A Comparative Analysis of Different Integration Approaches Between UMTS and WLAN Networks

Ismat Maarouf, *Member, IEEE,* Mohammed Aabed, *Member, IEEE,*

***Abstract*—The rapid evolution of wireless and mobile communication technologies raises the challenge of satisfying different technology features in a unified architecture. For example, while WLAN technologies provide QoS and high data rates, they lack the wide coverage and powerful mobility management supported by UMTS. On the other hand UMTS users suffer from lower QoS compared to WLAN. Looking at the combined benefit of both heterogeneous networks is the motivating factor of UMTS/WLAN integration, the very hot research topic in the literature. This paper proposes different solutions for UMTS/WLAN integration and vertical handover problems. The solution is based on an implementation of two coupling schemes that are later enhanced to support handover solutions. The work involves network simulation and comparative analysis for the proposed solutions, as well.**

***Index Terms*—UMTS, WLAN, 3GPP, internetworking, network architecture, QoS, beyond 3G.**

I. INTRODUCTION

Mobile and wireless telecommunication has become an essential characteristic of daily life. With the boost integration of these technologies in people's routine, higher expectations in terms of quality of service (QoS) and more requirements for better services arise. The existing generation of mobile and wireless network is capable of providing endless streams of voice and data information. Nonetheless, there are still limitations due to the deficiency in the integrity between the different protocols that operate these networks. The requirement and standards for beyond third generation (B3G) wireless networks demand for seamless mobility and transparent handover between heterogeneous networks. In other words, users will be able to roam into different kinds of networks without loosing connection and be able to access services provided by other networks as well.

This work is concerned with studying, characterizing and proposing the integration between Universal Mobile Telecommunication System (UMTS) and Wireless Local Area Networks (WLANs). UMTS is the 3G successor for the Global System for Mobile Communication (GSM). It was the European answer to the ITU IMT-2000 standard for 3G networks. UMTS utilizes W-CDMA air interface and a lot of the GSM infrastructure. On the other hand, WLANs is the cordless solution for packet switched services. They can be used either to replace wired LANs or as an extension of the wired LAN infrastructure. WLAN, which is commonly used to refer to the IEEE 802.11 series of standards, implement different versions of RAT with variations in data rates and interference tolerance potential. Using direct-sequence spreadspectrum (DSSS) in the 2.4-GHz band, IEEE 802.11b can provide 5.5 and 11 Mbps in addition to the 1-and 2-Mbps operation. In 802.11a, the supported data rates range from 6 to 54 Mbps with Orthogonal Frequency Division Multiplexing (OFDM) in the 5-GHz band. A higher rate extension in the 2.4-GHz band can be achieved with 802.11g technology.

The main contribution of this work is the provision of a comparative study between different integration approaches between UMTS and WLAN network. It firstly evaluates the integrated system under two coupling scheme i.e. open coupling and loose coupling. The study is then extended to evaluate three handover mechanisms over each proposed integration scheme. The handover approaches are mobile IP (network layer solution), mobile Stream Control Transmission protocol "mSCTP" (transport layer solution) and Session Initiation Protocol "SIP" (application layer protocol). The integration plans follow the requirements in the latest set of standards released by the 3rd Generation Partnership Project (3GPP), i.e. Release 7.

The rest of the paper is organized as follows: Section II provides a background material related to the integration problem. In section III, we provide some related work from the literature. This is followed by a detailed description of the design and implementation of the proposed integrated networks system solution in IV. Section V provides the simulation results and analysis for the integration approaches and the handover schemes. Finally, the paper concludes in section VI.

II. BACKGROUND

The scope of this work considers the integration between WLAN and UMTS. In what follows, a brief description of integration scenarios proposed in literature and the possible coupling schemes are discussed. After that, the protocols used in the handover solution, i.e. mobile IP, mSCTP and SIP, are briefly explained. Finally, some integration related functionalities are explained.

A. *Integration Scenarios and Coupling Schemes*

To meet WLAN/UMTS integration requirements, several integration scenarios are proposed in the literature. Theses scenarios focus on four general requirements, i.e. customer authentication, billing and tracking issues, service provisioning offers, and handover issues. According to [12] and [15], there are 6 different WLAN/UMTS inter networking scenarios:

- Scenario 1: Common Billing and Customer Care.
- Scenario 2: 3G-Based Access Control and Charging.
- Scenario 3: Access to 3G Packet-Switched Services.
- Scenario 4: Access to 3G Packet-Switched-Based Services with Service Continuity.
- Scenario 5: Access to 3G Packet-Switched-Based Services with Seamless Service Continuity.
- Scenario 6: Access to 3G Circuit-Switched-Based Services with Seamless mobility.

The scenario specifications play the major role in determining the level of integration or “coupling” scheme. In this work, the target is to simulate the specifications of scenario 5 yet other scenarios are indeed presented when studying a low level of coupling. Here, we describe briefly the possible coupling approaches implemented in this work.

- Open Coupling: In this scheme, there is no real integration between the two access technologies. The WLAN and UMTS networks are considered as two independent systems that share a single billing scheme between them. Although a common database is used between the two; separate authentication procedures are used. [3]
- Loose Coupling: In this approach, there is a common customer database and authentication procedure. The operator will still be able to utilize the same subscriber database for existing 3G clients and the new network (WLAN) clients, allowing centralized billing and maintenance for different technologies. WLAN users should utilize the common subscriber database without any user plane interface to the UMTS GSN nodes.[3]

B. *Mobile IP Overview*

Mobile IP is a network layer mobility management solution. The implementation of Mobile IP considers three main components

- Mobile Node
- Home Agent
- Foreign Agent

The mobile node (MN) is a device such as a cell phone which has the network roaming capabilities. The home agent (HA) is a router on the home network serving as the anchor point for communication with the mobile node; it tunnels packets from a device on the Internet, called a correspondent node (CN), to the roaming mobile node. The foreign agent (FA) is a router that may function as the point of attachment for the mobile node when it roams to a foreign network, delivering packets from the home agent to the mobile node. The FA assigns the MN an address called, the care-of address (CoA) which is the termination point of the tunnel toward the MN when it is in a foreign network.

The Mobile IP process has three main phases

- Agent Discovery: The MN discovers its FA and HA.
- Registration: The MN registers its current location with the FA and HA.
- Tunneling: A reciprocal tunnel is set up by the HA to the CoA to route packets to the MN.

When a mobile node moves to a foreign area, it first discovers the foreign agent. After that it is given a care of address CoA by that agent, which is delivered to the home agent. When a CN wants to talk to the MN, it contacts the home address. The packets are then intercepted by the HA and then tunneled to the FA which encapsulates the tunneled IP packet and deliver it to the MN. An optimization can be done if the CN is informed about the CoA of the MN. Then the packets can be directly sent to the FA without an interception from the HA. More details on mobile IP can be found in [2].

C. *mSCTP Overview*

SCTP is a transport protocol that is described at the same level of TCP and UDP. Like TCP, SCTP exhibits the feature of connection-orientation where an end to end connection is established priori to data transmission and the connection is maintained till the end of the user session. There are several important differences between SCTP and TCP. However, the main feature of SCTP over TCP that is of concern in the handover operation is multi-homing. Multi-homing, is the ability for single SCTP endpoint to support multiple IP addresses. SCTP does that by assigning multiple IP addresses belonging to different access networks to a single end point among which one is considered to be a primary address. Normally, a session will use the primary IP for the duration of the session. However, on failure detection, data chunks are retransmitted using the other IP addresses to guarantee reachability. Ultimately, a decision is made to make all transmission to one alternate until the primary address come to life again. [5] In the base version of SCTP, two mobile hosts (MH) A and B must establish an SCTP connection association (list of the A IP addresses + port number, list of B IP addresses + port number). This association is established before the start of a session and can not change during the session. This implies that if a

MH moves during an active session from one area of a specific IP to a new access network in which he does not have an IP recorded in the early association, he will lose the connection. This is actually the mobility case in UMTS/WLAN architecture when we look at the problem of vertical handover. The solution is to extend IP configuration during a session rather than restricting it to the connection establishment phase. This extension in SCTP is called mobile SCTP (mSCTP) and is proposed by DAR extension where IP address configuration messages (ASCONF) are exchanged between the MH and a fixed server (FS) using a client-server model. [8]

D. SIP Overview

SIP is a client-server application layer protocol that is used to establish and tear down unicast and multicast multimedia sessions. Basically, SIP entities are either user agents or network servers. A user agent is an end system that acts on behalf of someone who wants to participate in calls. In general, a user agent contains a user agent client protocol (UAC) and a user agent sever protocol (UAS) to allow peer to peer operations.

