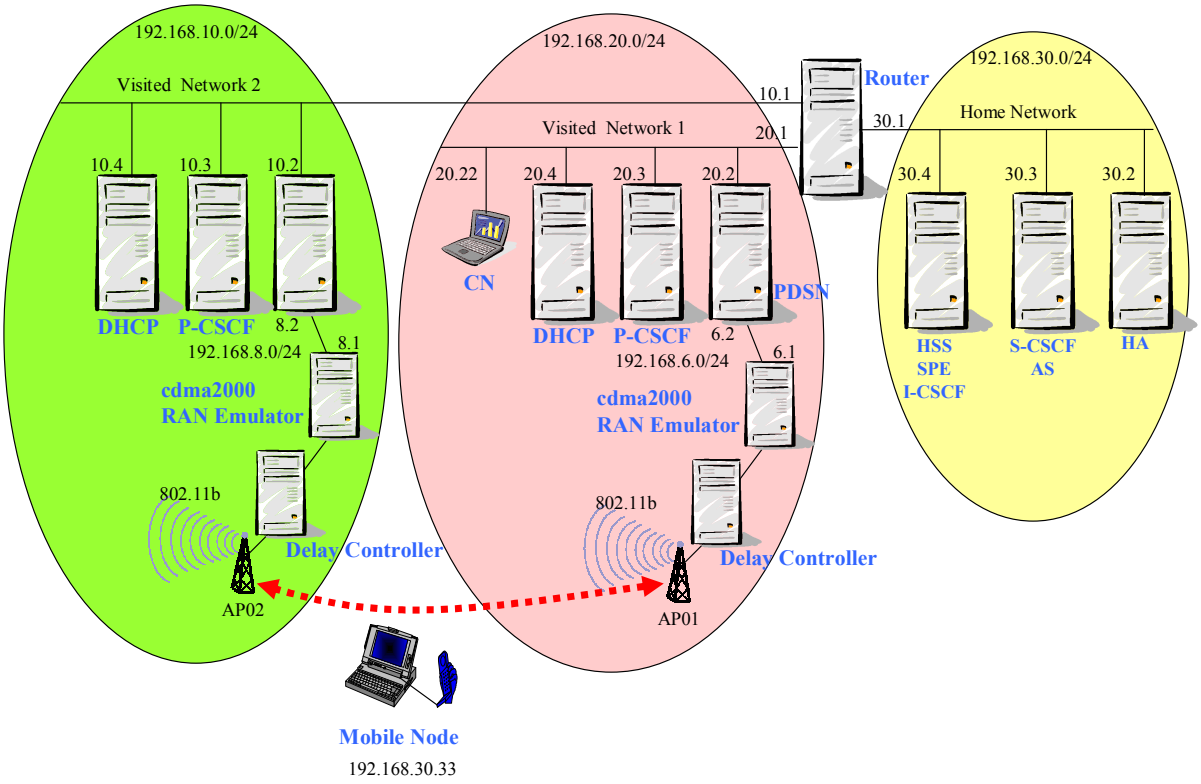

Figure 2. Prototype testbed architecture

Current version of the testbed has three domains. The home domain emulates the components owned and controlled by the carrier; users never roam to their home networks. In addition there are also two visiting domains, that emulate the roaming areas of the mobile nodes. The components in these areas may or may not be owned by the IMS carrier.

The home domain hosts the HSS, S-CSCF, I-CSCF and HA. Each of the visiting domains hosts the P-CSCF. Furthermore, each visiting domain is equipped with a DHCP server that provides the P-CSCF address to the mobile nodes in the corresponding domain via the use of DHCP INFORM message. In the absence of cellular CDMA RAN (Radio Access Network), we have emulated CDMA2000 access over 802.11 networks. The 802.11 AP in each visited network is connected to a CDMA2000 RAN emulator and this emulator is connected to a router that provides PDSN functionality. The traffic generated from the users of the visiting domain goes through the RAN emulator and the PDSN before reaching any of the destinations.

