Vertical Handoff Characterization for SIP and mSCTP Based UMTS-WLAN Integration Solutions

Syed Asadullah, Ashraf S. Mahmoud, Marwan Abu-Amara, Tarek Sheltami
Computer Engineering Department
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
Email: {sasad, ashraf, marwan, tarek}@kfupm.edu.sa

***Abstract* — It is desirable to integrate 3G Universal Mobile Telecommunication System (UMTS) and 802.11 wireless local area networks, especially at hot-spot locations such as hotels and airports. The efficiency of wireless data services can be maximized if the integration provides users with seamless roaming across the two types of networks. Seamless handoff between these two networks to maintain session continuity is a major challenge in WLAN-3G integration. To achieve this goal, integration architectures together with mobility solutions such mobile stream control transmission protocol (mSCTP) and session initiation protocol (SIP) have been proposed in the literature. In this paper, we implement through simulations an integration architecture and characterize the vertical handoff delay for both mobility solutions mSCTP and SIP as a function of network parameters. This study finds that mSCTP perform better in terms of handoff delay compared to SIP for the assumptions specified in this paper.**

***Index Terms* — WLAN, 3G, UMTS, mSCTP, SIP handoff, mobility.**

I. Introduction

The proliferation of wireless local area networks (WLANs) has provided network service providers with an option of integration with third-generation (3G) wireless wide area networks, such as Universal Mobile Telecommunications System (UMTS). Such integration allows mobile users to move among these heterogeneous networks in a seamless manner. However, the integration of 3G networks and WLANs presents some considerable challenges which include the demand for seamless handoff, continuity of data traffic and multimedia sessions across the two networks, central authentication system, billing, security etc. To deal with these challenges and to support this type of integration several 3G/WLAN interworking architectural scenarios have been proposed [1]. These architectures differ in terms of the extent of interoperability or services they provide.

Vertical handoff is a major factor in any UMTS/WLAN integration. Vertical handoff refers to the handoff between two heterogeneous networks such as UMTS-to-WLAN or WLAN-to-UMTS. It is desired that the data sessions that a mobile user maintains during handoff stay alive, that the handoff is seamless, and that the handoff spans only a very short period of time. To achieve seamless handoff, several mobility protocols have been proposed. The paper considers two mobility protocols; SIP and mSCTP. SIP is an application layer protocol while mSCTP works at the transport layer.

The paper is outlined as follows. In section 2 we briefly review the WLAN and UMTS integration architecture. In section 3 we discuss UMTS-to-WLAN vertical handoff and how it is handled using SIP and mSCTP. Similarly, we discuss WLAN-to-UMTS handoff scenario in section 4. In Section 5 we define the simulation setup and the parameters used. The simulation results and analysis for vertical handoff using SIP and mSCTP are presented in section 6. In section 7 we present the conclusions.

II. Network Architectures

The UMTS-WLAN integration architecture adopted in this paper is shown in Figure 1. The network clearly consists of the UMTS network, the WLAN network and the Internet service provider network. The following presents a brief description of the WLAN and UMTS networks and their main components.

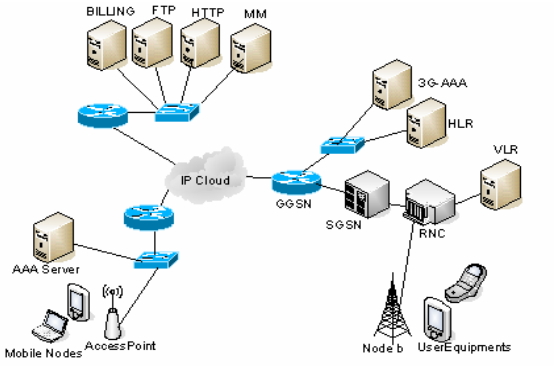

Figure 1: Integrated UMTS-WLAN network architecture.

WLAN Architecture: The 802.11b/g/a WLAN architecture basically consists of one or more Basic Service Sets (BSSs) also called access points (APs) and

client devices. The traffic for all the client devices that are associated with one AP flows through the particular AP. All the APs are connected to the network backbone through a switch or a hub.