In addition to user agents, SIP provides for two different types of network servers: proxy and redirect servers. A SIP proxy receives a request, determines which server to send it to (another proxy, redirect or a UAS) and then forwards the request. A redirect server receives requests, but instead of forwarding them to the next hop server, it tells the client to contact the next hop server directly. [6] In order to participate in SIP sessions, clients register with any type of SIP network servers. Since the SIP proxy server relays SIP messages, it is possible to use a domain name to find a user, rather than knowing the IP address or name of the host. In fact SIP clients are identified as in email addresses i.e. hostname@domain. Moreover, A SIP proxy can also be used to hide the location of the user. In addition, since a redirect server returns the location of the host rather than relaying the SIP message. This makes it possible to build highly scalable servers, since it only has to send back a response with the correct location, instead of participating in the whole transaction which is the case for the SIP proxy. Both the redirect and proxy server accepts registrations from users, in which the current location of the user is given. The location can be stored either locally at the SIP server, or in a dedicated location server. Deployment of SIP servers enables personal mobility, since a user can register with the server independently of location, and thus be found even if the user is changing location or communication device. [7]

III. RELATED WORK

In this section we summarize some of the related work in the literature.

In [13] , the authors evaluated the performance of dual-mode radio access protocol design through experimental simulations using various types of applications reflecting different QoS classes such as Voice over IP (VoIP) in GSM encoded format, FTP, and HTTP (web browsing). The parameters they measured regarding the integration problem included end-to-end packet delay, file upload time and HTTP page response time. They have shown some interesting results. For example, they found that in the UMTS coverage area, the FTP upload time is on an average 20.071 seconds and as the user enters into the WLAN coverage and switches transmission through the WLAN interface, the upload response time drops to an average of 0.6 seconds.

In [9], the authors studied and simulate the handover procedures using mSCTP. They described in details the handover process under two Fixed Server homing approaches: single homing where the server is configured with one IP and dual homing in which the FS has two IP addresses. The authors provided performance analysis regarding handover delays and overall throughput. They showed that dual homing performs better because duplicated buffered data transmission over both old and new paths resulted from the duality feature may help the receiver and sender to adapt to a sudden change in link characteristics easily and quickly during and after a vertical handover.

The work in [1] analyzes the handover problem using SIP. The authors provided numerical analysis regarding handover delays. They showed that the WLAN-to-UMTS handover incurs much larger delay than the UMTS-to-WLAN handover due to error-prone and bandwidth-limited wireless links in the former situation.

Mobile IP approach was discussed in [4]. The authors compared mobile IP approach with tight coupling and gateway approach. Their simulation showed that mobile IP suffers from long handover latency and might not be able to offer real-time services and applications. They suggested as a future work to improve handover speed in mobile IP. The tight coupling proved better performance but it is less flexible, as they explained in terms of standardization requirements.

IV. NETWORK DESIGN AND IMPLEMENTATION

In this section we start by describing our proposed network architecture with respect to coupling schemes. This is followed by a description of the different implementations for the three handover solutions

A. Network Coupling Design and Implementation

1) Open Coupling

According to [3] and [16], the open coupling WLAN/UMTS integration mainly targets scenario 1 and partially scenario 2 as discussed in [14]. In theses scenarios, the integration provides a common charging and billing system without any provision for integrating data services. This means that the IP network access for WLAN users will stay separate from the UMTS IP access. Moreover, UMTS users will not be able to benefit from the WLAN access network. Therefore, the integration requirement is very minimal and is restricted to the common customer care and billing system CC&BS. In order to have the common CC&BS feature, every network domain, i.e. WLAN and UMTS, must provide separate charging data records (CDR) that carries charging information for the users in that domain.

Authentication Considerations

For UMTS, the users are authenticated using SIM based authentication by the HLR. In our design, we assume that the HLR entity functionality is integrated with both the SGSN and AAA server.

For WLAN users, the WLAN AAA server will perform the authentication and accounting services independently. Thus, the authentication procedure in this architecture is totally isolated in the two networks.

Charging Considerations

For UMTS, CDRs are generated from the GSN nodes and AAA server. These entities are already interfaced with the CC&BS. For that matter, according to 3gpp specifications in [10] and [11], GSN nodes can deliver their CDR through a charging gateway or, alternatively, they can implement the gateway tunneling procedures to act as charging gateway. So, as a design goal of cost effectiveness with considerable minor changes in the GSN functionality, we decided to have the assumption that GSN nodes implement the charging gateway functionality (CGF) [11]. The records from AAA server will be tunneled and pass through SGSN and GGSN to the CC&BS.

For WLAN, AAA server should send CDRs to the CC&BS of the same service provider of the UMTS. Hence, AAA server is connected to a charging gateway (CG) that implements CGF. The CG, then, connects to the operators network and delivers the CDRs which will be tunneled to the CC&BS of the service provider.

Having these considerations, we propose the following design for the open coupling scenario in figure 1.

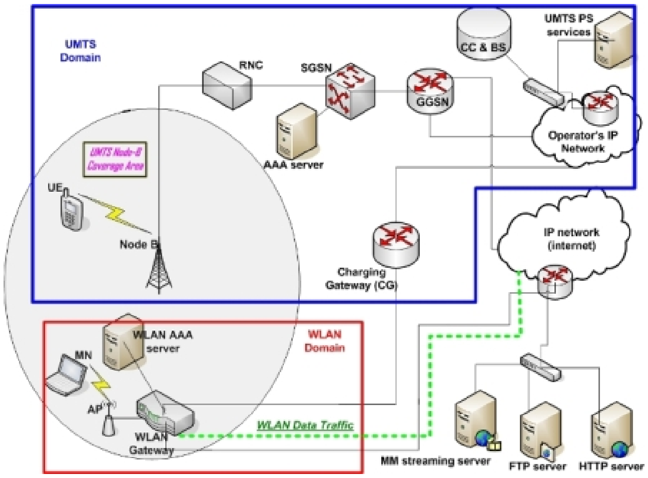


Figure 1 Open coupling

Design Details

As can be seen in figure 1, the WLAN domain consists of a mobile node MN, an access point AP connected to the WLAN gateway and a WLAN AAA server that is also connected to the WLAN gateway. The WLAN gateway has then a connection to an IP network, e.g. the Internet, which is independent of the core of UMTS network.

The UMTS domain will have its normal access to the UMTS packet switched PS services as well as the external IP access along the usual path, i.e. RNS, SGSN and GGSN. For authentication in WLAN, the AP will be running the AAA client application and communicating with the WALN AAA server via the WLAN gateway. For charging, in UMTS domain, AAA server will send tunneled CDR through the SGSN and then to the GGSN till it reaches the CC&BS in the operator's network. SGSN and GGSN will also send their CDRs along that path. Notice that we assumed the distributed CGF in SGSN and GGSN; hence no interface exists between these entities and a charging gateway.

WLAN users will generate their CDRs from the WLAN AAA, which will be tunneled through the WLAN gateway which is interfaces directly to a charging gateway CG. This CG implements the CGF functionality and delivers the CDRs to the operators CC&BS.

2) Loose Coupling

Loose coupling is in fact an enhanced version of the open coupling scheme. The loose coupling inherits the common charging system from the open coupling scheme with the addition of providing 3G based PS services for users in WLAN. Thus, it is required to provide a connectivity mechanism for the WLAN to the PS domain of the UMTS core as well as common authentication system but with the minimal impact on the WLAN side as recommended by 3GPP specifications in [10]. Loose coupling actually can be the basic approach of targeting all integration scenarios from 3 to 6 since the requirements on these scenarios are more concerned with the handover and QoS than the architecture. [14]

Data Routing Consideration

To provide 3G-based IP accessibility for WLAN users, their data traffic targeting such data networks will go through two new main entities:

WAG: WLAN Access Gateway is a router responsible for routing the WLAN traffic to an appropriate PLMN according to the enforcement policies identified for the subscriber. These policies are known after the authentication is done by the AAA procedures.