We focus on the measurements and behavior of the IMS system during a mobile's handoff. This study provides the necessary information to understand the details of the system's behavior during the users' handoff. This behavior can be characterized from the various functional operations that take place during the handoff and the relative time taken for each of these operations. The experimental measurements in this specific realistic testbed highlight the performance bottlenecks of the system and indicates what handoff actions may require

improvement for optimization. The following sections provide the details of the handoff experiments.

4.1 Handoff Experiments

User mobility contributes to handoff that results in performance degradation. We provide our analysis of handoff experiments in the IMS testbed that shows the effect of handoff. We illustrate various functions that take place during handoff and measure the corresponding amount of time taken for each of these operations. For completeness of the handoff analysis, we have implemented three representative handoff modes of operation: a) the non-optimized, b) the reactive and c) the proactive. Details of each of these mechanisms are provided in subsequent sections. Initially we provide a brief overview of some of the functions that take place during the handoff operation.

4.1.1 Functional operations during handoff

Following is a list of operations that are executed during a handoff.

Layer 2 Configuration: Whenever the mobile node connects to the new RAN, it goes through the process of re-establishing its PPPoE credentials and obtaining its PPP address. Since the current PPPoE access is provided over IEEE 802.11, it is also subjected to additional 802.11 related handoff delay.

After the layer 2 association is established, the MN initiates the PPPoE Active Discovery (PADI). Upon the reception of PPPoE Active Discovery Offer (PADO) the MN responds with a PPPoE Active Discovery Request (PADR) for selecting the network. This request is acknowledged with a PPPoE Active Discovery Session-Confirmation (PADS). After the discovery of the new network the MN goes through the user authentication process by using LCP, CHAP before an access is granted on the PPP link.

Layer 3 Configuration: Part of the MN's handoff process is the configuration of several layer 3 parameters, such as new care-of-address, default gateway, and P-CSCF server address. While the care-of-address could be provided by the FA, DHCP server or in a stateless manner for IPv6, server configuration parameters are obtained by the corresponding DHCP server in MN's new domain [8]. In the current testbed, the MN broadcasts a DHCP INFORM message to obtain new configuration information and the DHCP server of the corresponding IMS domain responds with a DHCP ACK message which contains the requested information (e.g., P-CSCF IP address).

Mobility Binding: Upon establishment of new PPPoE access, the MN can either listen to Foreign Agent's (FA) advertisements or can solicit Mobile IP advertisement for faster detection. Based on MIP advertisements, the MN determines that it is in a different network and registers its new care-of-address to the home agent. In the current prototype, we are operating on Foreign Agent CoA (FA-CoA) mode, so the IP address of the node does not change and the traffic will be directed to the MN through the address of FA.

Session Registration: After mobility binding and configuration operations are over, the mobile does register with the S-CSCF via the newly configured P-CSCF. This message is intercepted by

FA, is encapsulated and is tunneled to HA. HA sends this to P-CSCF. P-CSCF takes care of registering the mobile with S-CSCF via I-CSCF. Trombone routing caused by Mobile IP adds to the additional delay for the completion of mobile's re-registration. Reverse tunneling between FA and HA forces the SIP signaling destined to P-CSCF to traverse all the way to HA even if the mobile is closer to P-CSCF. Besides, additional encapsulation and decapsulation at HA and FA also contributes to the delay.

Security Association: According to 3GPP, security association needs to be established between MN and P-CSCF before any media can pass through the PDSN. AKA (Authentication and Key Agreement) is a challenge-response based mechanism that uses symmetric cryptography. By means of AKA both MN and P-CSCF can establish security association between the end points. In the testbed we use SIP Register and an out-of-bound protocol to implement an emulation of AKA. This out-of-bound protocol operates between P-CSCF and S-CSCF and obtains the required security key.

Session Maintenance: If the MN moves during an active session, maintenance of this session is needed. In particular, after the Re-registration is complete, the MN retransmits a Re-INVITE message to CN (Correspondent Node). The Re-INVITE message contains the SDP description of the active session. In case of Mobile IP as the mobility protocol, Re-INVITE will carry the home IP address in the SDP. However Re-INVITE's SDP parameters can be used to create new context. This Re-INVITE message is also subjected to trombone routing delay. However Re-INVITE operation does not add delay to the media handoff in case of proactive and reactive handoff because context generation is taken care of by means of context transfer. Only non-optimized handoff is affected due to Re-INVITE.