UMTS Network Architecture: The basic UMTS architecture consists of three domains: The User Equipment (UE), the UMTS Terrestrial Radio Access Network (UTRAN) and the Core Network (CN). The UE consists of the equipment used by the user to access UMTS services. The UTRAN consists of one or more Radio Network Sub-systems which can further have one or more Node Bs connected to one Radio Network Controller (RNC). The coverage area of Node B is called a cell. CN consists of Circuit switched networks for providing voice and circuit switched services, Packet Switched network (PS) for providing packet based services, and optionally a Home Location Register (HLR). HLR stores subscribers' location and maintains a list of services allowed for each subscriber.

Among several functional entities, the PS network consists of a Serving GPRS Support Node (SGSN) responsible for routing packets inside the PS as well as for mobility management, logical link management, and authentication and charging functions, and a Gateway GPRS Support Node (GGSN) acting as a gateway towards external packet switched networks. The external network could be a LAN, WAN, GPRS, ATM network etc.

III. UMTS-TO-WLAN VERTICAL HANDOFF

When a mobile node (MN) enters a foreign (WLAN) network, it identifies the presence of a WLAN by receiving the characteristic beacons from an AP. To start using the services of the new network, the MN has to authenticate itself with the WLAN. Many authentication methods exist and can be employed depending on the agreement between the UMTS/WLAN service providers. The paper uses authentication, authorization and accounting (AAA) authentication and extensible authentication protocol (EAP) signaling for WLAN authentication which is briefly discussed next.

WLAN Authentication: During the EAP signaling [2] for MN authentication and key exchange, the MN initially sends a request to the AP to connect to the WLAN and receives a response for authentication type from the AP. The MN submits an authentication request to the AP that is forwarded to the AAA server. The AAA server processes the request and sends challenge back to the client. The client submits its credentials to the AAA server that the AAA verifies and allows or denies access based on local user database or contacting an external user database. This is implementation dependent and can be EAP-MD5, EAP-MSCHAP v2 etc. AAA also sends a session key to the MN along with the response. The authentication is valid as long as the client is associated with the same AP. If the MN moves from one AP to another it re-authenticates itself during handoff. After successful authentication the MN begins to acquire an IP address from the WLAN network through dynamic host configuration protocol (DHCP) registration. Thus, any client that enters WLAN network needs to authenticate and acquire an IP address before it triggers its protocol specific handoff procedures. Two such procedures, SIP and mSCTP, are discussed next.

SIP: Session Initiation Protocol (SIP) [3] is basically a signaling protocol that offers a number of benefits, including extensibility and provision for call/session control. It is mainly used to establish, modify, and terminate multimedia sessions. Apart from the signaling function, SIP is an application layer protocol which inherently supports terminal mobility. SIP users are addressed using email-like addresses, sip:user@domain for example. The logical entities in SIP communication are user agents which are the nodes communicating using SIP, registrar server which is responsible for maintaining user agent access information, redirect server that keeps track of the user's location so as to redirect requests in case of location change, and proxy server responsible for relaying the messages. The proxy server can be Outbound, Inbound or Intermediate. The standard methods defined in SIP for setting up sessions between user agents are SIP — INVITE, ACK, BYE, OPTIONS, CANCEL and REGISTER. Figure 2 shows the SIP-based mobility management solution.

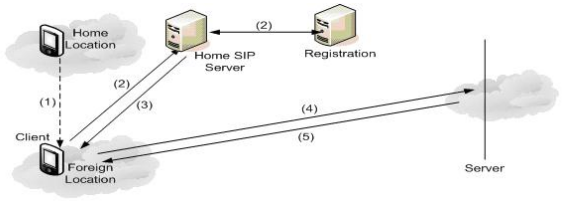

Figure 2: SIP-based mobility management solution.

When an MN moves into the coverage area of the foreign WLAN network, step (1), it authenticates itself with the new network, obtains a new IP address, and triggers the SIP message exchange. The MN then sends a location update to the home SIP server, step (2), so that new invite requests can be redirected to the correct address. The registrar processes the update message and forwards it to the location service, which stores the information. The home SIP server responds to the MN with an acknowledgment, step (3). In step (4), the MN sends a new SIP-INVITE request to the Correspondent Host's (CH) user agent using the same call identifiers as in the original connection setup. The request contains

the new address. Any new message from the server to the MN is sent to this new address. The CH's user agent acknowledges the request, step (5), and this completes the handoff procedure. The combined delay due to WLAN authentication, DHCP registration and SIP Location Update and INVITE procedures constitute the UMTS-to-WLAN handoff delay.