PDG: is another router, which is responsible for routing WLAN data traffic in the IP network or the PDN associated with PS services such as WAP and MMS. It can be thought as an end tunnel for WLAN user data traffic based on 3G IP access.

Authentication Consideration

For WLAN users, the WLAN AAA server should retrieve the subscription information from the HPLMN of the user to correctly do the authentication. Thus, the AAA server will act as a proxy server in this regard. The authentication messages exchanged between the AAA proxy and the AAA server in the home network will go through tunneling procedures within the IP network of the service provider. [16]

Charging Consideration

For UMTS users, charging will be the same as in the open coupling scheme. However, for WLAN users, the introduction of WAG and PDG for the new data traffic path, adds the requirement of collecting CDRs from these entities as well. By following the same assumptions in open coupling for UMTS users, the WAG and PDG can have the CGF integrated in them, especially that they are new entities and subject to more enhancements. We assume this integrated functionality in our design. On the other hand, for WLAN users who are not authorized to have the 3GPP IP accessibility, a CG is required to account for their data traffic. However, as their CDRs will be generated from the WLAN AAA server, which is interfaced with the WAG, the assumption of implementing the CGF in WAG will make it possible to configure the WAG as a CG for these CDRs.

Having these considerations, we propose the design of the loose coupling scheme as in figure 2.

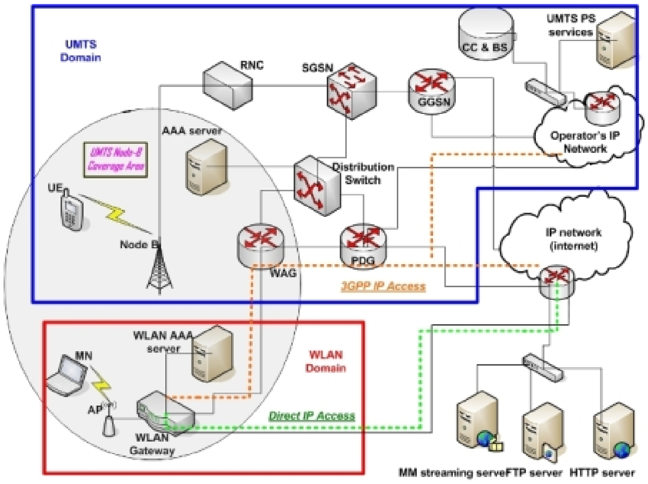


Figure 2 Loose coupling

Design Details

The WLAN domain in loose coupling has more entities in the structure and interfaces than the open coupling case. In loose coupling, we do not have an interface from the WLAN gateway to a separate CG as in open coupling. This is because WAG, introduced in loose coupling, will act as a CG. Moreover, a link connecting the WLAN gateway with the WAG for exchanging AAA common authentication procedure messages will be added in the loose coupling scenario. WLAN users can have a direct IP access route or a 3 gpp based access. In the latter case, traffic will go through the WAG and the PDG to the IP network. In the same manner, UMTS PS services network will be accessed.

The UMTS domain will have its normal access to the UMTS PS services as well as the external IP access along the usual path, i.e. RNS, SGSN and GGSN. For charging, in UMTS domain, the CDRs are handled in the same manner discussed under the open coupling scenario. However, for WLAN domain, three entities will generate the CDRs, i.e. WLAN AAA, WAG and PDG. We assumed that WAG and PDG implements CGF. Thus, WLAN AAA server will send tunneled CDR through the WAG and then to the PDG till it reaches the CC and BS in the operator's network. WAG and PDG will also send their CDRs along that path. Also, for WLAN users not accessing the UMTS core IP, their generated CDRs from the WLAN AAA server will consider the WAG as their CG.

B. *Handover Schemes Design and Implementation*

Before we describe our design and implementation of the handover solutions, we state the following notes and assumptions.

- This work simulates the handover operation at the target layer. Thus, the handover at the physical layer L1 and the data link layerL2 is not considered in the related calculations. This is based on the assumption that the handover operation at L1 and L2 is common among all handover schemes.

- We assume that the UE has a dual mode operation that triggers the new network signals.

- The handover delay or latency is calculated based on the handover transaction messages specified by the protocol. This is the common method used to mathematically account for the handover (HO) latency found in the literature. Thus, according to this definition of the handover delay, the latency is given by:

HO delay = ∑ (HO message trans. delays) + network delay

- In this work we are interested in the analysis of the handover operation in the direction from the UMTS to the WLAN. The other direction is one of the targets of a future work.

1) *mSCTP solution for vertical handover from UMTS to WLAN*

In the implementation of mSCTP as a mobility management solution, no addition or modification of network components is required. The only required change is to run mSCTP as a transport protocol on the end points, i.e. clients and servers. The following is a brief description of the mSCTP procedure to support vertical handover.

- A mobile client MC belongs initially to a UMTS network as his PLMN obtains an IP (IP_UMTS) that is used to communicate with the Fixed Server FS. This FS is usually an application server.
- When this MC enters the coverage domain of a WLAN to which he is subscribed as well, he obtains a new IP address, while he is still running under the UMTS coverage.
- The MC sends an ASCONF message to the FS asking for the addition of the new IP (IP_WLAN) while the session is still running under UMTS coverage. FS then sends an ACK message to the MC.
- When the WLAN signal becomes strong enough, MC decides to handover. Thus, he sends an ASCONF message to the FS to change the primary address from (IP_UMTS) to (IP_WLAN). FS, then, acknowledge this message.
- The traffic starts to flow based on WLAN IP as the destination address.

mSCTP Handover Design and Implementation

The implementation of mSCTP does not require any addition of new entities in the network. The handover procedure is implemented in already existing entities. The only modification required is to make the involved entities support the mSCTP protocol. In this work, since we are interested only in the handover procedure, we simulate only this part of the protocol.

Two entities are involved in the handover process, i.e. the end user (user equipment UE or mobile node MN) which is referred to as the mobile client MC, and the application server like HTTP or FTP server, known as the fixed server FS. Although it is possible to have the FS as a separate entity acting as a proxy, we preferred to target the minimum requirements from the cost effectiveness perspective.

The following steps indicate the handover operation from UMTS to WLAN under mSCTP protocol:

- A UE running an application triggers a WLAN signal and decides to start the handover operation at a predetermined time T since the L1 and L2 handover is not simulated.
- UE obtains the WLAN new IP. (Not accounted in the handover calculation)
- Handover transactions take place between the UE and the associated FS, which is the application server (HTTP, FTP or MM server).
- Finally, the UE becomes a MN and continues his application with the WLAN features.

Figure 3 indicates the mSCTP handover transactions.

2) *Mobile IP solution for vertical handover from UMTS to WLAN*

When the UMTS terminal i.e. user equipment (UE) decides to handover from UMTS to WLAN, it simply disables its UMTS protocols and uses the IP stack. If we Assume that UMTS is the UE home network and the UE has switched to the WLAN access mode so it becomes a WLAN mobile node (MN), then, the following steps are executed to perform the handover operation:

- MN sends an agent discovery solicitation in the foreign network, i.e. the WLAN network.