Media control on PDSN: Establishment of security association between P-CSCF and MN, and context generation by means of new context creation or context transfer from previous P-CSCF are preconditions before the PDSN can allow the media traffic. Once the new P-CSCF has a context and required security association with the mobile, it instructs the gate at the PDSN to allow the media traffic.

The faster re-establishment of security association in the new network and the faster context transfer or context creation helps to open the gate faster at PDSN. Faster gate opening at the new PDSN results in reduced handoff delay and less transient data loss between MN and CN. We describe the details of three different modes of operation that could be possible in a real deployment scenario.

4.1.2 Non-Optimized Handoff Mode

The most basic mode of handoff mode is the non-optimized. This mode is subjected to the maximum amount of handoff delay. In this mode of operation, a new context is created every time the mobile moves to the new domain. The message flow for the non-optimized operational mode is provided in Figure 3.

The MN completes all the handoff functions at layer 2 and layer 3 as described earlier. Specifically, after the MN establishes PPP access to the new domain, it performs the MIP binding functions and it gets the new domain configuration information via DHCP. Then the SIP related handoff functions are performed, starting with the SIP re-registration and the security association establishment using emulated AKA. If the MN moves during an active session, session maintenance is carried out with the transmission of an encrypted SIP Re-INVITE message that carries the SDP description of the ongoing session. Upon receipt of this message, P-CSCF creates a new context for the same mobile and instructs the gate at PDSN to open. This results in the resumption of the media in the new access network.

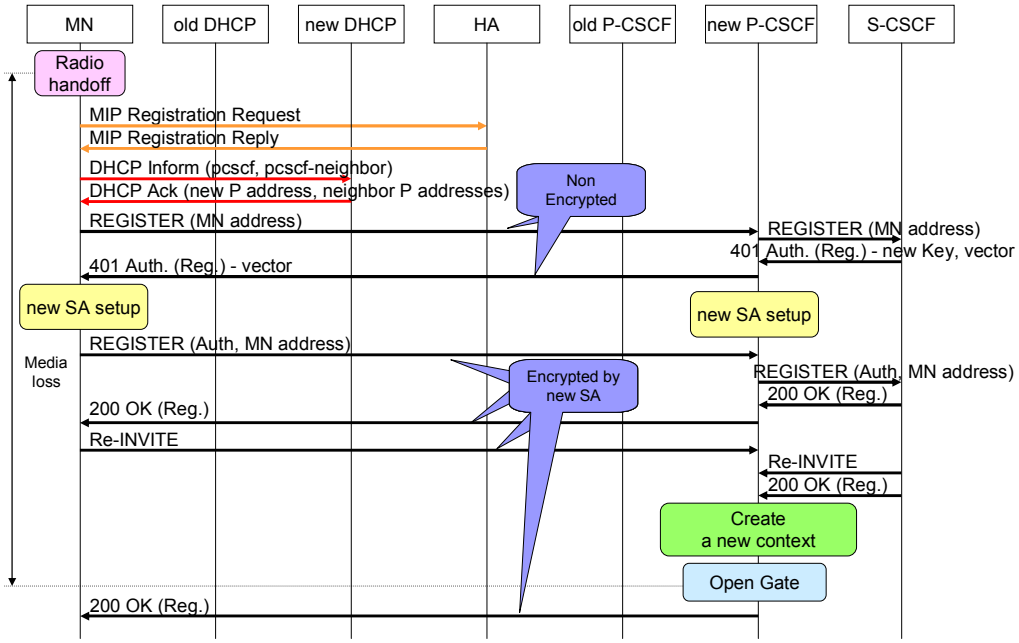

Figure 3. Detailed message flow for the Non-optimized handoff mode

4.1.3 Reactive Handoff Mode

In reactive mode of operation, all the L2 and L3 operations take place like non-optimized mode. The detailed message flow is provided on figure 4. By comparing with figure 3 (non-optimized mode) the difference between the two handoff operational modes are evident. In particular, the session maintenance information message (e.g., Re_INVITE) that carries the SDP description of the active session does not play any role in context creation and thus does not affect the media handoff delay. The context created in the new domain's P-CSCF is transferred from the old domain's P-CSCF. The objective of this approach is to reduce the handoff delay by eliminating the dependence on the session maintenance messages Re-INVITE and 200 OK.