mSCTP: Mobile Stream Control Transmission Protocol (mSCTP) [4] is an extension of SCTP [5] with the addition of the Dynamic Address Reconfiguration (DAR) feature that is also referred to as ADDIP. The mSCTP protocol provides mobility management at the transport layer. An mSCTP enabled node has the provision of acquiring and holding multiple IP addresses while keeping the end-to-end connection intact. During initiation of the connection, a list of addresses is exchanged between the endpoints. One address (primary) is used as the destination for normal transmission and the other addresses are used for retransmissions only. With mSCTP nodes can add, delete and change the primary address dynamically while still being connected. The major advantage of mSCTP over other mobility protocols is that it does not require any additional infrastructure but the entities communicating at both ends must support mSCTP. The Address Configuration (ASCONF) message transactions for mSCTP mobility management are shown is Figure 3.

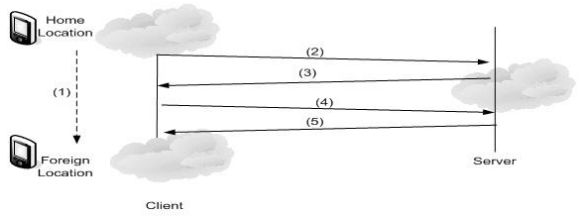

Figure 3: mSCTP Mobility Management.

When a Mobile Node moves into the coverage area of the foreign WLAN network, step (1), it authenticates itself with the new network and obtains an IP address from the local DHCP server through standard registration procedure. The MN then informs the entity at the other end (CH) about the new IP address through the ASCONF message ADD-IP, step (2). The CH updates this IP address in the list and sends an acknowledgement, step (3). The MN further informs the CH to set the newly assigned IP address to the primary IP address, step (4), which the CH responds to with an acknowledgment, step (5). The new primary IP address now becomes the destination address for further communication. With this the handoff procedure is complete and the time required to complete the ASCONF messages (steps 2 to 5) constitutes the UMTS-to-WLAN handoff delay. Due to the dual homing feature of mSCTP, the MN continues to communicate with the CN through its UMTS IP while it is authenticating itself with the WLAN and acquiring an IP address from the DHCP server in the new location. Hence, the WLAN authentication and DHCP registration delay does not add to the handoff delay in UMTS-to-WLAN handoff using mSCTP.

IV. WLAN-TO- UMTS VERTICAL HANDOFF

When an MN moves from a WLAN to a UMTS network, it is required to perform two key functions prior to initiating a handoff [6], namely the GPRS Attach and the Packet Data Protocol (PDP) Context Activation [7][8]. This establishes a data connection setup used to carry protocol specific handoff messages. The GPRS attach is analogous to the WLAN authentication and PDP Context Activation is analogous to the DHCP registration in the UMTS-to-WLAN handoff. The messages involved in the GPRS Attach and PDP Context Activation procedures are shown in Figure 4.

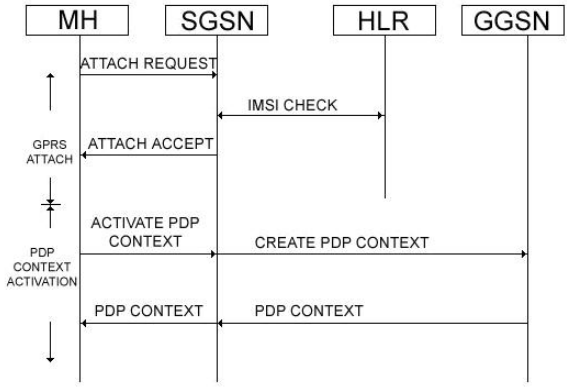

Figure 4: GPRS Attach & PDP Context Activation.

GPRS Attach: As part of the GPRS Attach procedure [7], the MN sends an Attach Request message to the SGSN with the MN's international subscriber identifier (IMSI). The SGSN uses the IMSI to authenticate the MN with its HLR. Successful authentication is followed by the SGSN sending a location update to the HLR. The SGSN finally completes the procedure by sending an Attach Accept message to the MN. This establishes a logical association between the MN and the SGSN.