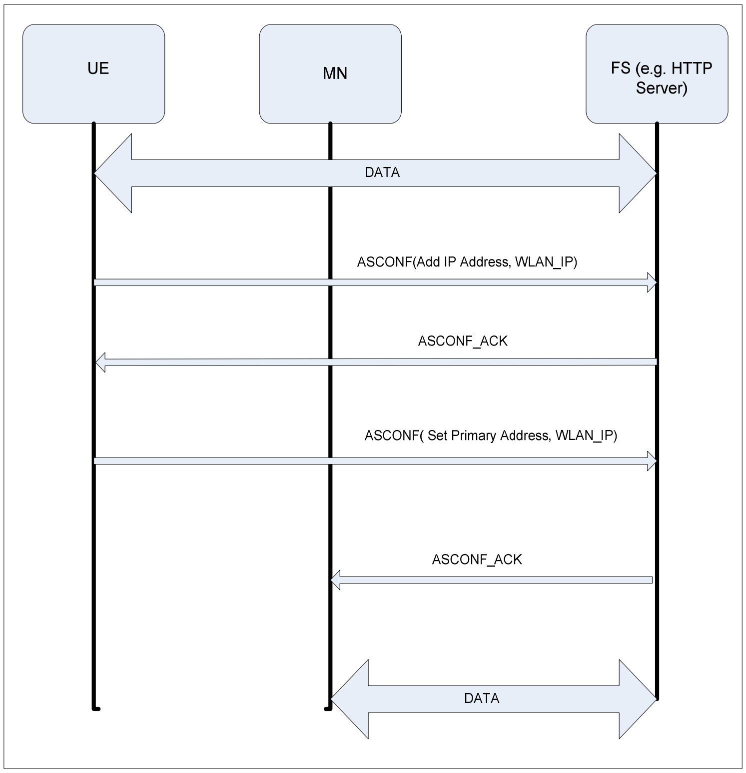


Figure 3 mSCTP Handover Transactions

- The FA responds by sending agent advertisement message that contains the CoA.
- MN sends a registration request to the FA, which forwards it to his HA to update the MN CoA.
- The HA sends a registration reply to the MN that is delivered via the FA.
- Data starts to flow to the MN by the interception of the HA or directly through the FA if route optimization is adopted.

Mobile IP Handover Design and Implementation

As mentioned previously, in mobile IP, the handover operation involves three entities:

- The mobile roaming node, which is in our case the UE.
- The home agent HA.
- The Foreign agent FA.

Both HA and FA are new entities that have to be added to the network. The adopted scenario assumes that the user is roaming in his HPLMN in a cell that covers the WLAN domain as well. As a result, the suitable choice for the HA is to place it at the nearest point in the UMTS where the user subscription information is located, i.e. HLR. Thus we implement the HA as a server connected to the SGSN that provides the connectivity to HLR as well as the routing roles. Notice that this HA can act as a FA.

FA should be associated with the WLAN. There are two options; either to place at the packet data gateway (PDG) or at the wireless LAN access gateway (WAG) which are routing elements identified by the UMTS/WLAN integration requirements as per 3GPP release 7. However, we choose to associate the FA as a server connected to the WAG for the following reasons:

- In open coupling, PDG does not exist while WAG is emulated by the WLAN gateway.
- In loose coupling, WAG (FA) communicates with SGSN (HA) at the same level while the PDG is ahead of WAG that adds an additional level of delay.

The following steps indicate the handover operation under mobile IP protocol:

- A UE running an application triggers a WLAN signal and decides to start the handover operation at a predetermined time T since the L1 and L2 handover is not simulated.
- UE obtains the WLAN new IP. (Not accounted in the handover calculation)
- UE switches to the WLAN MN mode.
- Handover transactions, take place between the MN and the associated FA and HA.
- It is assumed that there is a route optimization following the HO operation that prevents triangular routing and makes data go directly to the MN without passing through the HA.

Figure 4 indicates the Mobile IP transactions.

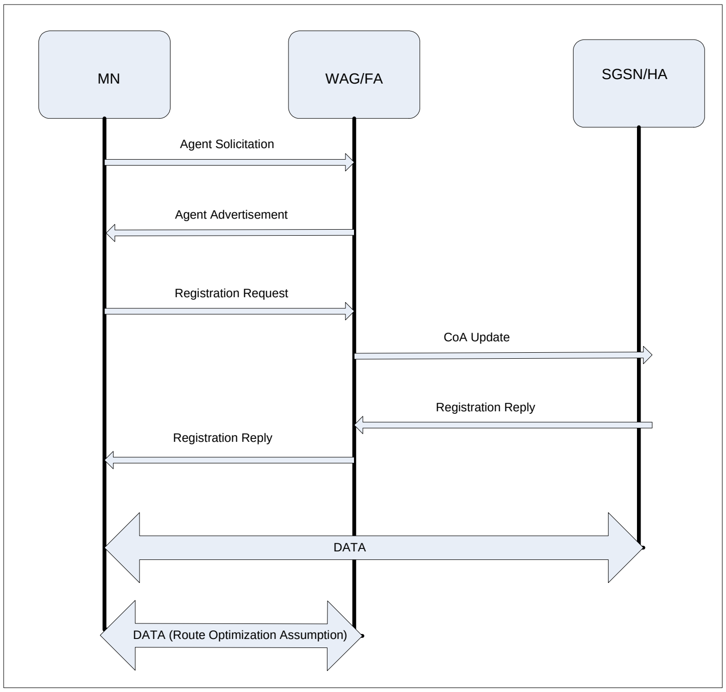


Figure 4 Mobile IP Handover Transactions

3) *SIP solution for vertical handover from UMTS to WLAN*

SIP is an application layer protocol to handle mobility issues. The implementation of SIP requires the addition of SIP servers that will keep track of mobile hosts MH. However, this function can be integrated in the application servers as well. When an MH moves from one network to another and wants to handoff during an active session, it should send an INVITE message to the SIP server updating its location within the contact field and getting the redirect information for the correspondent CH. After that, it sends a re-INVTE message to the CH to tell it about its new contact information.

In UMTS to WLAN handover, SIP encounters two phases.

- The MH needs to acquire its new IP from the DHCP server.
- The normal SIP invite operation is exchanged between the MH and the CH.

SIP Handover Design and Implementation

In the implementation of SIP, the handover operation from the 3GPP to the WLAN system does not need any addition of new entities to network. The handover operation needs the involvement of two entities only:

- The mobile roaming node, which is in our case the UE or Mobile Host (MH).
- The correspondent host, CH, which is the second party involved in the current session, such as the service server.

Even though the SIP protocol requires the addition of new servers for the support of the handover from the WLAN to the 3GPP system, these new entities are not considered in our model. This is due to the scope of this study that is only focused on the one direction of the handover, i.e. from the UMTS to the WLAN. The implementation assumes that SIP is supported by all the involved parties in the session.

The following steps indicate the handover operation under SIP protocol:

- DHCP registration procedure: This phase consists of four messages.
    - The UE/MH sends a DHCP DISCOVER message once it identifies the presence of the WLAN.
    - The proper DHCP server sends a DHCP OFFER to the MH.
    - A DHCP REQUEST is sent by the MH to the correspondent DHCP server as a confirmation to the offer.
    - Finally, The DHCP server sends a DHCP ACK to the MH with the required registration info such as the IP address.
- SIP message exchange: There are two main transactions in this phase.
    - The MH sends a SIP INVIT directly to the CH to reestablish the connection given the new address and location.
    - Once the CH is updated with the new information it sends a SIP OK directly to the MH while starting the data flow.

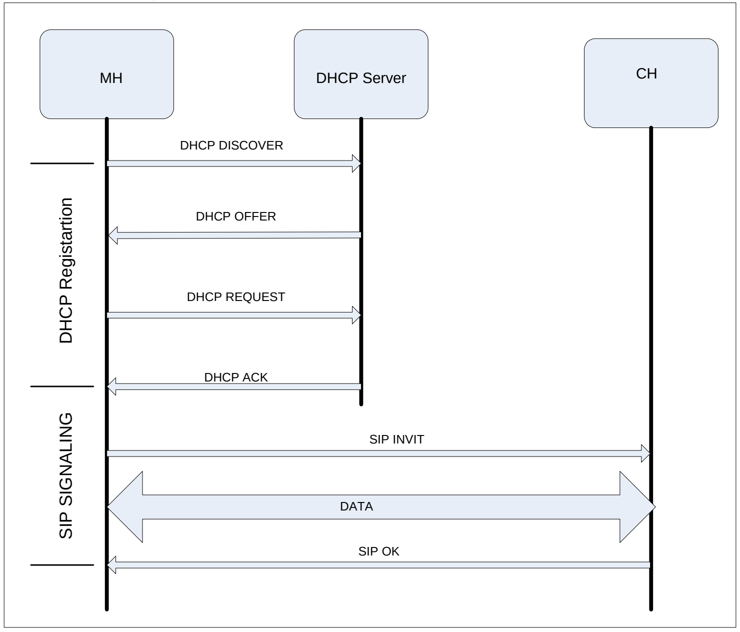


Figure 5 SIP Handover Transactions

Figure 5 shows the order and direction of the previously mentioned messages.