After the radio handoff and establishment of PPPoE access in the new domain, MN performs the regular MIP binding and obtains the required configuration information using DHCP. MN initiates a SIP REGISTER message via the new P-CSCF. When this message reaches the S-

CSCF, the S-CSCF informs the old P-CSCF to transfer the context of the active session to the new P-CSCF. At this point the old P-CSCF transfers the context to the new P-CSCF, the context is created in the new domain. After the completion of the SA setup (e.g., AKA) between MN and new P-CSCF the gate opens at PDSN and the session resumes.

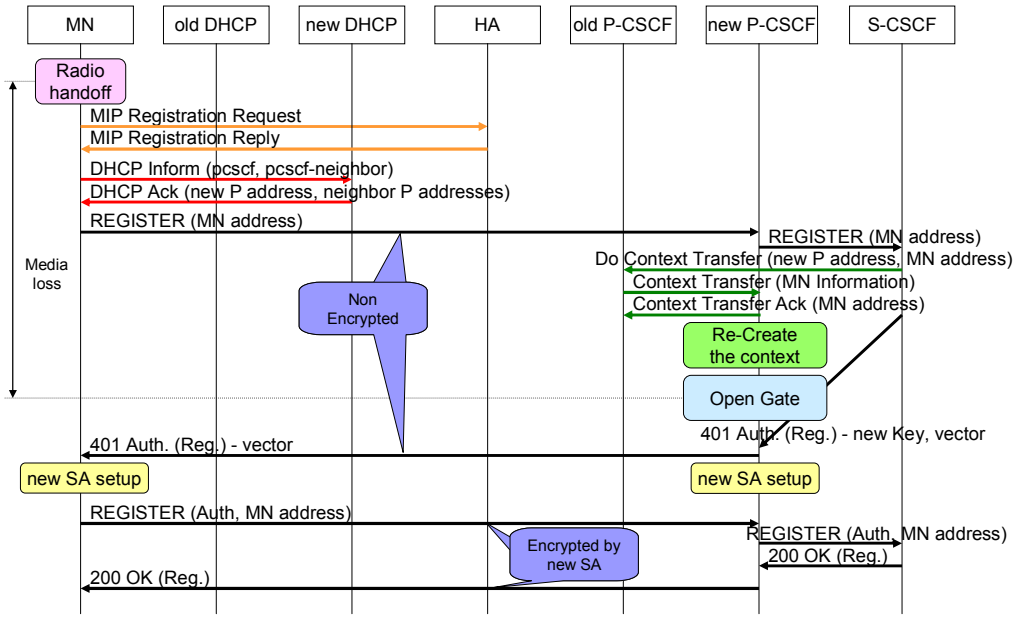

Figure 4: Detailed message flow for the Reactive handoff mode

4.1.4 Proactive Handoff Mode

The most optimized handoff operational mode with respect to performance is the proactive mode. As its name implies, the context creation in the new IMS domain and security association with the new P-CSCF happen while the MN is still at the old network. Even though this specific handoff mode provides the least media delay, it heavily depends upon the discovery of neighboring P-CSCFs ahead of time and accuracy of its movement profile. It is proposed that information service discovery methods such IEEE 802.21 [9] or IEEE 802.11u [10] can provide the information about the neighboring networks. Various techniques such as signal strength threshold can be used to determine mobile's precise movement pattern.