PDP Context Activation: After the GPRS attach, the MN must activate a PDP address (or IP address) to begin packet data communication. Activation of a PDP address creates an association between the MN's current SGSN and the GGSN that anchors the PDP address. A record of such an association is known as the PDP context. PDP context transfer is initiated by the MN by sending a PDP Context Activation message to the SGSN. After receiving the Activation message, the

SGSN discovers the appropriate GGSN & selects a GGSN capable of performing the function. The SGSN and GGSN create special paths for the transfer of the respective mobility protocol messages. Once the MN gets attached to the UMTS network and activates its PDP context, then dependent on the type of handoff procedure used, the MN triggers either a SIP handoff procedure or an mSCTP handoff procedure similar to what was discussed in section 3. If a SIP handoff procedure is triggered, then the time required for GPRS Attach and PDP context activation plus the delay due to SIP REGISTER/UPDATE messages constitute the handoff delay for WLAN-to-UMTS handoff. On the other hand, if an mSCTP handoff procedure is triggered, then the handoff delay for WLAN-to-UMTS handoff is the delay due to the ASCONF messages.

V. Simulation Setup & Parameters

A SIP-based integration architecture is design and implemented using Opnet Modeler as shown in Figure 5. For handoff between heterogeneous networks it is required that the mobile node has a dual mode network interface which supports both UMTS and WLAN networks. As Opnet does not support such interface we have built custom tasks to simulate the handoff instances. A custom task is a collection of communication transactions between network entities. In our simulation we have customized tasks for AAA Authentication, DHCP address lease, SIP Handoff, mSCTP handoff, and UMTS Authentication. The exact transactions, as stated by the corresponding RFCs, required to perform these tasks were built into the simulation code.

The handoff task is simulated as follows. Two separate nodes are used in WLAN and UMTS network which serve the purpose of one MN having dual mode network interface. Suppose a node is moving from UMTS to WLAN network, the first set of handoff transactions are initiated between UMTS node and the Correspondent Host. Once the MN moves into the WLAN area the remaining transactions are carried out between the corresponding WLAN node and correspondent host. Hence in this work we interested in quantifying the time required to complete the transactions that take place over the network and is recorded using the above mentioned tasks. These transactions vary depending on the mobility protocol being used and hence we compare the delay incurred using different mobility protocols. In our simulation the packet sizes for mSCTP and SIP transactions, WLAN and UMTS authentication etc. were taken from related sources [2][3][4][9][10]. The implemented architecture includes entities such as the mobile nodes, DHCP server, SIP servers, interconnecting routers and switches etc. For our analysis, the correspondent host is the HTTP_Server.

Comparison of handoff delays at protocol level does not require actual simulation of physical layer and hence Ethernet nodes were used as mobile nodes to save simulation time and support higher traffic.

For handoff delay comparison the simulations were run with WLAN and UMTS users using the services from their respective domains. One node was selected to initiate the handoff when it moves from UMTS to WLAN network or vice versa. Mobility in both directions was considered. For each traffic load the handoff delay was noted at five different instances during simulation and the average of these five values is considered to be the effective handoff delay for the load.

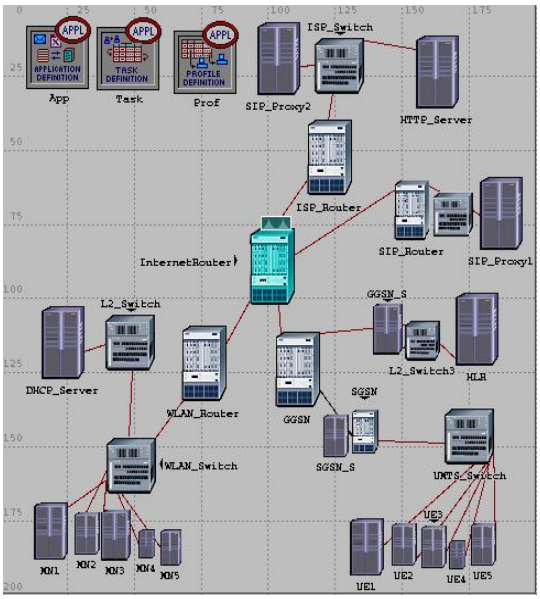

Figure 5: SIP-based integration architecture implementation using Opnet Modeler.

VI. Handoff Delay Comparison

Figure 6 shows the comparison of handoff delay vs traffic load for mSCTP and SIP. Figure 7 shows effect of background load on the handoff delay. Results are presented in this section for mobility in both directions.

WLAN-UMTS: The top curve in the Figure 6.1 is the handoff delay for WLAN-UMTS using SIP where the values for handoff delay are around 3 – 3.5 seconds. The delay for mSCTP is much less than in SIP. This can be attributed to the fact that when the Mobile node moves into the UMTS network requires GPRS attach and PDP context activation which consumes considerable amount of time. An mSCTP enabled node also goes through these steps but with a difference that it still maintains connectivity through its WLAN IP. Hence the GPRS attach and PDP context activation do not add to the handoff delay in case of mSCTP.