V. SIMULATION RESULTS AND ANALYSIS

Our evaluation study is divided into two phases. In the first phase, we analyses the integrated system under the two coupling schemes without handover consideration. In the second phase, we compare the integrated systems after the implementation of the handover schemes.

A. *Loose and Open Coupling Architectures Simulation Results and Analysis*

In this section, we report the simulation results for the adopted infrastructures for both loose and open coupling. Note that the purpose in this phase is to evaluate the network design and characterize its performance.

Therefore, there is no consideration for any handover scenarios or any mobility management protocols in this phase. Furthermore, the performance metrics used in this part of the work are application response time and total application throughput.

We simulated seven traffic combinations considering three application types: HTTP, FTP and video conferencing. For each of these combinations, three different load points are considered in the WLAN side, i.e. 1000, 4000, and 7000 users. Nevertheless, the simulations did not show any significant effect for changing the load on the application response time or the throughput. All the considered load points generated almost the same figures with minor differences. This can be clearly noticed in the figures 6, 7, 8, 9, 10 and 11. Tables 1, 2 and 3 show the application throughput and response time for each application mix in the case of 7000 users.

Application Mix Percentages (FTP-HTTP-Multimedia)	File Response Time (seconds)		Received Throughput (Byte/s)	
	Loose	Open	Loose	Open
20-40-40	42.399	42.4359	45717.18	45717.177
10-45-45	21.625	21.574	22860.03	22860.034
40-30-30	84.167	84.3377	91431.46	91431.463
35-10-55	73.7419	73.7419	80002.89	80002.89
10-60-30	21.625	21.574	22860.03	22860.034
35-55-10	73.7419	73.7419	80002.89	80002.89
10-30-60	21.625	21.574	22860.03	22860.034

Table 1 : FTP Traffic Results

Application Mix Percentages (FTP-HTTP-Multimedia)	Page Response Time (seconds)		Received Throughput (Byte/s)	
	Loose	Open	Loose	Open
20-40-40	12.1164	14.0246	45746	45740.893
10-45-45	15.6662	12.85714	51460.16	51460.16
40-30-30	10.509731	8.76511	34316.777	34316.7
35-10-55	3.205	4.0533	11390	11454.957
10-60-30	20.8677	20.8257	68211.9	68596.671
35-55-10	19.146592	15.642244	62889.26	62889.22
10-30-60	10.509731	8.7695	34316.777	34316.7

Table 2 : HTTP Traffic Results

Application Mix Percentages (FTP-HTTP-Multimedia)	End-to-end Packet Delay (seconds)		Received Throughput (Byte/s)	
	Loose	Open	Loose	Open
20-40-40	0.01134	0.010844	45921	45921
10-45-45	0.011620	0.011532	51429	51429
40-30-30	0.011662	0.0115547	34286	62721
35-10-55	0.011714	0.011527	62721	62721
10-60-30	0.011662	0.011547	34286	34286
35-55-10	0.01134	0.011568	11429	11429
10-30-60	0.0114	0.011594	68393	68393

Table 3 : Video Conferencing Traffic Results

Table 4 shows the different specification for the applications in each traffic mix.

Application Mix Percentages (FTP-HTTP-Multimedia)	FTP		HTTP		Video Conferencing	
	File Size (Byte)	Inter request time (Seconds)	Page Size (Byte)	Inter request time (Seconds)	Packet Size (Byte)	Number of Requests (packets/s)
20-40-40	3.20e+07	360	1.07e+07	60	5000	35.556
10-45-45	1.60e+07	360	1.20e+07	60	5000	40
40-30-30	6.40e+07	360	8.00e+06	60	5000	26.667
35-10-55	5.60e+07	360	2.67e+06	60	5000	48.889
10-60-30	1.60e+07	360	1.60e+07	60	5000	26.667
35-55-10	5.60e+07	360	1.47e+07	60	5000	8.8889
10-30-60	1.60e+07	360	8.00e+06	60	5000	53.333

Table 4 : Traffic Mixes Specifications

Our simulation results show no significant difference between the open and loose coupling (see figures 12, 13, 14 and 15). The only factor that contributes to the response time and throughput is the packet size or the page size in the HTTP case. The effect of the load, i.e. number of users in the network, was transparent to both response time and throughput according to our OPNET simulations. This can be explained by the choice of our physical links in the core network that would make their utilization roughly around 45% with the 7000 users.

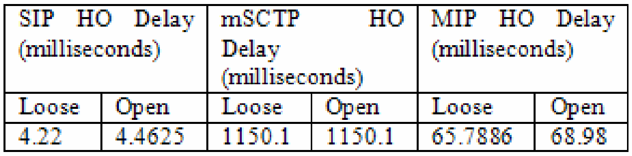
B. Handover Delay Simulation Results and Analysis

In this phase, we simulated the three previously reported handover schemes with the two integration approaches i.e. loose and open coupling. The simulation considers different scenarios in which a user equipment UE encounters a handover operation while he is running an HTTP, FTP and a multimedia application. Theses applications are run under the same variable traffic mix reported in table 4 with the same traffic loads used in the first phase.

Two metrics are used to study the vertical handover operation in this work.

- Handover delay: we define handover delay as the amount of time required to complete the transaction messages exchange in the handover operation for different protocols. This represents the time of data flow disruption at the target layer.
- Response time: it is the page response time for HTTP. In FTP the metric considers the file download response time. In multimedia application we measure the end-to-end packet delay. In all applications, the measurement is taken as the average overall value for the application.

Table 5 summarizes the handover delay results. (Notice that these figures are independent of the application type, traffic size and network load)



SIP HO Delay (milliseconds)		mSCTP HO Delay (milliseconds)		MIP HO Delay (milliseconds)	
Loose	Open	Loose	Open	Loose	Open
4.22	4.4625	1150.1	1150.1	65.7886	68.98

Table 5: Handover latency for different handover schemes

Based on the handover latency results, the system can be analyzed as follows. The simulation shows that SIP provides the best results in term of handover delay. As the table shows, the protocol works with almost the same response in both loose and open coupling with an average delay of 4.3 milliseconds. This is an expected result due to the fact that this protocol has one intermediate message that intercepts the data flow in the handover case. Note that if the 3GPP coverage includes the WLAN zone, the UE/MH will continue to receive the data from the CH through the old path, i.e. UMTS route. The session will be disconnected once the MH sends the SIP INVIT.

- Handover in mobile IP in the case of open coupling seems slightly longer than the case in loose coupling. This is because the handover transactions are all done in the WLAN domain. As a result, in the loose coupling case, the signaling path between the HA (SGSN) and the FA (WAG) is direct and shorter than the path in open coupling in which the transactions encounter a longer path through the IP network.
- mSCTP handover exhibits the highest delay among the other handover schemes due to:
    - mSCTP has more transactions (4 messages) compared with SIP (2 messages).
    - Most of the transactions (three out of four) in mSCTP are exchanged while the terminal is in the UMTS mode. In this mode, the data rate is much lower than the case in the WLAN mode in which the MH in the SIP operation exchanges its messages with CH.
    - In contrast to mobile IP, mSCTP transactions propagate along the data path and, thus, suffer similar congestion and link utilization problems faced by application traffic.
    - mSCTP clients should communicate with the application servers, and thus, handover messages always travel the longest path in the network. In contrast, mobile IP transactions are exchanged among FA, HA and MN which are physically connected in a more optimized topology.

With respect to response time figures, according to our definition of handover latency, the only effect of handover operation on the response time will be decreasing the effective simulation period as the handover part of the simulation will not have a significant traffic. This implies that as the handover latency increases, the effective simulation period in which we have traffic will decrease and, as a consequence, the overall response time should also decrease. However, these differences are not expected to be very significant or accurately applicable as they depend on the simulation environment. Basically, this implies that there should be no difference in response time for the application between the results before and after considering the handover. Thus, in this work, the difference in response time among handover schemes does not impact the decision of selecting the best handover scheme. For example, in figures 16 and 17, the page response time of 60% load of HTTP in MIP is 21.583 seconds whereas it is 18.76 seconds in mSCTP in which we have longer handover delay. Also in the case of multimedia traffic, the end-to-end packet delay ranges between 11.5 ms and 11.8 ms in the case of MIP while it has a smaller range of 11.5 ms to 11.7 ms in mSCTP as indicated in figures 19 and 20.