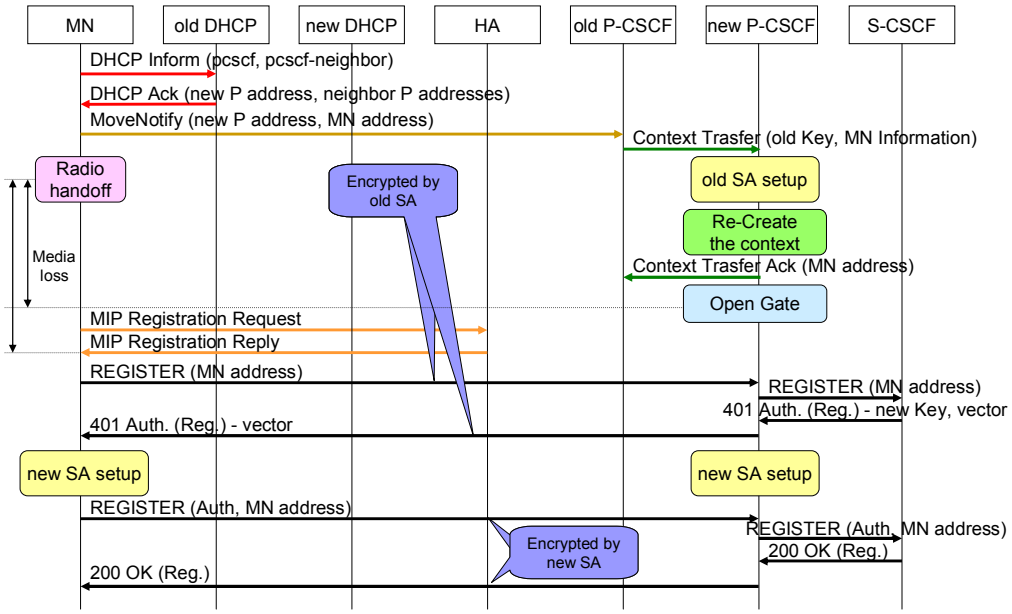

Figure 5: Detailed message flow for the Proactive handoff mode of operation

Figure 5 provides the detailed message flows of proactive handoff. Prior to MN's radio handoff, some of the handoff functions are done proactively in the old network. Specifically, the MN utilizing the DHCP INFORM, acquires the addresses of P-CSCFs from the neighboring IMS domains. In this case, DHCP server is equipped with the information about the servers in the neighboring domains [8]. After the MN has identified the new neighboring domain it is likely to move, it informs its current P-CSCF about the address of its new P-CSCF. The current P-CSCF transfers the context of the active session (e.g., SDP, CDR information) to the new P-CSCF. Similarly, new security association is established between new P-CSCF and the mobile by using a MoveNotify message to S-CSCF. This operation emulates a proactive AKA operation. Thus the new PDSN opens up its gate for this specific mobile's media even before the mobile has moved. After the mobile re-establishes its connection in the new network, and completes the MIP operation, media starts flowing. Mobile's SIP related signaling such as Re-REGISTER or Re-INVITE do not affect the media handoff delay here.

Based on the message flow description, it is obvious that the proactive handoff operational mode is very promising. We compare the results of each handoff operation in the following section.

4.1.5 Performance Results

In Figure 6, we plot the delays associated with the different handoff functions that contribute to the overall handoff delay for three different handoff scenarios. On the average, the mobile was subjected to 3666 ms delay for proactive handoff, 9685 ms delay for reactive handoff and 12526 ms delay for non-optimized handoff. Number of packets lost is proportional to the handoff delay and depends on the traffic generation rate. As is evident, proactive handoff does not have any

delay due to DHCP, context transfer and SIP-based security association compared to reactive or non-optimized case. On the other hand non-optimized case is subjected to maximum delay due to additional signaling messages during SIP-based security association and context creation phase. Layer 2 delay, PPPoE delay and MIP binding delay more or less remain same for all three handoff scenarios. Although proactive handoff takes the least amount of time, there is still room for improvement in optimizing some of functional components such as layer 2 association, PPP activation [3], [11]. A different mobility protocol such as MIPv6 [5] or SIP-based mobility [6] may result in smaller binding update delay.