UMTS-WLAN: Even in the case of a UMTS-WLAN handoff mSCTP performs better than SIP as mSCTP requires fewer transactions than SIP for handoff completion. It can also be noted from the results presented below that the difference between the handoff times (SIP and mSCTP) is not significantly high initially but with increase in the traffic on the network the performance of SIP handoff degrades fast. This is mainly due to the additional communication and signaling between SIP entities viz. SIP Proxy, Registrar etc. for location update and registration. With increasing load this communication takes longer time to traverse through the network. The background load figures also indicate that SIP handoff time increases dramatically with increase in the background traffic on the network.

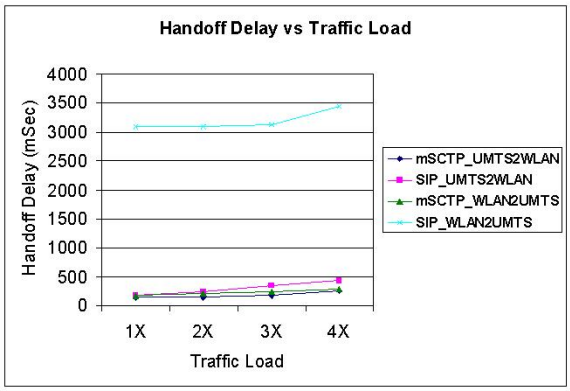

Figure 6: Handoff Delay vs Traffic Load comparison for mSCTP and SIP.

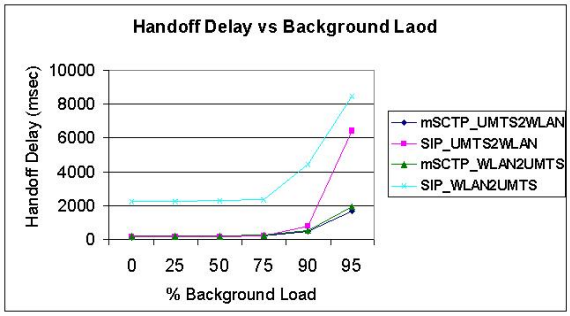

Figure 7: Handoff Delay vs Background Load comparison for mSCTP and SIP.

VII. CONCLUSION

It is quite evident that there are still several challenges to be addressed to enable seamless integration of wireless LAN and UMTS networks. Various architectures have been proposed to meet these challenges. This paper provides an implementation using simulation platform (Opnet) for an integration architecture that supports two mobility schemes: mSCTP and SIP. We also use our implementation to quantify the vertical handoff delay for such network. Based on the results, we can also conclude that with respect to mobility mSCTP provides faster vertical handoffs than SIP. Moreover SIP also requires additional entities at various levels like SIP proxy server, registration server and intermediate proxy servers while mSCTP does not have any of these requirements. However end nodes must be mSCTP enabled to support dual homing feature and achieve faster handoffs.

REFERENCES

[1] Apostolis K. Salkintzis, "Interworking Techniques and Architectures For WLAN/3g Integration Toward 4g Mobile Data Networks," *IEEE Wireless Communications,* June 2004

[2] RFC 3748 - Extensible Authentication Protocol (EAP)

[3] RFC 3261 - SIP: Session Initiation Protocol

[4] SCTP Extensions for Dynamic Reconfiguration of IP Addresses and Enforcement of Flow and Message Limits Network Working Group, *INTERNET-DRAFT,* June 6, 2001

[5] Seok Joo Koh, Moon Jeong Chang, and Meejeong Lee, "mSCTP for Soft Handover in Transport Layer," *IEEE COMMUNICATIONS LETTERS,* Vol. 8, No. 3, March 2004

[6] Farhan Siddiqui, Sherali Zeadally, "Mobility management across hybrid wireless networks:Trends and challenges," *Computer Communications 29*, 2006, 1363 - 1385

[7] Wei Wu, Nilanjan Banerjee, "SIP-Based Vertical Handoff Between WWANs and WLANs," *IEEE Wireless Communications*, June 2005

[8] Nilanjan Banerjee, Arup Acharya, T. J. Watson, "Seamless SIP-Based Mobility for Multimedia Applications," *IEEE Network,* March/April 2006.

[9] RFC 4187 - Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA)

[10] RFC 2960 - Stream Control Transmission Protocol