VI. CONCLUSION

In this work we performed an extensive comparative study of different UMTS/WLAN integration alternatives. These alternatives consider open and loose coupling schemes each with three possible handover solutions, i.e. SIP, mSCTP and MIP. Using OPNET simulations, we were able to evaluate these solutions. The results show that open and loose coupling have almost similar performance figures in terms of throughput and response time. Moreover, the SIP handover solution exhibits the least handover latency among other. Thus, the recommended integration scheme is to adopt loose coupling, as it provides more services compared to open coupling, with SIP as the mobility management solution.

As a future and continuation work, the handover operation from WLAN to UMTS will be studied. Moreover, possible optimization approaches for mobility management schemes would be considered to provide more QoS features.

REFERENCES

[1] WEI WU; NILANJAN BANERJEE; KALYAN BASU and SAJAL K. DAS. Sip-based vertical handoff between wwans and wlans. *IEEE Wireless Communications*, pages 66–72, June 2005.

[2] Introduction to mobile ip [ip tunneling]. http://www.cisco.com/en/US/tech/tk827/tk369/ technologies white paper09186a00800c9906.shtml#53663, October 2001.

[3] Reshan Samarasinghe; Vasilis Friderikos and A.H. Aghvami. Aanalysis of intersystem handover: Umts fdd & wlan. Centre for Telecommunications. Research; King's College London, London Communications Symposium, 8th-9th September 2003.

[4] Shiao-Li Tsao; Chia-Ching Lin. Design and evaluation of umts-wlan interworking strategies. In *Vehicular Technology Conference*, volume 2,pages 777–781. IEEE, September 2002.

[5] L. Ong and J. Yoakum. An introduction to the stream control transmission protocol (sctp). http://www.ietf.org/rfc/rfc3286.txt, May 2002.

[6] Henning G. Schulzrinne and Jonathan D. Rosenberg. The session initiation protocol: Providing advanced telephony services across the internet. *Bell Labs Technical Journal*, pages 144–160, October-December 1998.

[7] Elin Wedlund and Henning Schulzrinne. Mobility support using sip. *Second ACM/IEEE International Conference on Wireless and Mobile Multimedia (WoWMoM99), (Seattle, Washington)*, August 1999.

[8] Wei Xing; Holger Karl; Adam Wolisz and Harald Muller. M-sctp: Design and prototypical implementation of an end-to-end mobility concept. In *Proc. of 5th Intl. Workshop The Internet Challenge: Technology and Applications*, October 2002.

[9] LI MA; FEI YU; and VICTOR C. M. LEUNG. A new method to support umts/wlan vertical handover using sctp. *IEEE Wireless Communications*, pages 44–51, August 2004.

[10] 3GPP TS 23.234. 3gpp system to wireless local area network (wlan) internetworking;. System Description Release 7, 3rd Genration Partnership Project, December 2005.

[11] 3GPP TS 32.015. Telecommunication management; charging management; call and event data for the packet switched (ps) domain. Technical specification, 3rd Generation Partnership Project, Decmber 2003.

[12] 3GPP TR 22.934 6.2.0. Feasibility study on 3gpp syrtem to wlan lntewarking (release 6). Technical specification, 3rd Genration Partnership Project, September 2002.

[13] F. Siddiqui; S. Zeadally; S. Fowler. A novel architecture for roaming between 3g and wireless lans. pages 101–105. IEEE Conference, 1st International Conference on Multimedia Services Access Networks, June 2005.

[14] Yi-Bing Lin and Ai-Chun Pang. *Wireless and Mobile All-IP Networks*. Wiley Publishing, Inc., 2005.

[15] APOSTOLIS K. SALKINTZIS. Interworking techniques and architectures for wlan/3g integration toward 46 mobile data networks. *IEEE Wireless Communications*, pages 50–61, June 2004.

[16] Apostolis K. Salkintzis. Wlan/3g interworking architectures for next generation hybrid data networks. *IEEE International Conference on Communications*, 7:3984 – 3988, June 2004.