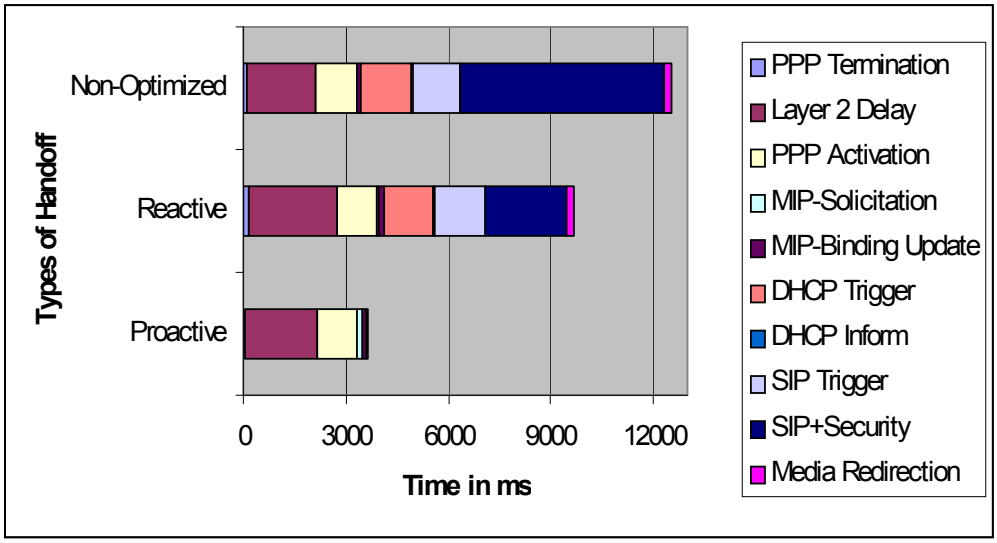

Figure 6. Handoff result analysis for three types of handoff

Similarly, in case of reactive handoff, besides layer 2 delay, major component of the delays came from SIP registration, security association and context transfer. Since we have used Mobile IP, these results are inclusive of inherent trombone routing delays that could be mitigated.

Thus an efficient optimized AKA protocol, faster context transfer and mitigation of trombone routing can optimize the handoff operation.

5. Conclusions

Most of the vendors focus on providing a platform for flexible services but give little importance to the underlying networking issues. However, there are some important and complicated issues that should be investigated in terms of how the protocols and functional components interact with each other in an MMD architecture. These could best be realized by way of a testbed prototype. We discussed a specific target architecture for the testbed prototype according to 3GPP2's MMD specification. In addition to building a MMD compliant mobility testbed, our contribution includes analysis of different handoff mechanisms and associated functional modules that contribute to the handoff delays. Our testbed can be used as an optimized realistic platform for the deployment and evaluation of roaming services. We believe that details of the

implementation and analysis of the handoff results can be beneficial to CDMA2000 wireless service providers.

References

[1] 3GPP TS 23.218: 3rd Generation Partnership Project: Technical Specification Group Core Network; IP Multimedia Session handling; IM call Model"
[2] 3GPP2 X.S0013-004-0 v2.0: All-IP Core Network Multimedia Domain: IP Multimedia Call Control based on SIP and SDP
[3] Kagalkar et al., "PPP Migration: A Technique for Low-Latency Handoff in CDMA2000 Networks," ACM Mobiquitous 2005
[4] C. Perkins et al, "IP Mobility Support in IPv4," IETF RFC 3344
[5] D. Johnson et al, "Mobility Support in IPv6," IETF RFC 3775
[6] Elin Wedlund, Henning Schulzrinne, "Mobility Support using SIP," in IEEE/ACM Multimedia Conference WOMOM 1999.
[7] T. Chiba et al, "Gap Analysis and Architecture Alternatives for 3GPP2 Networks," Submitted to IEEE Vehicular Technology Magazine
[8] H. Schulzrinne, "Dynamic Host Configuration Protocol (DHCP-for-IPv4) Option for Session Initiation Protocol," IETF RFC 3361
[9] "Draft IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handover Services, IEEE P802.21/D00.01", A contribution to IEEE 802.21 WG
[10] www.ieee.org
[11] Arunesh Mishra, et al **"**An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process," in ACM SIGCOMM Computer Communications Review (ACM CCR), Vol 33 Issue 2, April, 2003