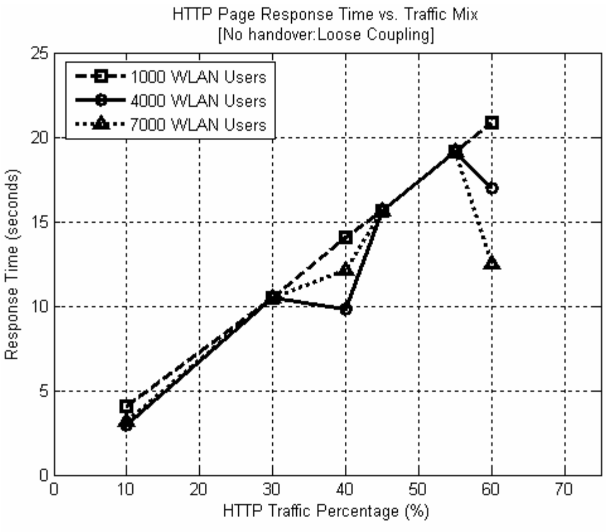


Figure 6 HTTP response time for loose coupling

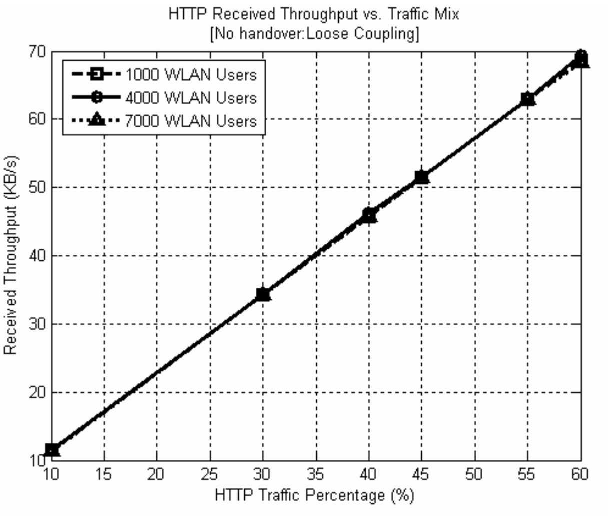


Figure 7 HTTP throughput for loose coupling

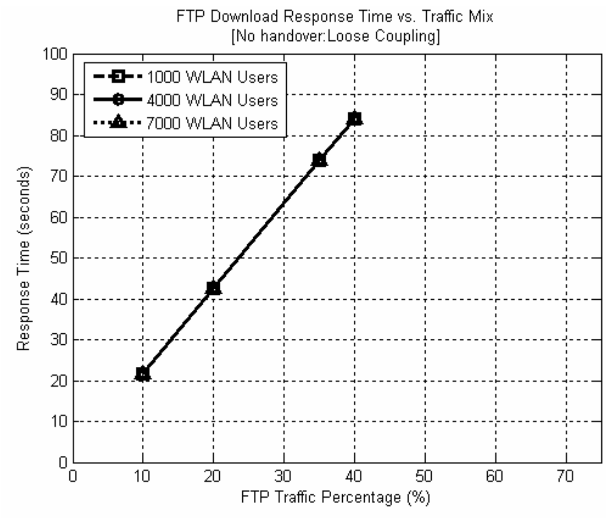


Figure 8 FTP response time for loose coupling

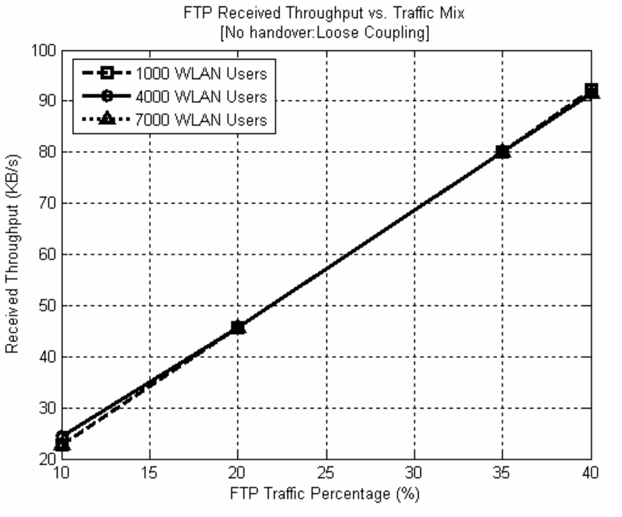


Figure 9 FTP throughput for loose coupling

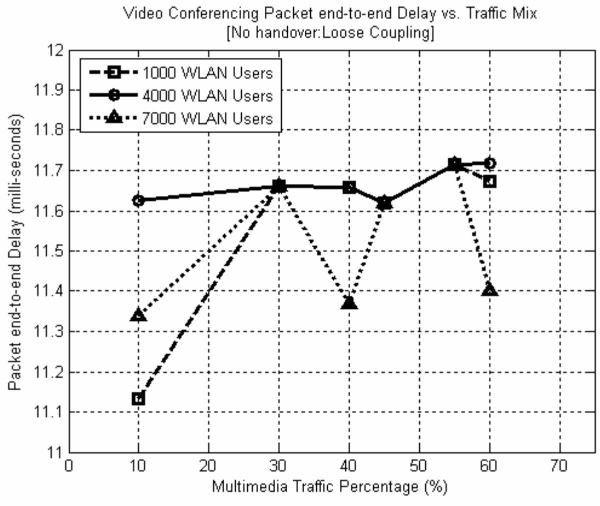


Figure 10 MM response time for loose coupling

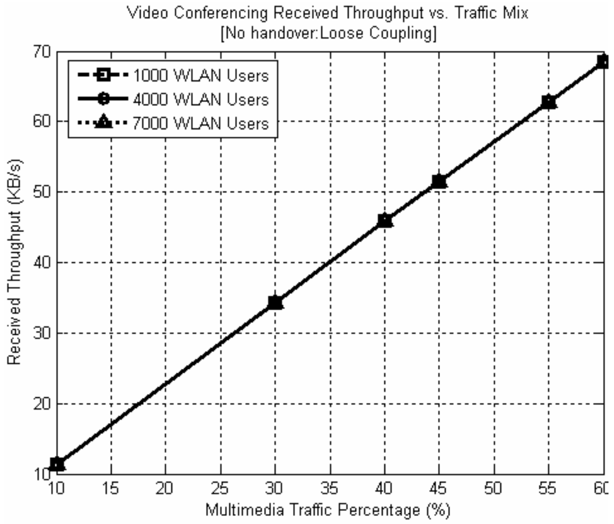


Figure 11 MM throughput for loose coupling

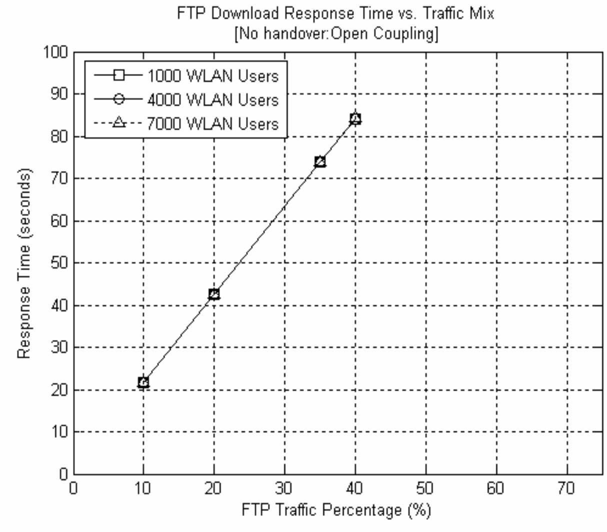


Figure 12 FTP response time for open coupling

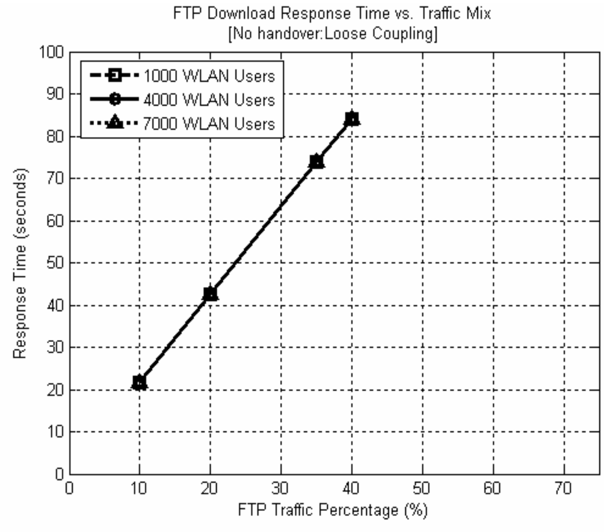


Figure 13 FTP response time for loose coupling

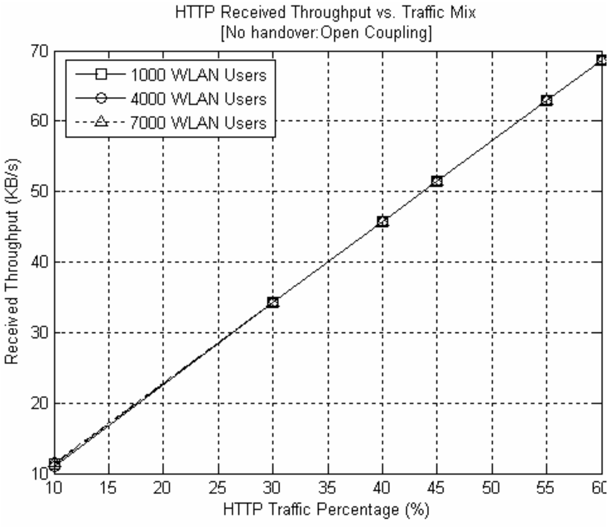


Figure 14 HTTP response time for open coupling

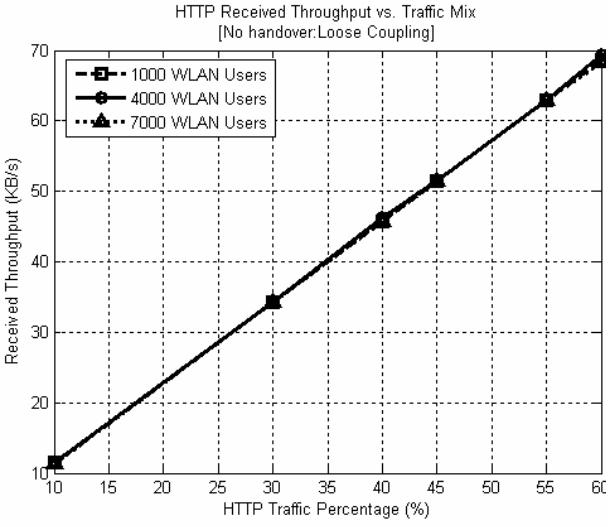


Figure 15 HTTP response time for loose coupling

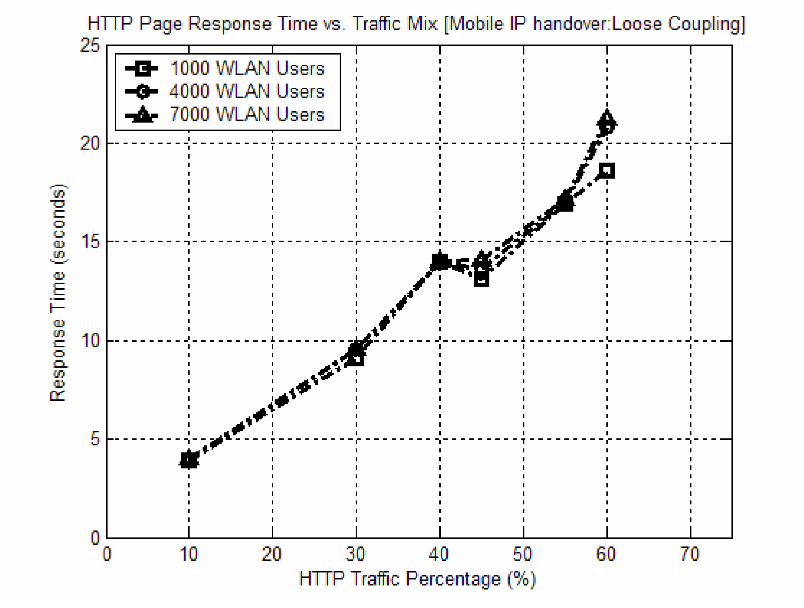


Figure 16 HTTP page response time for MIP

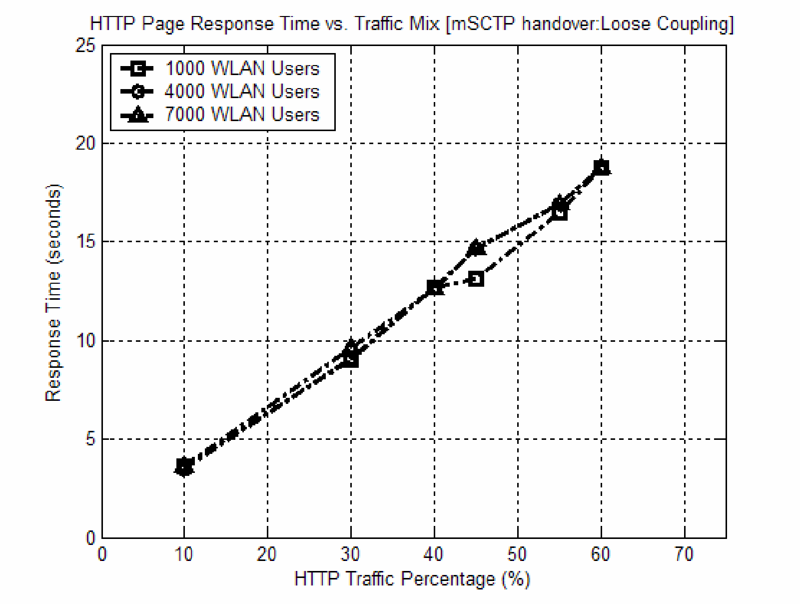


Figure 17 HTTP page response time for mSCTP

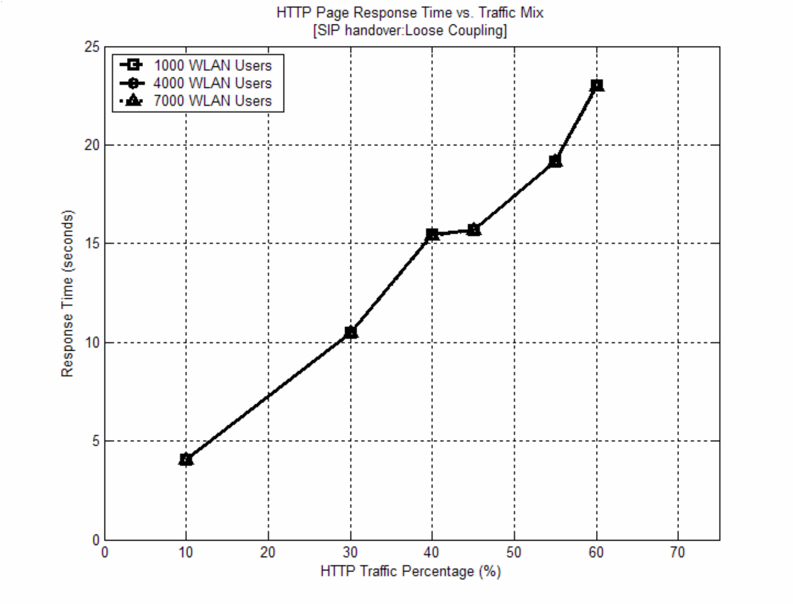


Figure 18 HTTP page response time for SIP

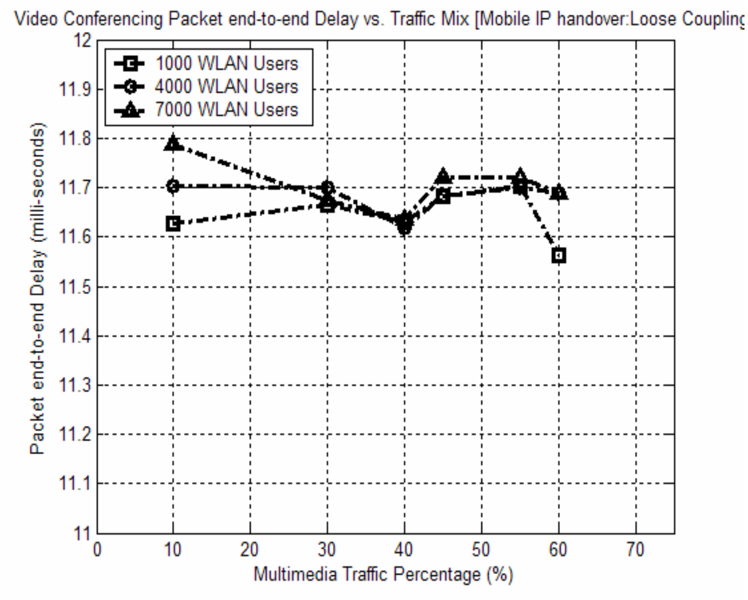


Figure 19 multimedia end-end packet delay for MIP

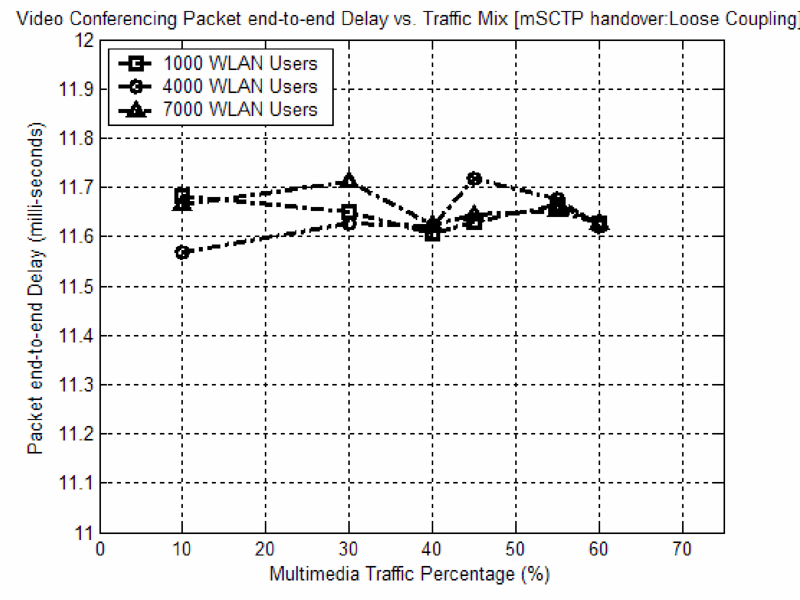


Figure 20 multimedia end-end packet delay for mSCTP

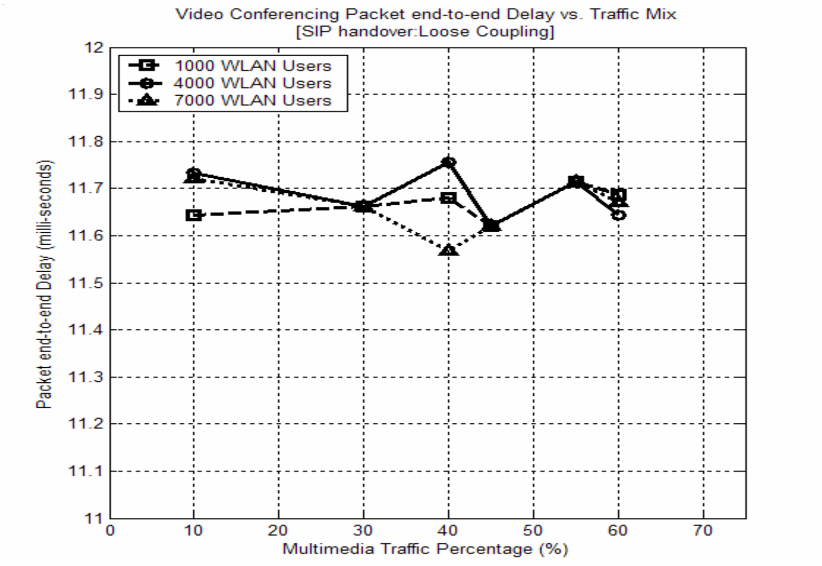


Figure 21 multimedia end-end packet delay for SIP